

SELF-EXCITING HURDLE MODELS FOR TERRORIST ACTIVITY

BY MICHAEL D. PORTER AND GENTRY WHITE

GeoEye Analytics and University of Queensland

A predictive model of terrorist activity is developed by examining the daily number of terrorist attacks in Indonesia from 1994 through 2007. The dynamic model employs a shot noise process to explain the self-exciting nature of the terrorist activities. This estimates the probability of future attacks as a function of the times since the past attacks. In addition, the excess of nonattack days coupled with the presence of multiple coordinated attacks on the same day compelled the use of hurdle models to jointly model the probability of an attack day and corresponding number of attacks. A power law distribution with a shot noise driven parameter best modeled the number of attacks on an attack day. Interpretation of the model parameters is discussed and predictive performance of the models is evaluated.

1. Introduction. Since the attacks of September 11, 2001 there has been a substantial increase in the amount of resources dedicated to combating terrorism. As a single example, the Congressional Research Services estimates that as of September 2010 \$1.2 trillion USD has been spent on the Global War on Terror, including \$28.5 billion USD on enhanced security [Belasco (2010)]. Despite this increase in spending, there are no accurate measures of the effectiveness of these counter-terrorism efforts [Lum, Kennedy and Sherley (2006), Perl (2007)]. According to the Congressional Research Service, one of the keys to combating terrorism effectively is to understand, and have correctly specified models for, terrorist activity:

“Better understanding of the dynamics of terrorism allows for a more complete picture of the complexities involved in measuring success or failure and can assist the 110th Congress as it coordinates, funds, and oversees anti-terrorism policy and programs” [Perl (2007)].

A dynamic model for terrorism should describe terrorist activity in a way that is consistent with both the theoretical framework for terrorism and the observed data. Defining accurate models of terrorist activity is important not just from the standpoint of assessing the effectiveness of counter-terrorism efforts but also for use as a tool to predict the future risk of terrorist attack.

A majority of publications in the literature use one of three basic modeling approaches to terrorism. The work of Enders and Sandler (1993, 2000, 2002, 2006)

Received June 2011; revised September 2011.

Key words and phrases. Self-exciting, hurdle model, shot noise, terrorism, Indonesia, Hawkes process, Riemann zeta, point process, probability gain.

SELF-EXCITING HURDLE MODELS FOR TERRORIST ACTIVITY

BY MICHAEL D. PORTER AND GENTRY WHITE

GeoEye Analytics and University of Queensland

A predictive model of terrorist activity is developed by examining the daily number of terrorist attacks in Indonesia from 1994 through 2007. The dynamic model employs a shot noise process to explain the self-exciting nature of the terrorist activities. This estimates the probability of future attacks as a function of the times since the past attacks. In addition, the excess of nonattack days coupled with the presence of multiple coordinated attacks on the same day compelled the use of hurdle models to jointly model the probability of an attack day and corresponding number of attacks. A power law distribution with a shot noise driven parameter best modeled the number of attacks on an attack day. Interpretation of the model parameters is discussed and predictive performance of the models is evaluated.

1. Introduction. Since the attacks of September 11, 2001 there has been a substantial increase in the amount of resources dedicated to combating terrorism. As a single example, the Congressional Research Services estimates that as of September 2010 \$1.2 trillion USD has been spent on the Global War on Terror, including \$28.5 billion USD on enhanced security [Belasco (2010)]. Despite this increase in spending, there are no accurate measures of the effectiveness of these counter-terrorism efforts [Lum, Kennedy and Sherley (2006), Perl (2007)]. According to the Congressional Research Service, one of the keys to combating terrorism effectively is to understand, and have correctly specified models for, terrorist activity:

“Better understanding of the dynamics of terrorism allows for a more complete picture of the complexities involved in measuring success or failure and can assist the 110th Congress as it coordinates, funds, and oversees anti-terrorism policy and programs” [Perl (2007)].

A dynamic model for terrorism should describe terrorist activity in a way that is consistent with both the theoretical framework for terrorism and the observed data. Defining accurate models of terrorist activity is important not

Received June 2011; revised September 2011.

Key words and phrases. Self-exciting, hurdle model, shot noise, terrorism, Indonesia, Hawkes process, Riemann zeta, point process, probability gain.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in <i>The Annals of Applied Statistics</i> , 2012, Vol. 6, No. 1, 106–124. This reprint differs from the original in pagination and typographic detail.

just from the standpoint of assessing the effectiveness of counter-terrorism efforts but also for use as a tool to predict the future risk of terrorist attack.

A majority of publications in the literature use one of three basic modeling approaches to terrorism. The work of Enders and Sandler (1993, 2000, 2002, 2006) and Barros (2003) uses time series analysis techniques including intervention analysis and vector autoregressive (VAR) models. Other more recent papers have applied group based trajectory analysis [Nagin (2005)] to analyze terrorist data. Examples of this include LaFree, Morris and Dugan (2010) who incorporate a zero-inflated Poisson distribution to account for an excess of zero counts. Other approaches use Cox proportional hazards models to directly model the time between attacks [Dugan, LaFree and Piquero (2005), LaFree, Dugan and Korte (2009)]. These last two papers attempt to account for the recent event history as influencing current events by constructing an ad hoc term that describes the recent density of attacks. While this is a step toward incorporating the dynamics related to the process history, it requires prespecification of the form of dependence.

The desire to include terms to account for recent event history is rooted in the theoretical understanding of terrorism and other politically motivated violence. The clustering hypothesis, concurrent with contagion theory [Midlarsky (1978)], explains politically motivated violent activity as a series of nonindependent events, where each event influences the probability of subsequent events. This hypothesis is well accepted in criminological and sociological approaches to terrorism. Clustering effects and contagious behavior have been demonstrated in military coups [Li and Thompson (1975)], international terrorism [Midlarsky, Crenshaw and Yoshida (1980)], airline hijackings [Holden (1986)], race riots [Myers (2000)] and insurgent activity [Townesley, Johnson and Ratcliffe (2008)].

In examining terrorism data in general there are two major issues to address. The first is that as terrorist attacks are usually rare, the daily number of terrorist attacks are often zero, but due to large coordinated attacks there are some extreme values. This combination of a large number of zeros and extreme values is poorly modeled by standard probability distributions. Second, the timing of terrorist incidents appear to be clustered, rendering standard models that assume independence unsuitable.

The first of these issues is addressed using a hurdle model [Mullahy (1986)], also known as the two-part model [Heilbron (1994)]. The hurdle model is a two component model that allows separate specifications of the probability of a zero count and the probability of a nonzero count. This allows the hurdle model to accommodate a large number of zeros in addition to some extreme counts. The hurdle model is used in a variety of applications, including in ecology for modeling counts of rare species [Welsh et al. (1996)], in public health for modeling smoking behavior [Jones (1994)], in political science for modeling proportional representation in minority electorates [Marschall, Ruhil and Shah (2010)] and network change detection

[[Heard et al. \(2010\)](#)]. For terrorism modeling, the hurdle model is preferred to the zero-inflated model [[Lambert \(1992\)](#)] which assumes that the extra zeros observed are due to censoring. The hurdle model assumes that the extra zeros are due to a separate process (the “hurdle”), which must be overcome before the number of corresponding incidents are determined. This is more reasonable for terrorism data, as it can often be assumed that sparsity of attacks is because they are indeed rare, not because they are unobserved.

The second issue regarding the clustering behavior of terrorist activity is addressed by incorporating a self-exciting component [[Hawkes \(1971\)](#)]. The self-exciting component specifies that the probability of an event is a function of the time (and possibly other aspects) of all previous events, such that the effect on the probability decreases over time. This can help account for the clustering and dynamic nature of terrorism. For example, [Holden \(1986, 1987\)](#) developed a self-exciting Poisson model that better represented the dynamics of airline hijacking and [Mohler et al. \(2011\)](#) a self-exciting space-time model for residential burglaries.

This paper combines these two concepts to examine the use of self-exciting hurdle models for terrorist activity in Indonesia and Timor-Leste between 1994 and 2007. The hurdle component is modeled as a self-exciting Bernoulli process where the form of self-excitation is dictated by the data. The nonzero counts are modeled by an extreme value distribution with parameters that are also a function of the event history. The corresponding self-exciting hurdle model is capable of making predictions and providing information about the risk of terrorist activity without any additional covariate information other than the event history. This capability is important as covariate information for terrorism data is often either missing or unreliable.

2. Data and exploratory analysis. The Global Terrorism Database [[LaFree and Dugan \(2007\)](#)] is an open-source publicly available database of over 87,000 terrorist events around the world from 1970 through 2008. The database is continually updated with new information and is believed to be the most comprehensive database of its kind. The data used here are a subset of the GTD consisting of daily counts of terrorist attacks in Indonesia¹ from January 1, 1994 through December 31, 2007.

2.1. Exploratory data analysis. The data from a training period of 1994 through 2000 was examined to inform model construction. Of the 2,557 days considered in this analysis, there were 250 terrorist attacks on 158 unique event days. [Figure 1](#), displaying the daily and semi-yearly counts of attacks, illustrates the nature of the terrorist activity. In particular, the attacks appear clustered with long stretches between attacks followed by a period of in-

¹This includes data from Timor-Leste (East Timor) which became a sovereign state in 2002.

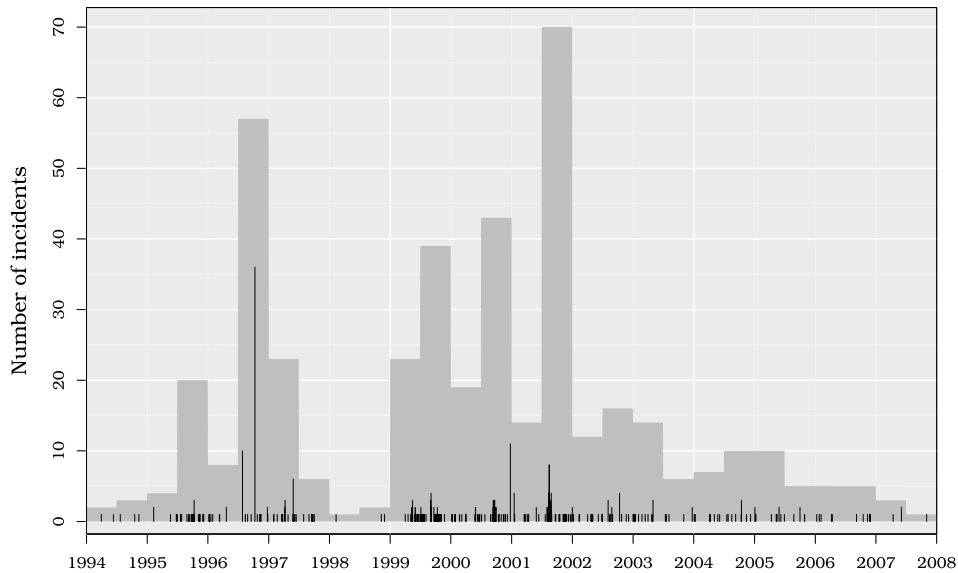


FIG. 1. *Daily number of terrorist events in Indonesia. The histogram shows the cumulative number of attacks in each six month period.*

creased activity. In addition, while most days have no attacks (93.8%), some days have multiple attacks (possibly due to planned coordinated attacks).

Table 1 shows the distribution of the observed attacks per day compared to the expected values under a Poisson and negative binomial probability distribution fitted with maximum likelihood. While the Poisson cannot capture any of the tail behavior, the negative binomial adapts better to the tail distribution but consequently underestimates the number of days with only 1 attack and overestimates the number of days with 2 and 3 attacks. A chi-square goodness-of-fit test yields a p -value of 0.00016, suggesting that

TABLE 1
Distribution of observed and expected number of Indonesian terrorist attacks per day (1994–2000)

# Attacks	# Days	Poisson	Neg.Bin
0	2,399	2,319	2,401
1	130	227	103
2	16	11	31
3	7	0	12
4	1	0	5
>4	4	0	5
AIC	–	1,988.0	1,481.8

the negative binomial is not a sufficient model. In addition, the tail is comprised of several extreme values (36, 11, 10 and 6 attacks on a single day) that exacerbate the general lack of fit. The large number of zeros coupled with the presence of several extreme values necessitate more complex models.

Figure 1 also reveals the possibility that the attack days are clustered. This is examined with Ripley’s K -function [Dixon (2002)]. The K -function is a second-order function that can reveal if clustering (or inhibition) is present in a point pattern. It is defined loosely as

$$K(t) \propto \mathbb{E}[\# \text{ of future events } \leq t \text{ from a randomly chosen event}].$$

Because this measure will be effected by an inhomogeneous attack day probability, we estimate it [Baddeley, Møller and Waagepetersen (2000), Veen and Schoenberg (2006)]:

$$\hat{K}(t) = T^{-1} \sum_{i=1}^N \frac{w_i}{\hat{p}_i} \sum_{j>i} \frac{\mathbb{1}(t_j - t_i \leq t)}{\hat{p}_j},$$

where T is the total observational time, \hat{p}_i is the estimated probability of at least one attack at day t_i , and w_i is a one-sided edge correction. The estimated probability is given by the model in (9) that includes trend and seasonality components (i.e., model BL₅ in Table 2; see Section 4).

Figure 2 plots $\hat{K}(t) - t$ which has expected value 0 if the estimated probabilities are correct and there is no unaccounted for clustering. The observed values fall well above the 95% pointwise confidence intervals (obtained from a parametric bootstrap, 1,000 simulations), showing that the data display clustering behavior not accounted for by the baseline model.

These characteristics of the terrorism data require flexible models that can handle the temporal attack patterns, clustering, an excess of nonattack days and the possibility of multiple coordinated attacks on the same day. Self-exciting hurdle models are developed in the next section to represent the complex nature of such terrorist activity.

3. Hurdle models and the self-exciting process.

3.1. *Hurdle models.* The hurdle models of Mullahy (1986) refer to a class of two component models for discrete count data. These models assume that two different processes drive the zero and nonzero counts, respectively. The *hurdle* component of the model corresponds to the probability that the count is nonzero (i.e., a terrorist action will occur), while the *count* component corresponds to the distribution of positive counts (i.e., the number of corresponding terrorist events). When the two components are combined, the hurdle model produces a probability mass function on the nonnegative integers. Additionally, the hurdle approach facilitates the use of two sim-

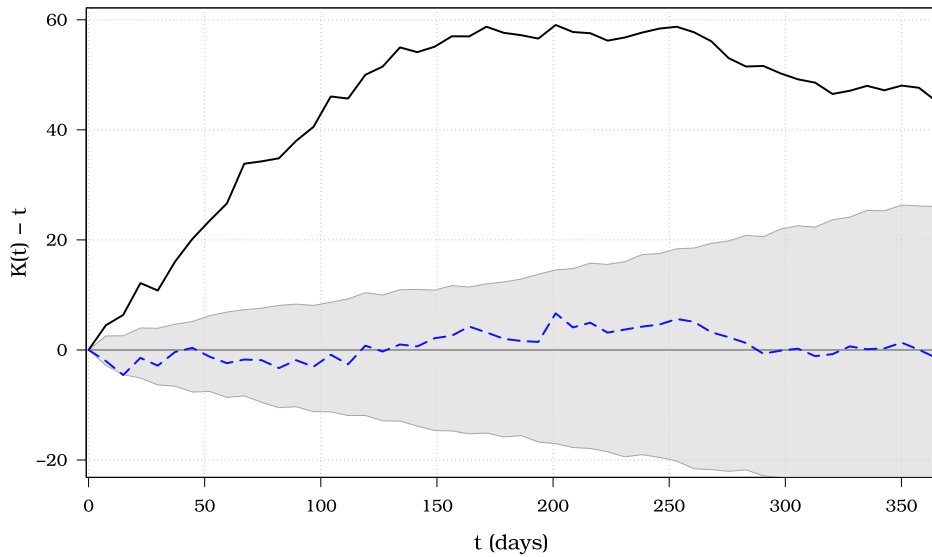


FIG. 2. Plot of $\hat{K}(t) - t$ for Indonesian terrorism (1994–2000). The solid line is from the baseline model (BL_5) with the gray band showing the 0.025 and 0.975 quantiles estimated from 1,000 simulations. The dashed line is from the self-exciting model (SE_1) presented in Section 4.

pler models in place of one more complicated model. This will be especially appealing in model fitting, as the likelihood may be separable, allowing for independent evaluation of each component.

For modeling the daily counts of a terrorist process, let Y_t be the number of events on day t and E_t be the indicator for an *event day* such that $E_t = 1$ if $Y_t \geq 1$ (i.e., there is at least one terrorist attack on day t) and $E_t = 0$ if $Y_t = 0$ (i.e., there are no attacks on day t). Also, let t_i denote the i th event day (not event, but day with at least one attack).

Letting $\mathcal{H}_t = \{Y_s : s \leq t\}$ be the internal history of the terrorist process, the hurdle component is modeled as a Bernoulli random variable with hurdle probability $p_t = \Pr(E_t = 1 | \mathcal{H}_{t-1})$ specified conditionally on the past history of the event process. A separate count model is constructed for $Y_t | E_t = 1$ with density $f_t(y) = \Pr(Y_t = y | E_t = 1, \mathcal{H}_{t-1})$, for $y \in \{1, 2, \dots\}$. Because the count component is constructed conditionally on there being at least one event, it will have no support on 0 [i.e., $f_t(0) = 0$]. Combining the hurdle and count components gives the full density

$$(1) \quad f_t^*(y) = \Pr(Y_t = y | \mathcal{H}_{t-1}) = \begin{cases} f_t(y)p_t, & y > 0, \\ 1 - p_t, & y = 0. \end{cases}$$

This has a similar form to a zero-inflated model [Lambert (1992)] but differs in that for the zero-inflated model $\Pr(Y_t = 0 | \mathcal{H}_{t-1}) = (1 - p_t)f_t(0)$.

Zero-inflated models usually arise when there is a random censoring process that prevents observations at certain times. Alternatively, instead of assuming that terrorist events occurred but were not observed, the hurdle approach explicitly models the probability that no events occurred.

The log-likelihood for a hurdle model can be decomposed as the sum of two terms $\log L = \log L_1 + \log L_2$, where for T observations

$$\begin{aligned}
 (2) \quad \log L_1 &= \sum_{t=1}^T E_t \log p_t + (1 - E_t) \log(1 - p_t), \\
 \log L_2 &= \sum_{t=1}^T E_t \cdot \log f_t(y_t) \\
 (3) \quad &= \sum_{t: E_t=1} \log f_t(y_t).
 \end{aligned}$$

The first term (in the form of a Bernoulli process) represents the hurdle component and the second term is the usual sum for a set of observations coming from the density f_t and represents the number of events per event day. The form of the likelihood is equivalent to a discrete time version of a marked point process [Daley and Vere-Jones (2003)], where the marks are the number of attacks on an event day. A particular benefit of this representation emerges when f_t and p_t share no common parameters, allowing their log-likelihoods to be handled separately for parameter estimation.

3.2. Self-exciting process. It has been suggested that some terrorist processes exhibit self-excitation or contagion behavior [Holden (1986), Dugan, LaFree and Piquero (2005), LaFree, Dugan and Korte (2009)]. A self-exciting point process [Hawkes (1971)] is one where the realization of events increases the short-term probability of observing future events, much in the same manner that one contagious individual can infect other individuals (while they are still infectious) or how major earthquakes lead to aftershocks. This type of model can be written in the form of a cluster process [Hawkes and Oakes (1974)] and used to explain the apparent clustering of terrorist activities. Specifically, we consider a generalized *shot noise process* [Rice (1977)], where the self-exciting component $S_t \geq 0$ will be a nonnegative function of the past history \mathcal{H}_{t-1} of the form

$$\begin{aligned}
 (4) \quad S_t &= \sum_{s < t} E_s \alpha_s g(t - s) \\
 &= \sum_{i: t_i < t} \alpha_i g(t - t_i).
 \end{aligned}$$

The magnitude parameter, α_i , determines the influence that the i th event day has on the self-exciting process. It may be a function of other associated

information about the event day (e.g., number of events, number killed, success indicator, group attribution). Based on the results of the exploratory data analysis, we restrict our attention to the case where $\alpha_i \geq 0$, although inhibition effects could be obtained if negative values were permitted.

The decay function $g(\cdot)$ specifies the shape of the excitation based on the time since the previous event days. To aid in parameter interpretation and estimation, we specify $g(\cdot)$ to be a proper probability mass function with strictly positive support such that $g(u) \geq 0$, $g(u) = 0$ for $u < 1$, and $\sum_{u=1}^{\infty} g(u) = 1$. This ensures that the influence of an event will eventually diminish and makes α_i solely responsible for the magnitude of the effects contributed by the i th event.

Any discrete distribution can be used for the decay function, provided its support is limited to the set of positive integers. Standard distributions, like Poisson or negative binomial, may require shifting or truncation to avoid having support on zero. For example, the (mean specified) shifted negative binomial is

$$(5) \quad g(u; \mu, r) = \frac{\Gamma(r + u - 1)}{\Gamma(r)(u - 1)!} \left(\frac{r}{\mu - 1 + r} \right)^r \left(\frac{\mu - 1}{\mu - 1 + r} \right)^{u-1},$$

which has a mean of $\mu > 1$, and size parameter $r > 0$. The corresponding shifted geometric distribution is given by $g(u; \mu, r = 1)$ and the shifted Poisson density can be defined as the limiting case when $r \rightarrow \infty$.

Figure 3 shows the geometric and negative binomial decay functions along with the shot noise processes corresponding to 8 event days. This illustrates how the form of the shot noise process from (4) is similar to kernel intensity estimation [Diggle (1985)], but with weights and one-sided (predictive) kernels.

3.3. Self-exciting Bernoulli hurdle process. The event days, E_t , are modeled as a self-exciting Bernoulli process where the probability of an event day is excited (increased) by the occurrence of previous events. Specifically, we consider the hurdle probability on day t to be a function of the process

$$(6) \quad X_t = B_t + S_t,$$

where B_t is a baseline process and S_t is the self-exciting component given by (4). The baseline can be a function of any exogenous variables that could have an effect on the process (e.g., social, political or economic conditions and events, counter-terrorism efforts, etc.), but it will not include any information coming from the internal history.

To ensure the hurdle probability $p_t \in [0, 1]$, a transformation $\eta(p_t) = X_t$ is used. If the baseline process is nonnegative (i.e., $B_t \geq 0$), then $X_t \geq 0$ and an appealing transformation is $\eta(p) = -\log(1 - p)$, corresponding to the hurdle

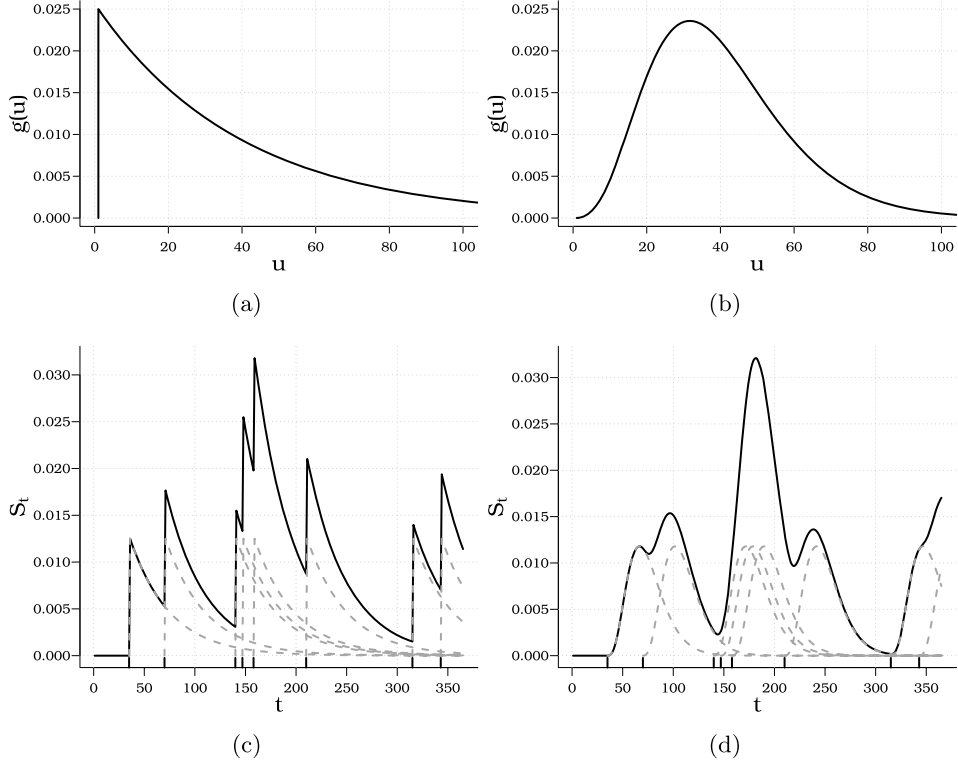


FIG. 3. (a) and (b) show the decay functions for the shifted geometric ($\mu = 40$) and shifted negative binomial ($\mu = 40, r = 5$), respectively. (c) and (d) show the corresponding shot noise processes (solid lines) and scaled component kernels (dotted lines) for 8 events with $\alpha_i = \alpha = 0.5$.

probability

$$(7) \quad p_t = 1 - e^{-X_t}.$$

While other transformations (e.g., logit) could be used, this one provides several benefits. It enjoys a likelihood function that is computationally convenient, substituting (7) into (2) obtains the Bernoulli log-likelihood

$$(8) \quad \log L_1 = \sum_{1 \leq t_i \leq T} \log(e^{X_{t_i}} - 1) - \sum_{t=1}^T X_t.$$

The shot noise component of the second sum ($\sum_{t=1}^T X_t = \sum_{t=1}^T B_t + \sum_{t=1}^T S_t$) can be simplified by recognizing that

$$\sum_{t=1}^T S_t = \sum_{i: t_i < T} \alpha_i G(T - t_i),$$

where G is the cumulative density function of g . This significantly reduces computation, as it is only calculated for the event days and not over all days in the observation window.

This transformation also allows for a straightforward interpretation of the parameters in X_t . Notice that p_t is also the probability that a Poisson random variable with rate X_t is greater than 0. When events are rare and X_t is small, E_t can be modeled approximately by a Poisson distribution. According to the superposition property of the Poisson, the event days arise from two sources, the baseline process with rate B_t and the shot noise process with rate S_t . The rate from the shot noise can be further decomposed into its elements $\alpha_i g(t - t_i)$ for $\{i : t_i < t\}$. Thus, under the Poisson approximation, event day i will generate an expected $\sum_{u=1}^{\infty} \alpha_i g(u) = \alpha_i$ number of additional event days and the decay function $g(\cdot)$ controls when the extra event days will occur. While these interpretations will not hold exactly under the Bernoulli model, they do provide an indication of how each component effects the process.

3.4. Survival functions. This formulation for the hurdle process also provides simple expressions for some properties of the time until the next event day. For a given time t_0 , let τ be the time of the next event day. The survival function $V_{t_0}(u) = \Pr(\tau - t_0 > u)$ gives the probability that the next event day will be more than u days away. For the self-exciting hurdle process, this becomes

$$\begin{aligned} V_{t_0}(u) &= \Pr(E_{t_0+1} = E_{t_0+2} = \dots = E_{t_0+u} = 0) \\ &= \prod_{\delta=1}^u (1 - p_{t_0+\delta}) \\ &= \exp\left\{-\sum_{\delta=1}^u X_{t_0+\delta}\right\} \end{aligned}$$

for $u \in \{1, 2, \dots\}$, where $X_{t_0+\delta}$ is calculated assuming no new events have occurred since t_0 . The survival function can be used to calculate the expected time until the next event day

$$\begin{aligned} \mathbb{E}_{t_0}[\tau] &= \mathbb{E}[\tau - t_0] = 1 + \sum_{u=1}^{\infty} V_{t_0}(u) \\ &= 1 + \sum_{u=1}^{\infty} \exp\left\{-\sum_{\delta=1}^u X_{t_0+\delta}\right\}. \end{aligned}$$

The survival function specification of this model can be useful for prediction, and making inference about future attacks given the current history.

4. Results. In order to build the models and evaluate their predictive performance, the data are partitioned into two time periods. The first time period from 1994 through 2000 is used to construct the models and estimate their parameters. The second period from 2001 through 2007 is used to assess the predictive performance of the models.

4.1. *Event day modeling.* Recall that the hurdle probability is constructed from the sum of a baseline and self-exciting processes (6). In order to capture potential seasonality and other large scale trends in the data, the baseline process B_t is defined as

$$(9) \quad \log(B_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + A_1 \sin(2\pi t/\omega) + A_2 \cos(2\pi t/\omega),$$

where the seasonal terms have a period of $\omega = 365.25$ days. This results in five baseline parameters $\theta_B = (\beta_0, \beta_1, \beta_2, A_1, A_2) \in \mathbb{R}^5$. The log transform ensures $B_t \geq 0$. We considered self-exciting processes of the form

$$(10) \quad S_t(\alpha, \mu, r) = \alpha \sum_{i: t_i < t} g(t - t_i; \mu, r),$$

where $g(u; \mu, r)$ is a shifted negative binomial decay function (5) and the magnitude α is a constant. This results in three parameters for the self-exciting component $\theta_S = (\alpha, \mu, r) \in \mathbb{R}_+^3$.

The combined model for $X_t = B_t + S_t$ results in up to 8 parameters to estimate. Estimation is carried out by maximizing the log-likelihood function given in (8). Since there is no explicit solution available, we used R's numerical optimization routine `nlminb` [R Development Core Team (2011)] to obtain estimates. For all models, the estimates converged within seconds from a variety of starting points.

Table 2 shows the evaluated models and their corresponding AIC scores. The lowest AIC belongs to the four component model with a constant baseline (SE₁), providing essentially a discrete time version of the Hawkes model [Hawkes (1971), Ozaki (1979)]. This AIC is much lower than the best baseline only model (BL₅) which includes the periodic terms. Figure 2 plots the weighted K -function for SE₁ and BL₅ showing the self-exciting model is much tighter around 0 indicating a better fit over the baseline only model [Diggle (1979)]. Figure 4 shows the estimated hurdle probability in the training period for both models.

The parameters of the self-exciting model (SE₁) suggest a baseline hurdle probability of about 0.012 when the shot noise term drops to zero. However, according to the Poisson approximation, every event day will generate an expected $\alpha = 0.89$ additional event days. The generated event day is expected to occur in $\mu = 37.54$ days, but the time until occurrence has a median of only 16 days.

TABLE 2

Parameter estimates for the event day models for the training period (1994–2000). The BL are the baseline only and the SE are the self-exciting models

Model	AIC	β_0	β_1	β_2	A_1	A_2	α	μ	r
BL ₁	1,187.77	-2.75							
BL ₂	1,149.54	-8.2	8.05						
BL ₃	1,151.25	-2.83	-8.01	11.91					
BL ₄	1,171.22	-2.82			-0.42	-0.31			
BL ₅	1,136.80	-8.02	7.69		-0.35	-0.32			
BL ₆	1,138.60	-3.53	-5.74	9.95	-0.35	-0.32			
SE ₁	1,075.21	-4.41					0.89	37.54	0.45
SE ₂	1,077.16	-5.17	1.23				0.87	36.55	0.45
SE ₃	1,078.71	15.31	-63.2	50.24			0.85	35.69	0.45
SE ₄	1,077.03	-4.34			-0.41	-0.40	0.86	38.55	0.44
SE ₅	1,078.76	-5.99	2.68		-0.38	-0.42	0.82	36.57	0.43
SE ₆	1,079.60	20.46	-79.33	62.91	-0.43	-0.45	0.80	37.28	0.43

4.2. *Count modeling.* It can be obtained from Table 1 that most event days (92.4%) are comprised of only 1 or 2 attacks. However, several days (1.9%) have more than 9 attacks (with one day having 36 recorded terrorist attacks). This suggests count distributions that have an initial rapid decay,

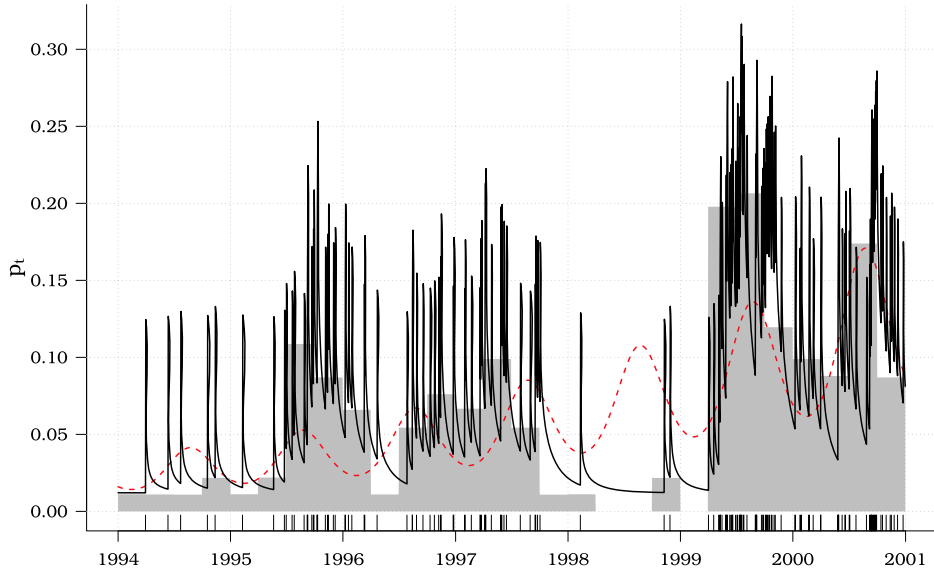


FIG. 4. Hurdle probabilities for the best fitting self-exciting model SE₁ (solid line), the best fitting baseline only model BL₅ (dotted line), and the observed piecewise constant probabilities (blocks). The rug signifies the event days.

but still possess a long tail to accommodate the extreme values. One such discrete distribution is the Riemann zeta or discrete Pareto distribution. This power law distribution has recently been employed to model the severity (e.g., number killed or injured) of terrorist attacks [Clauset, Young and Gleditsch (2007)].

Using it to model the event day counts, under an i.i.d assumption, gives the probability mass function

$$f_t(y; s) = \frac{y^{-s}}{\zeta(s)}, \quad y \in \{1, 2, \dots\},$$

where $\zeta(s) = \sum_{x=1}^{\infty} x^{-s}$ is the Riemann zeta function and the parameter $s \in (1, \infty)$. This one parameter model is easy to estimate numerically via maximum likelihood [Goldstein, Morris and Yen (2004), Seal (1952)] by finding the s that satisfies the equation $\frac{1}{n} \sum_{i=1}^n \log(y_i) = -\frac{\zeta'(s)}{\zeta(s)}$. The training period data gives rise to an estimated parameter of $\hat{s} = 2.86$. The bootstrap Kolmogorov–Smirnov test of Clauset, Shalizi and Newman (2009) gives a p -value of 0.69, suggesting the zeta distribution provides a suitable fit.

While the constant parameter zeta distribution appears to provide a good fit, there may be better models, especially if the distribution is allowed to vary over time. In particular, as for the hurdle component, distributions that vary over time in response to a shot noise process are considered. An additional driving process (with shot noise) was implemented for the count model, namely,

$$X_t^c(\beta^c, \alpha^c, \mu^c) = \beta^c + \alpha^c \sum_{i: t_i < t} Y_i \cdot g^c(t - t_i; \mu^c),$$

where β^c is a constant baseline, $\alpha_i^c = \alpha^c \cdot Y_i$ varies in response to the number of attacks on the i th event day, and the decay function $g^c(u; \mu^c)$ is the one parameter shifted geometric function with mean μ^c . This results in three parameters for the count component $\theta_C = (\beta^c, \alpha^c, \mu^c)$. Note that this is a separate process than the one specified for the hurdle component given by (9) and (10). The additional flexibility afforded by the hurdle specification allows such additional complexity to be introduced for the count component of the model.

There are two primary hypotheses about how the count distribution behaves in response to the event history. First, counts could respond in a self-exciting manner, where recent event and multi-event days increase the likelihood of further multi-event days. Alternatively, due to the depletion of resources or increased anti-terrorism measures after large attacks, counts may respond in a more self-inhibiting manner, where recent multi-event days lower the chances of multi-events day in the near future.

To evaluate the self-exciting hypothesis, let $s_t = (1 - e^{-X_t^c})^{-1}$ since smaller values of s lead to more mass on the extreme values. Alternatively, $s_t = e^{X_t^c}$

TABLE 3

Parameter estimates for the count models for the training period (1994–2000). C_z , C_{se} and C_{si} are the constant parameter, self-exciting and self-inhibiting forms, respectively, of the Riemann zeta distribution

Model	AIC	s	β	α	μ
C_z	241.00	2.86			
C_{se}	240.48	$(1 - e^{-X_t})^{-1}$	0.375	0.20	2.13
C_{si}	244.72	e^{X_t}	1.00	0.34	55.18

for the self-inhibiting model. Table 3 shows the estimated parameters and AIC scores for the fitted count models. The lower AIC suggests that the self-exciting count model has more validity than the self-inhibiting model for the training data. Figure 5 shows the estimated probability of counts under the constant parameter (C_z) and self-exciting (C_{se}) models. While the self-exciting model puts more mass on multi-event days when X_t^c is large, the decay is rapid and the self-exciting behavior is short lived—only for a few days, before the shot noise term approaches 0 and its effects become negligible.

4.3. *Testing model predictions.* Based on the training period, models are selected and fit for both the hurdle and count components. To evaluate how

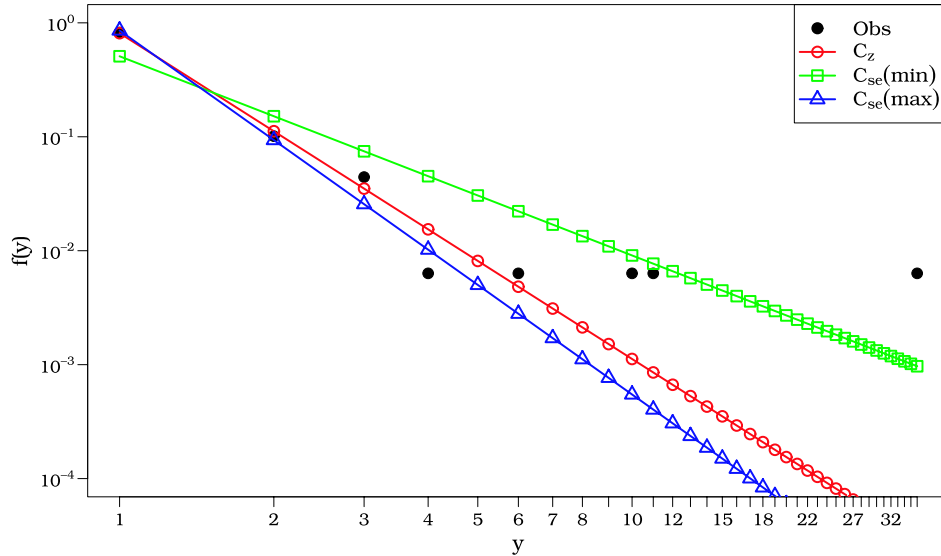


FIG. 5. Estimated count probabilities for the constant parameter model (C_z with $s = 2.86$) and for the self-exciting count model (C_{se}) at the baseline state ($s = 3.19$) and most excited state ($s = 1.75$).

well this can predict future terrorist activity, the models are applied to the terrorist incidents in the testing period (2001–2007) using the following procedure:

- (1) At day t predict the probability of an event day and conditional number of attacks on day $t + 1$, based on the current history \mathcal{H}_t .
- (2) Observe Y_{t+1} and re-estimate the model parameters based on the updated history \mathcal{H}_{t+1} .
- (3) Repeat steps (1) and (2) through the testing period to construct a final model based on all data between 1994–2007.

This update and forecast scheme results in a set of daily predictions $\hat{p}_t = \Pr(E_t = 1 | \mathcal{H}_{t-1})$ and $\hat{f}_t(y) = \Pr(Y_t = y | E_t = 1, \mathcal{H}_{t-1})$ conditioned on the past history of the process.

The predictive capability of the models are evaluated by comparing the log probability gain, or predictive log-likelihood ratio, of each model [Daley and Vere-Jones (2004)]. This score is a relative measure of the improvement in predicative capability for a model compared to a reference or null model. The interpretation is that values greater than 0 show a relative improvement in predictions over the reference model, with the highest values indicating the preferred model.

For the hurdle component, the log probability gain is

$$G = \sum_{t \in \mathcal{T}_2} E_t \log \frac{\hat{p}_t}{\hat{\pi}_t} + (1 - E_t) \log \frac{1 - \hat{p}_t}{1 - \hat{\pi}_t},$$

where \hat{p}_t comes from the model of interest and $\hat{\pi}_t$ from the constant baseline only model (BL₁), and $\mathcal{T}_2 = [01 \text{ Jan } 2001, 31 \text{ Dec } 2007]$ covers the testing period. Table 4 compares the log probability gains G and the AIC scores from the training period, and shows that the results from the test period were similar to those in the training period, supporting the use of AIC for model selection.

TABLE 4

The hurdle component log probability score, G , using the constant baseline only model (BL₁) as the reference model. The AIC scores are from the training period (see Table 2)

Model	AIC	G	Model	AIC	G
BL ₁	1,187.77	0.00	SE ₁	1,075.21	26.96
BL ₂	1,149.54	-19.10	SE ₂	1,077.16	25.47
BL ₃	1,151.25	20.69	SE ₃	1,078.71	24.96
BL ₄	1,171.22	-1.39	SE ₄	1,077.03	26.22
BL ₅	1,136.80	-18.21	SE ₅	1,078.76	24.56
BL ₆	1,138.60	21.63	SE ₆	1,079.60	24.57

TABLE 5
*The count component log probability score, G^c ,
 using the constant parameter zeta model (C_z)
 as the reference model*

	C_s	C_{se}	C_{si}
G^c	0.00	8.07	-0.27

The constant baseline self-exciting model (SE_1) has the largest log probability gain and all self-exciting models outperform all of the baseline only models. The best baseline only model for the testing period was the full model with seasonality and a quadratic trend (BL_6). This is in contrast to the training period results where the best baseline only model had a linear trend (BL_5). The linear trend model failed to perform as well during the testing period, as the general attack rate peaked around the transition point and started decreasing during the remainder of the testing period, creating a significant quadratic trend over the combined observation period. Alternatively, the self-exciting models can adapt rapidly to sudden changes in attack rate, yielding better overall predictive ability.

The count component uses the log probability gain score

$$G^c = \sum_{i: t_i \in \mathcal{T}_2} \log \frac{\hat{f}(y_i)}{\hat{h}(y_i)},$$

where \hat{f} is the model of interest and \hat{h} is the reference model. Table 5 shows the values of G^c using the constant parameter zeta model (C_z) as the reference model. These results show the self-exciting zeta model (C_{se}) as the preferred model for predicting the counts. As with the hurdle component, the predictive performance of the count models followed the AIC scores from the training period. However, the model with the self-exciting component does predict substantially better than the constant component model—more so than would be expected from the AIC scores in the training period (see Table 3). This may be due to changes in the attack behavior during the testing period. Figure 1 shows that while there were a few large counts during 2001, there were very few multi-attack days after. This change in the pattern of multiple attacks accounts for the improved predictive capabilities of the self-exciting model over the constant parameter model. The self-exciting count model is able to quickly adapt to changing conditions and consequently makes better predictions.

In summary, the daily counts of Indonesian terrorist attacks were modeled with a self-exciting hurdle model. The separation of the model into two components allowed both the attack rate and attack characteristics to

be more faithfully represented. It was found that including a self-exciting process benefited both the hurdle and count components by allowing the models to quickly adapt to changing attack behavior. It is this property as a predictive tool that makes the self-exciting hurdle model especially useful in an applied setting where prediction is of primary interest.

5. Conclusion. In practice, policy makers are faced with limited resources, financially, materially and in personnel. The allocation of these resources depends on an understanding of the future risk of terrorist activity. As an example, in the wake of the attacks of September 11, 2001, policy makers placed additional security in the form of National Guard troops in all US commercial airports. This was done in hopes of re-establishing public trust in the safety of air travel while the intelligence community was assessing the risk of subsequent attacks [Price and Forrest (2009)]. The obvious question in that case was when to shift resources from increased airport security to investing in long-term strategies to improve intelligence capabilities. As this case illustrates, only by clearly understanding the risk of terrorist activity can policy makers make informed decisions about allocating resources.

This paper presents a two-component self-exciting model for the analysis and prediction of future terrorist activity. It was found that the best model used a constant baseline with a self-exciting term for the hurdle component and a Riemann zeta distribution with shot noise driven parameter for the count component. The improvement offered by the self-exciting term provides support for the contagion theory and suggests a significant short term increase in terrorism risk after an attack. The use of a power law distribution for the number of same day attacks corresponds with the power law behavior of the number killed in terrorist attacks [Clauset, Young and Gleditsch (2007)]. This is to be expected, as multiple coordinated attacks would tend to indicate better planned, and hence more severe, terrorist activities.

The self-exciting hurdle model adheres to the theoretical concept for a contagion effect to terrorism as manifested in the clustering of data while providing good fit and predictive capabilities without relying on exogenous variables. The model provides a simple structure and interpretation of the parameters useful for understanding the dynamic nature of the terrorist activity. This provides an appropriate starting point for exploring additional covariate effects, including analysis concerning the effectiveness of counter-terrorism activities, geography, political or economical factors. As the attack clustering may be partially attributable to a stochastic baseline driven by such exogenous processes [Holden (1986)], further analysis could include covariates in the baseline model (9) to test the impact on the self-exciting component.

By focusing on the timing of terrorist attacks on the islands of Indonesia, other aspects of the attacks, such as location, attack type and group responsible, were not considered. Including such additional information could lead

to more detailed models that better explain the nuances of the terrorist activity. For example, the self-exciting component could include a spatial proximity term using the models of Mohler et al. (2011) or multiple self-exciting terms could be included that represent how attacks from one terrorist group influence the subsequent attacks of other groups.

The utility of the models presented here and their ease of implementation and interpretation make them a potentially useful tool in security related fields. The results show that the risk of terrorist activity can vary greatly over short periods of time, thus policy responses in terms of resource allocation, security and counter-terrorism responses should reflect this as well as addressing the more long-term trends in risk. For example, understanding the short-term variations in risk would allow a more effective deployment and assessment of additional airport security measures in the wake of a terrorist attack, while a grasp of the longer term trends could help guide the development and assessment of more strategic counter-terrorism resources, such as increasing the number of foreign language experts available for translation or developing effective de-radicalization programs to prevent the growth of terrorist groups.

REFERENCES

- BADDELEY, A. J., MØLLER, J. and WAAGEPETERSEN, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **54** 329–350. [MR1804002](#)
- BARROS, C. P. (2003). An intervention analysis of terrorism: The Spanish ETA case. *Defence and Peace Economics* **6** 401–412.
- BELASCO, A. (2010). The cost of Iraq, Afghanistan, and other Global War on Terror Operations Since 9/11. Technical Report RL33110, Congressional Research Services. Available at <http://www.fas.org/sgp/crs/natsec/RL33110.pdf>.
- CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51** 661–703. [MR2563829](#)
- CLAUSET, A., YOUNG, M. and GLEDITSCH, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution* **51** 58–87.
- DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. I*, 2nd ed. Springer, New York.
- DALEY, D. J. and VERE-JONES, D. (2004). Scoring probability forecasts for point processes: The entropy score and information gain. *J. Appl. Probab.* **41A** 297–312. [MR2057581](#)
- DIGGLE, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35** 87–101.
- DIGGLE, P. (1985). A kernel method for smoothing point process data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **34** 138–147.
- DIXON, P. M. (2002). Ripley’s K function. In *Encyclopedia of Econometrics* (A. H. EL-SHAARAWI and W. W. PIEGORSCH, eds.) **2** 1796–1803. Wiley, Chichester.
- DUGAN, L., LAFREE, G. and PIQUERO, A. (2005). Testing a rational choice model of airline hijackings. *Criminology* **43** 1031–1066.
- ENDERS, W. and SANDLER, T. (1993). The effectiveness of antiterrorism policies: A vector-autoregression-intervention analysis. *The American Political Science Review* **4** 829–844.

- ENDERS, W. and SANDLER, T. (2000). Is transnational terrorism becoming more threatening? *Journal of Conflict Resolution* **44** 307–332.
- ENDERS, W. and SANDLER, T. (2002). Patterns of transnational terrorism, 1970–1999: Alternative time-series estimates. *International Studies Quarterly* **2** 145–165.
- ENDERS, W. and SANDLER, T. (2006). *The Political Economy of Terrorism*. Cambridge Univ. Press, New York.
- GOLDSTEIN, M. L., MORRIS, S. A. and YEN, G. G. (2004). Problems with fitting to the power-law distribution. *Eur. Phys. J. B* **41** 255–258.
- HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. [MR0278410](#)
- HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. [MR0378093](#)
- HEARD, N. A., WESTON, D. J., PLATANIOTI, K. and HAND, D. J. (2010). Bayesian anomaly detection methods for social networks. *Ann. Appl. Stat.* **4** 645–662. [MR2758643](#)
- HEILBRON, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biom. J.* **36** 531–547.
- HOLDEN, R. T. (1986). The contagiousness of aircraft hijacking. *The American Journal of Sociology* **91** 874–904.
- HOLDEN, R. T. (1987). Time series analysis of a contagious process. *J. Amer. Statist. Assoc.* **82** 1019–1026. [MR0922168](#)
- JONES, A. M. (1994). Health, addiction, social interaction and the decision to quit smoking. *Journal of Health Economics* **13** 93–110.
- LAFREE, G. and DUGAN, L. (2007). Introducing the global terrorism database. *Terrorism and Political Violence* **19** 181–204.
- LAFREE, G., DUGAN, L. and KORTE, R. (2009). The impact of British counter terrorist strategies on political violence in Northern Ireland: Comparing deterrence and backlash models. *Criminology* **47** 17–45.
- LAFREE, G., MORRIS, N. A. and DUGAN, L. (2010). Cross-national patterns of terrorism: Comparing trajectories for total, attributed and fatal attacks, 1970–2006. *British Journal of Criminology* **50** 622–649.
- LAMBERT, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34** 1–14.
- LI, R. P. and THOMPSON, W. R. (1975). The “Coups contagion” hypothesis. *The Journal of Conflict Resolution* **19** 63–88.
- LUM, C., KENNEDY, L. W. and SHERLEY, A. J. (2006). Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. *Journal of Experimental Criminology* **2** 489–516.
- MARSCHALL, M. J., RUHIL, V. S. ANIRUDH and SHAH, P. R. (2010). The new radical calculus: Electoral institutions and black representation in local legislatures. *American Journal of Political Science* **54** 107–124.
- MIDLARSKY, M. I. (1978). Analyzing diffusion and contagion effects: The urban disorders of the 1960s. *The American Political Science Review* **72** 996–1008.
- MIDLARSKY, M. I., CRENSHAW, M. and YOSHIDA, F. (1980). Why violence spreads: The contagion of international terrorism. *International Studies Quarterly* **24** 341–365.
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. [MR2816705](#)
- MULLAHY, J. (1986). Specification and testing of some modified count data models. *J. Econometrics* **33** 341–365. [MR0867980](#)

- MYERS, D. J. (2000). The diffusion of collective violence: Infectiousness, susceptibility, and mass media networks. *American Journal of Sociology* **106** 178–208.
- NAGIN, D. (2005). *Group-Based Modeling of Development*. Harvard Univ. Press, Cambridge, MA.
- OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. [MR0541960](#)
- PERL, R. (2007). Combating terrorism: The challenge of measuring effectiveness. Technical Report RL33160, Congressional Research Services. Available at <http://fpc.state.gov/documents/organization/57513.pdf>.
- PRICE, J. C. and FORREST, J. S. (2009). *Practical Aviation Security: Predicting and Preventing Future Threats*. Butterworth-Heinemann, Oxford, UK.
- R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- RIE, J. (1977). On generalized shot noise. *Adv. in Appl. Probab.* **9** 553–565. [MR0483114](#)
- SEAL, H. L. (1952). The maximum likelihood fitting of the discrete Pareto law. *Journal of the Institute of Actuaries* **78** 115–121.
- TOWNSLEY, M., JOHNSON, S. D. and RATCLIFFE, J. H. (2008). The time dynamics of insurgent activity in Iraq. *Security Journal* **21** 139–146.
- VEEN, A. and SCHOENBERG, F. P. (2006). Assessing spatial point process models using weighted K -functions: Analysis of California earthquakes. In *Case Studies in Spatial Point Process Modeling* (A. BADDELEY, P. GREGORI, J. MATEU, R. STOICA and D. STOYAN, eds.). *Lecture Notes in Statist.* **185** 293–306. Springer, New York. [MR2232135](#)
- WELSH, A., CUNNINGHAM, R. B., DONNELLY, C. F. and LINDENMAYER, D. B. (1996). Methods for analysing data with extra zero: ZIP regression models with application for surveys of rare species. *Ecological Modelling* **88** 297–308.

GEOEYE ANALYTICS
7921 JONES BRANCH DRIVE, SUITE 600
MCLEAN, VIRGINIA
USA
E-MAIL: porter.mike@geoeye.com

INSTITUTE FOR SOCIAL SCIENCE RESEARCH
ARC CENTRE OF EXCELLENCE IN POLICING
AND SECURITY
UNIVERSITY OF QUEENSLAND
ST. LUCIA, QUEENSLAND 4072
AUSTRALIA
E-MAIL: gentry.white@uq.edu.au

SELF-EXCITING HURDLE MODELS FOR TERRORIST ACTIVITY

BY MICHAEL D. PORTER AND GENTRY WHITE

GeoEye Analytics
The University of Queensland

A predictive model of terrorist activity is developed by examining the daily number of terrorist attacks in Indonesia from 1994 through 2007. The dynamic model employs a shot noise process to explain the self-exciting nature of the terrorist activities. This estimates the probability of future attacks as a function of the times since the past attacks. In addition, the excess of non-attack days coupled with the presence of multiple coordinated attacks on the same day compelled the use of hurdle models to jointly model the probability of an attack day and corresponding number of attacks. A power law distribution with a shot noise driven parameter best modeled the number of attacks on an attack day. Interpretation of the model parameters is discussed and predictive performance of the models is evaluated.

1. Introduction. Since the attacks of September 11, 2001 there has been a substantial increase in the amount of resources dedicated to combating terrorism. As a single example, the Congressional Research Services estimates that as of September 2010 \$1.2 trillion USD has been spent on the Global War on Terror, including \$28.5 billion USD on enhanced security (Belasco, 2010). Despite this increase in spending there are no accurate measures of the effectiveness of these counter-terrorism efforts (Lum C. and Sherley, 2006; Perl, 2007). According to the Congressional Research Service, one of the keys to combating terrorism effectively is to understand, and have correctly specified models for, terrorist activity

“Better understanding of the dynamics of terrorism allows for a more complete picture of the complexities involved in measuring success or failure and can assist the 110th Congress as it coordinates, funds, and oversees anti-terrorism policy and programs” Perl (2007).

A dynamic model for terrorism should describe terrorist activity in a way that is consistent with both the theoretical framework for terrorism and the observed data. Defining accurate models of terrorist activity is important not

AMS 2000 subject classifications: Primary, ; secondary

Keywords and phrases: self-exciting, hurdle model, shot noise, terrorism, Indonesia, Hawkes process, Riemann zeta, point process, probability gain

1 just from the standpoint of assessing the effectiveness of counter-terrorism 1
2 efforts but also for use as a tool to predict the future risk of terrorist attack. 2

3 A majority of publications in the literature use one of three basic model- 3
4 ing approaches to terrorism. The work of [Enders and Sandler \(1993, 2000,](#) 4
5 [2002, 2006\)](#), and [Barros \(2003\)](#) uses time series analysis techniques includ- 5
6 ing intervention analysis and vector autoregressive (VAR) models. Other 6
7 more recent papers have applied group based trajectory analysis ([Nagin,](#) 7
8 [2005](#)) to analyze terrorist data. Examples of this include [LaFree, Morris and](#) 8
9 [Dugan \(2010\)](#) who incorporate a zero-inflated Poisson distribution to ac- 9
10 count for an excess of zero counts. Other approaches use Cox proportional 10
11 hazards models to directly model the time between attacks ([Dugan, LaFree](#) 11
12 [and Piquero, 2005; LaFree, Dugan and Korte, 2009](#)). These last two papers 12
13 attempt to account for the recent event history as influencing current events 13
14 by constructing an ad hoc term that describes the recent density of attacks. 14
15 While this is a step toward incorporating the dynamics related to the process 15
16 history, it requires prespecification of the form of dependence. 16

17 The desire to include terms to account for recent event history is rooted 17
18 in the theoretical understanding of terrorism and other politically moti- 18
19 vated violence. The clustering hypothesis, concurrent with contagion theory 19
20 ([Midlarsky, 1978](#)), explains politically motivated violent activity as a series 20
21 of non-independent events, where each event influences the probability of 21
22 subsequent events. This hypothesis is well accepted in criminological and 22
23 sociological approaches to terrorism. Clustering effects and contagious be- 23
24 havior have been demonstrated in military coups ([Li and Thompson, 1975](#)), 24
25 international terrorism ([Midlarsky, Crenshaw and Yoshida, 1980](#)), airline 25
26 hijackings ([Holden, 1986](#)), race riots ([Myers, 2000](#)), and insurgent activity 26
27 ([Townesley, Johnson and Ratcliffe, 2008](#)). 27

28 In examining terrorism data in general there are two major issues to 28
29 address. The first is that as terrorist attacks are usually rare, the daily num- 29
30 ber of terrorist attacks are often zero, but due to large coordinated attacks 30
31 there are some extreme values. This combination of a large number of zeros 31
32 and extreme values is poorly modeled by standard probability distributions. 32
33 Second, the timing of terrorist incidents appear to be clustered, rendering 33
34 standard models that assume independence unsuitable. 34

35 The first of these issues is addressed using a hurdle model ([Mullahy, 1986](#)), 35
36 also known as the two-part model ([Heilbron, 1994](#)). The hurdle model is a 36
37 two component model that allows a separate specifications of the probability 37
38 of a zero count and the probability of a non-zero count. This allows the hurdle 38
39 model to accommodate a large number of zeros in addition to some extreme 39
40 counts. The hurdle model is used in a variety of applications, including in 40

41

ecology for modeling counts of rare species (Welsh et al., 1996), in public health for modeling smoking behavior (Jones, 1994), in political science for modeling proportional representation in minority electorates (Marschall, Ruhil and Shah, 2010), and network change detection (Heard et al., 2010). For terrorism modeling, the hurdle model is preferred to the zero-inflated model (Lambert, 1992) which assumes that the extra zeros observed are due to censoring. The hurdle model assumes that the extra zeros are due to a separate process (the “hurdle”), which must be overcome before the number of corresponding incidents are determined. This is more reasonable for terrorism data as it can often be assumed that sparsity of attacks is because they are indeed rare, not because they are unobserved.

The second issue regarding the clustering behavior of terrorist activity is addressed by incorporating a self-exciting component (Hawkes, 1971). The self-exciting component specifies that the probability of a event is a function of the time (and possibly other aspects) of all previous events, such that the effect on the probability decreases over time. This can help account for the clustering and dynamic nature of terrorism. For example, Holden (1986, 1987) developed a self-exciting Poisson model that better represented the dynamics of airline hijacking and Mohler et al. (2011) a self-exciting space-time model for residential burglaries.

This paper combines these two concepts to examine the use of self-exciting hurdle models for terrorist activity in Indonesia and Timor-Leste between 1994 and 2007. The hurdle component is modeled as a self-exciting Bernoulli process where the form of self-excitation is dictated by the data. The non-zero counts are modeled by an extreme value distribution with parameters that are also a function of the event history. The corresponding self-exciting hurdle model is capable of making predictions and providing information about the risk of terrorist activity without any additional covariate information other than the event history. This capability is important as covariate information for terrorism data is often either missing or unreliable.

2. Data and Exploratory Analysis. The Global Terrorism Database (LaFree and Dugan, 2007) is an open-source publicly available database of over 87,000 terrorist events around the world from 1970 through 2008. The database is continually updated with new information and is believed to be the most comprehensive database of its kind. The data used here are a subset of the GTD consisting of daily counts of terrorist attacks in Indonesia¹ from January 1, 1994 through December 31, 2007.

¹This includes data from Timor-Leste (East Timor) which became an sovereign state in 2002.

2.1. *Exploratory Data Analysis.* The data from a training period of 1994 through 2000 was examined to inform model construction. Of the 2557 days considered in this analysis, there were 250 terrorist attacks on 158 unique event days. Figure 1, displaying the daily and semi-yearly counts of attacks, illustrates the nature of the terrorist activity. In particular, the attacks appear clustered with long stretches between attacks followed by a period of increased activity. In addition, while most days have no attacks (93.8%), some days have multiple attacks (possibly due to planned coordinated attacks).

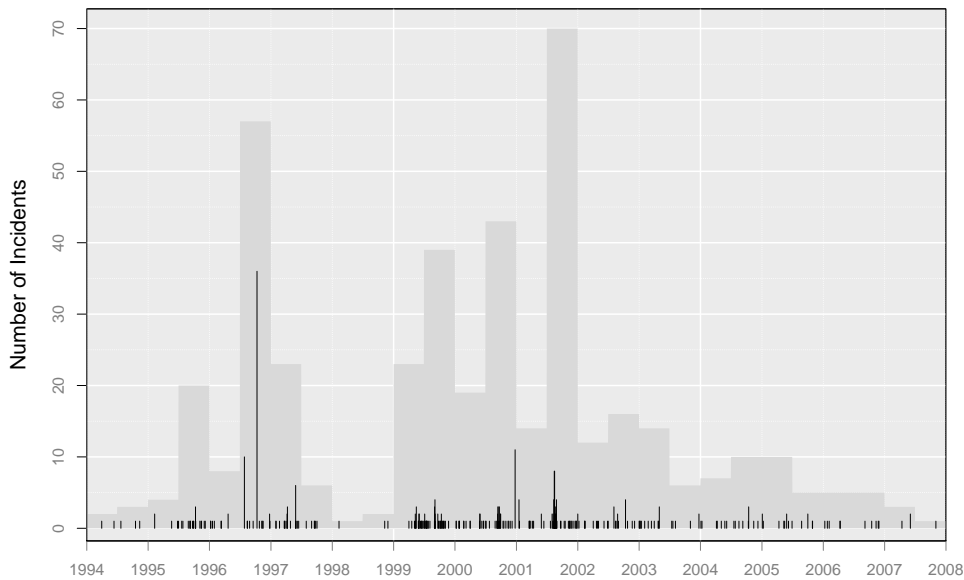


Fig 1: Daily number of terrorist events in Indonesia. The histogram shows the cumulative number of attacks in each six month period.

Table 1 shows the distribution of the observed attacks per day compared to the expected values under a Poisson and negative binomial probability distribution fitted with maximum likelihood. While the Poisson cannot capture any of the tail behavior, the negative binomial adapts better to the tail distribution but consequently underestimates the number days with only 1 attack and overestimates the number of days with 2 and 3 attacks. A chi-square goodness-of-fit test yields a p -value of 0.00016 suggesting that the negative binomial is not a sufficient model. In addition, the tail is comprised of several extreme values (36,11,10, and 6 attacks on a single day) that exacerbate the general lack of fit. The large number of zeros coupled with the presence of several extreme values necessitate more complex models.

TABLE 1
Distribution of observed and expected number of Indonesian terrorist attacks per day (1994-2000).

# Attacks	# Days	Poisson	Neg.Bin
0	2399	2319	2401
1	130	227	103
2	16	11	31
3	7	0	12
4	1	0	5
>4	4	0	5
AIC	-	1988.0	1481.8

Figure 1 also reveals the possibility that the attack days are clustered. This is examined with Ripley's K -function (Dixon, 2002). The K -function is a second order function that can reveal if clustering (or inhibition) is present in a point pattern. It is defined loosely as

$$K(t) \propto \mathbb{E}[\# \text{ of future events } \leq t \text{ from a randomly chosen event}]$$

Because this measure will be effected by an inhomogeneous attack day probability, we estimate it (Baddeley, Møller and Waagepetersen, 2000; Veen and Schoenberg, 2006)

$$\hat{K}(t) = T^{-1} \sum_{i=1}^N \frac{w_i}{\hat{p}_i} \sum_{j>i} \frac{\mathbb{1}(t_j - t_i \leq t)}{\hat{p}_j}$$

where T is the total observational time, \hat{p}_i is the estimated probability of at least one attack at day t_i , and w_i is a one-sided edge correction. The estimated probability is given by the model in (9) that includes trend and seasonality components (i.e model BL₅ in Table 2).

Figure 2 plots $\hat{K}(t) - t$ which has expected value 0 if the estimated probabilities are correct and there is no unaccounted for clustering. The observed values fall well above the 95% pointwise confidence intervals (obtained from a parametric bootstrap, 1000 simulations) showing that the data display clustering behavior not accounted for by the baseline model.

These characteristics of the terrorism data require flexible models that can handle the temporal attack patterns, clustering, an excess of non-attack days, and the possibility of multiple coordinated attacks on the same day. Self-exciting hurdle models are developed in the next section to represent the complex nature of such terrorist activity.

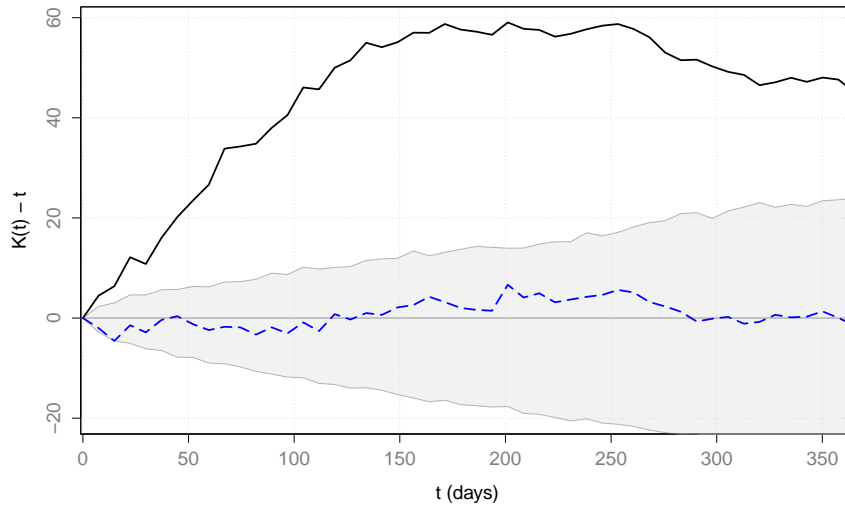


Fig 2: Plot of $\hat{K}(t) - t$ for Indonesian terrorism (1994-2000). The solid line is from the baseline model (BL_5) with the gray band showing the 0.025 and 0.975 quantiles estimated from 1000 simulations. The dashed line is from the self-exciting model (SE_1) presented in Section 4.

3. Hurdle Models and the Self-Exciting Process.

3.1. *Hurdle Models.* The hurdle models of Mullahy (1986) refer to a class of two component models for discrete count data. These models assume that two different processes drive the zero and non-zero counts respectively. The *hurdle* component of the model corresponds to the probability that the count is non-zero (i.e. a terrorist action will occur), while the *count* component corresponds to the distribution of positive counts (i.e. the number of corresponding terrorist events). When the two components are combined, the hurdle model produces a probability mass function on the non-negative integers. Additionally, the hurdle approach facilitates the use of two simpler models in place of one more complicated model. This will be especially appealing in model fitting as the likelihood may be separable, allowing for independent evaluation of each component.

For modeling the daily counts of a terrorist process, let Y_t be the number of events on day t and E_t be the indicator for an *event day* such that $E_t = 1$ if $Y_t \geq 1$ (i.e. there is at least one terrorist attack on day t) and $E_t = 0$ if $Y_t = 0$ (i.e. there are no attacks on day t). Also, let t_i denote the i^{th} event day (not event, but day with at least one attack).

Letting $\mathcal{H}_t = \{Y_s : s \leq t\}$ be the internal history of the terrorist process,

the hurdle component is modeled as a Bernoulli random variable with hurdle probability $p_t = \Pr(E_t = 1 | \mathcal{H}_{t-1})$ specified conditionally on the past history of the event process. A separate count model is constructed for $Y_t | E_t = 1$ with density $f_t(y) = \Pr(Y_t = y | E_t = 1, \mathcal{H}_{t-1})$, for $y \in \{1, 2, \dots\}$. Because the count component is constructed conditionally on there being at least one event, it will have no support on 0 (i.e. $f_t(0) = 0$). Combining the hurdle and count components gives the full density

$$(1) \quad f_t^*(y) = \Pr(Y_t = y | \mathcal{H}_{t-1}) = \begin{cases} f_t(y)p_t & y > 0 \\ 1 - p_t & y = 0 \end{cases}$$

This has a similar form to a zero-inflated model (Lambert, 1992) but differs in that for the zero-inflated model $\Pr(Y_t = 0 | \mathcal{H}_{t-1}) = (1 - p_t)f_t(0)$. Zero-inflated models usually arise when there is a random censoring process that prevents observations at certain times. Alternatively, instead of assuming that terrorist events occurred but were not observed, the hurdle approach explicitly models the probability that no events occurred.

The log-likelihood for a hurdle model can be decomposed as the sum of two terms $\log L = \log L_1 + \log L_2$, where for T observations

$$(2) \quad \log L_1 = \sum_{t=1}^T E_t \log p_t + (1 - E_t) \log(1 - p_t)$$

$$(3) \quad \begin{aligned} \log L_2 &= \sum_{t=1}^T E_t \cdot \log f_t(y_t) \\ &= \sum_{t: E_t=1} \log f_t(y_t) \end{aligned}$$

The first term (in the form of a Bernoulli process) represents the hurdle component and the second term is the usual sum for a set of observations coming from the density f_t and represents the number of events per event day. The form of the likelihood is equivalent to a discrete time version of a marked point process (Daley and Vere-Jones, 2003) where the marks are the number of attacks on an event day. A particular benefit of this representation emerges when f_t and p_t share no common parameters allowing their log-likelihoods to be handled separately for parameter estimation.

3.2. Self-exciting Process. It has been suggested (Holden (1986); Dugan, LaFree and Piquero (2005); LaFree, Dugan and Korte (2009)), that some terrorist processes exhibit self-excitation or contagion behavior. A self-exciting point process (Hawkes, 1971) is one where the realization of events increase

the short-term probability of observing future events, much in the same manner that one contagious individual can infect other individuals (while they are still infectious) or how major earthquakes lead to aftershocks. This type of model can be written in the form of a cluster process (Hawkes and Oakes, 1974) and used to explain the apparent clustering of terrorist activities. Specifically we consider a generalized *shot noise process* (Rice, 1977), where the self-exciting component $S_t \geq 0$ will be a non-negative function of the past history \mathcal{H}_{t-1} of the form:

$$(4) \quad \begin{aligned} S_t &= \sum_{s < t} E_s \alpha_s g(t - s) \\ &= \sum_{i: t_i < t} \alpha_i g(t - t_i) \end{aligned}$$

The magnitude parameter, α_i determines the influence that the i^{th} event day has on the self-exciting process. It may be a function of other associated information about the event day (e.g. number of events, number killed, success indicator, group attribution). Based on the results of the exploratory data analysis we restrict our attention to the case where $\alpha_i \geq 0$, although inhibition effects could be obtained if negative values were permitted.

The decay function $g(\cdot)$ specifies the shape of the excitation based on the time since the previous event days. To aid in parameter interpretation and estimation, we specify $g(\cdot)$ to be a proper probability mass function with strictly positive support such that $g(u) \geq 0$, $g(u) = 0$ for $u < 1$, and $\sum_{u=1}^{\infty} g(u) = 1$. This ensures that the influence of an event will eventually diminish and makes α_i solely responsible for the magnitude of the effects contributed by the i^{th} event.

Any discrete distribution can be used for the decay function provided its support is limited to the set of positive integers. Standard distributions, like Poisson or negative binomial, may require shifting or truncation to avoid having support on zero. For example, the (mean specified) shifted negative binomial is

$$(5) \quad g(u; \mu, r) = \frac{\Gamma(r + u - 1)}{\Gamma(r) (u - 1)!} \left(\frac{r}{\mu - 1 + r} \right)^r \left(\frac{\mu - 1}{\mu - 1 + r} \right)^{u-1}$$

which has a mean of $\mu > 1$, and size parameter $r > 0$. The corresponding shifted geometric distribution is given by $g(u; \mu, r = 1)$ and the shifted Poisson density can be defined as the limiting case when $r \rightarrow \infty$.

Figure 3 shows the geometric and negative binomial decay functions along with the shot noise processes corresponding to 8 event days. This illustrates

1 how the form of the shot noise process from (4) is similar to kernel inten- 1
 2 sity estimation (Diggle, 1985), but with weights and one-sided (predictive) 2
 3 kernels. 3
 4

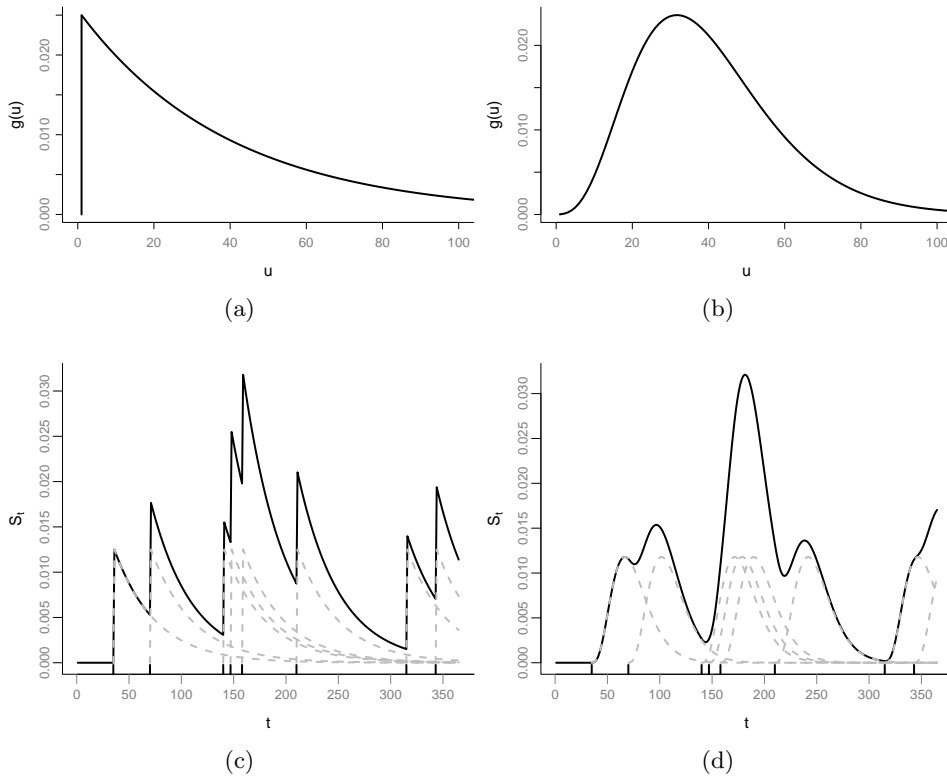


Fig 3: (a) and (b) show the decay functions for the shifted geometric ($\mu = 40$) and shifted negative binomial ($\mu = 40, r = 5$) respectively. (c) and (d) show the corresponding shot noise processes (solid lines) and scaled component kernels (dotted lines) for 8 events with $\alpha_i = \alpha = 0.5$.

3.3. *Self-Exciting Bernoulli Hurdle Process.* The event days, E_t are modeled as a self-exciting Bernoulli process where the probability of an event day is excited (increased) by the occurrence of previous events. Specifically, we consider the hurdle probability on day t to be a function of the process

$$(6) \quad X_t = B_t + S_t$$

where B_t is a baseline process and S_t is the self-exciting component given by (4). The baseline can be a function of any exogenous variables that could

1 have an effect on the process (e.g. social, political or economic conditions 1
 2 and events, counter-terrorism efforts, etc.), but it will not include any infor- 2
 3 mation coming from the internal history. 3

4 To ensure the hurdle probability $p_t \in [0, 1]$, a transformation $\eta(p_t) = X_t$ 4
 5 is used. If the baseline process is non-negative (i.e. $B_t \geq 0$), then $X_t \geq 0$ 5
 6 and an appealing transformation is $\eta(p) = -\log(1 - p)$ corresponding to the 6
 7 hurdle probability 7

$$8 \quad (7) \quad p_t = 1 - e^{-X_t}. \quad 8$$

9 While other transformations (e.g. logit) could be used, this one provides 9
 10 several benefits. It enjoys a likelihood function that is computationally 10
 11 convenient, substituting (7) into (2) obtains the Bernoulli log-likelihood 11
 12 12
 13 13

$$14 \quad (8) \quad \log L_1 = \sum_{1 \leq t_i \leq T} \log(e^{X_{t_i}} - 1) - \sum_{t=1}^T X_t. \quad 14$$

15 The shot noise component of the second sum ($\sum_{t=1}^T X_t = \sum_{t=1}^T B_t + \sum_{t=1}^T S_t$) 15
 16 can be simplified by recognizing that 16
 17 17

$$18 \quad \sum_{t=1}^T S_t = \sum_{i:t_i < T} \alpha_i G(T - t_i) \quad 18$$

19 where G is the cumulative density function of g . This significantly reduces 19
 20 computation as it is only calculated for the event days and not over all days 20
 21 in the observation window. 21

22 This transformation also allows for a straightforward interpretation of the 22
 23 parameters in X_t . Notice that p_t is also the probability that a Poisson ran- 23
 24 dom variable with rate X_t is greater than 0. When events are rare and X_t 24
 25 is small, E_t can be modeled approximately by a Poisson distribution. Accord- 25
 26 ing to the superposition property of the Poisson, the event days arise from 26
 27 two sources, the baseline process with rate B_t and the shot noise process 27
 28 with rate S_t . The rate from the shot noise can be further decomposed into 28
 29 its elements $\alpha_i g(t - t_i)$ for $\{i : t_i < t\}$. Thus, under the Poisson approxi- 29
 30 mation, event day i will generate an expected $\sum_{u=1}^{\infty} \alpha_i g(u) = \alpha_i$ number of 30
 31 additional event days and the decay function $g(\cdot)$ controls when the extra 31
 32 event days will occur. While these interpretations will not hold exactly under 32
 33 the Bernoulli model, they do provide an indication of how each component 33
 34 effects the process. 34
 35 35
 36 36
 37 37
 38 38
 39 39
 40 40
 41 41

3.4. *Survival Functions.* This formulation for the hurdle process also provides simple expressions for some properties of the time until the next event day. For a given time t_0 , let τ be the time of the next event day. The survival function $V_{t_0}(u) = \Pr(\tau - t_0 > u)$ gives the probability that the next event day will be more than u days away. For the self-exciting hurdle process, this becomes

$$\begin{aligned} V_{t_0}(u) &= \Pr(E_{t_0+1} = E_{t_0+2} = \dots = E_{t_0+u} = 0) \\ &= \prod_{\delta=1}^u (1 - p_{t_0+\delta}) \\ &= \exp \left\{ - \sum_{\delta=1}^u X_{t_0+\delta} \right\} \end{aligned}$$

for $u \in \{1, 2, \dots\}$, where $X_{t_0+\delta}$ is calculated assuming no new events have occurred since t_0 . The survival function can be used to calculate the expected time until the next event day

$$\begin{aligned} \mathbb{E}_{t_0}[\tau] &= \mathbb{E}[\tau - t_0] = 1 + \sum_{u=1}^{\infty} V_{t_0}(u) \\ &= 1 + \sum_{u=1}^{\infty} \exp \left\{ - \sum_{\delta=1}^u X_{t_0+\delta} \right\}. \end{aligned}$$

The survival function specification of this model can be useful for prediction, and making inference about future attacks given the current history.

4. Results. In order to build the models and evaluate their predictive performance the data are partitioned into two time periods. The first time period from 1994 through 2000 is used to construct the models and estimate their parameters. The second period from 2001 through 2007 is used to assess the predictive performance of the models.

4.1. *Event Day Modeling.* Recall that the hurdle probability is constructed from the sum of a baseline and self-exciting processes (6). In order to capture potential seasonality and other large scale trends in the data the baseline process B_t is defined as

$$(9) \quad \log(B_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + A_1 \sin(2\pi t/\omega) + A_2 \cos(2\pi t/\omega).$$

where the seasonal terms have a period of $\omega = 365.25$ days. This results in five baseline parameters $\theta_B = (\beta_0, \beta_1, \beta_2, A_1, A_2) \in \mathbb{R}^5$. The log transform

TABLE 2
Parameter estimates for the event day models for the training period (1994-2000). The BL are the baseline only and the SE are the self-exciting models.

Model	AIC	β_0	β_1	β_2	A_1	A_2	α	μ	r
BL ₁	1187.77	-2.75							
BL ₂	1149.54	-8.2	8.05						
BL ₃	1151.25	-2.83	-8.01	11.91					
BL ₄	1171.22	-2.82			-0.42	-0.31			
BL ₅	1136.80	-8.02	7.69		-0.35	-0.32			
BL ₆	1138.60	-3.53	-5.74	9.95	-0.35	-0.32			
SE₁	1075.21	-4.41					0.89	37.54	0.45
SE ₂	1077.16	-5.17	1.23				0.87	36.55	0.45
SE ₃	1078.71	15.31	-63.2	50.24			0.85	35.69	0.45
SE ₄	1077.03	-4.34			-0.41	-0.40	0.86	38.55	0.44
SE ₅	1078.76	-5.99	2.68		-0.38	-0.42	0.82	36.57	0.43
SE ₆	1079.60	20.46	-79.33	62.91	-0.43	-0.45	0.80	37.28	0.43

ensures $B_t \geq 0$. We considered self-exciting processes of the form

$$(10) \quad S_t(\alpha, \mu, r) = \alpha \sum_{i:t_i < t} g(t - t_i; \mu, r)$$

where $g(u; \mu, r)$ is a shifted negative binomial decay function (5) and the magnitude α is a constant. This results in three parameters for the self-exciting component $\theta_S = (\alpha, \mu, r) \in \mathbb{R}_+^3$.

The combined model for $X_t = B_t + S_t$ results in up to 8 parameters to estimate. Estimation is carried out by maximizing the log-likelihood function given in (8). Since there is no explicit solution available, we used R's numerical optimization routine `nlm` (R Development Core Team, 2011) to obtain estimates. For all models, the estimates converged within seconds from a variety of starting points.

Table 2 shows the evaluated models and their corresponding AIC scores. The lowest AIC belongs to the four component model with a constant baseline (SE₁), providing essentially a discrete time version of the Hawkes model (Hawkes, 1971; Ozaki, 1979). This AIC is much lower than the best baseline only model (BL₅) which includes the periodic terms. Figure 2 plots the weighted K -function for SE₁ and BL₅ showing the self-exciting model is much tighter around 0 indicating a better fit over the baseline only model (Diggle, 1979). Figure 4 shows the estimated hurdle probability in the training period for both models.

The parameters of the self-exciting model (SE₁) suggest a baseline hurdle probability of about 0.012 when the shot noise term drops to zero. However,

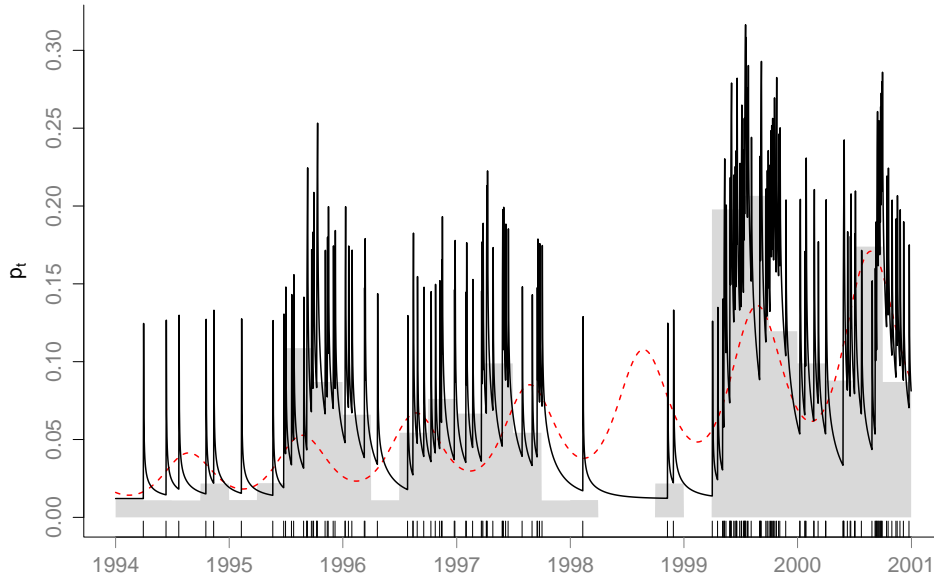


Fig 4: Hurdle probabilities for the best fitting self-exciting model SE_1 (solid line), the best fitting baseline only model BL_5 (dotted line), and the observed piecewise constant probabilities (blocks). The rug signifies the event days.

according to the Poisson approximation, every event day will generate an expected $\alpha = 0.89$ additional event days. The generated event day is expected to occur in $\mu = 37.54$ days, but the time until occurrence has a median of only 16 days.

4.2. *Count Modeling.* It can be obtained from Table 1 that most event days (92.4%) are comprised of only 1 or 2 attacks. However, several days (1.9%) have more than 9 attacks (with one day having 36 recorded terrorist attacks). This suggests count distributions that have an initial rapid decay, but still possess a long tail to accommodate the extreme values. One such discrete distribution is the Riemann zeta or discrete Pareto distribution. This power law distribution has recently been employed to model the severity (e.g. number killed or injured) of terrorist attacks (Clauset, Young and Gleditsch, 2007).

Using it to model the event day counts, under an i.i.d assumption, gives the probability mass function

$$f_t(y; s) = \frac{y^{-s}}{\zeta(s)} \quad y \in \{1, 2, \dots\}$$

where $\zeta(s) = \sum_{x=1}^{\infty} x^{-s}$ is the Riemann zeta function and the parameter $s \in (1, \infty)$. This one parameter model is easy to estimate numerically via maximum likelihood (Seal, 1952; Goldstein, Morris and Yen, 2004) by finding the s that satisfies the equation $\frac{1}{n} \sum_{i=1}^n \log(y_i) = -\frac{\zeta'(s)}{\zeta(s)}$. The training period data gives rise to an estimated parameter of $\hat{s} = 2.86$. The bootstrap Kolmogorov-Smirnov test of Clauset, Shalizi and Newman (2009) gives a p -value of 0.69, suggesting the zeta distribution provides a suitable fit.

While the constant parameter zeta distribution appears to provide a good fit, there may be better models especially if the distribution is allowed to vary over time. In particular, as for the hurdle component, distributions that vary over time in response to a shot noise process are considered. An additional driving process (with shot noise) was implemented for the count model, namely

$$X_t^c(\beta^c, \alpha^c, \mu^c) = \beta^c + \alpha^c \sum_{i:t_i < t} Y_i \cdot g^c(t - t_i; \mu^c)$$

where β^c is a constant baseline, $\alpha_i^c = \alpha^c \cdot Y_i$ varies in response to the number of attacks on the i^{th} event day, and the decay function $g^c(u; \mu^c)$ is the one parameter shifted geometric function with mean μ^c . This results in three parameters for the count component $\theta_C = (\beta^c, \alpha^c, \mu^c)$. Note that this is a separate process than the one specified for the hurdle component given by (9) and (10). The additional flexibility afforded by the hurdle specification allows such additional complexity to be introduced for the count component of the model.

There are two primary hypotheses about how the count distribution behaves in response to the event history. First, counts could respond in a self-exciting manner, where recent event and multi-event days increase the likelihood of further multi-event days. Alternatively, due to the depletion of resources or increased anti-terrorism measures after large attacks, counts may respond in a more self-inhibiting manner, where recent multi-event days lower the chances of multi-events day in the near future.

To evaluate the self-exciting hypothesis, let $s_t = (1 - e^{-X_t^c})^{-1}$ since smaller values of s lead to more mass on the extreme values. Alternatively, $s_t = e^{X_t^c}$ for the self-inhibiting model. Table 3 shows the estimated parameters and AIC scores for the fitted count models. The lower AIC suggests that the self-exciting count model has more validity than the self-inhibiting model for the training data. Figure 5 shows the estimated probability of counts under the constant parameter (C_z) and self-exciting (C_{se}) models. While the self-exciting model puts more mass on multi-event days when X_t^c is large, the decay is rapid and the self-exciting behavior is short lived - only

TABLE 3
 Parameter estimates for the count models for the training period (1994-2000). C_z , C_{se} , and C_{si} are the constant parameter, self-exciting, and self-inhibiting forms (respectively) of the Riemann zeta distribution.

Model	AIC	s	β	α	μ
C_z	241.00	2.86			
C_{se}	240.48	$(1 - e^{-X_t})^{-1}$	0.375	0.20	2.13
C_{si}	244.72	e^{X_t}	1.00	0.34	55.18

for a few days, before the shot noise term approaches 0 and its effects become negligible.

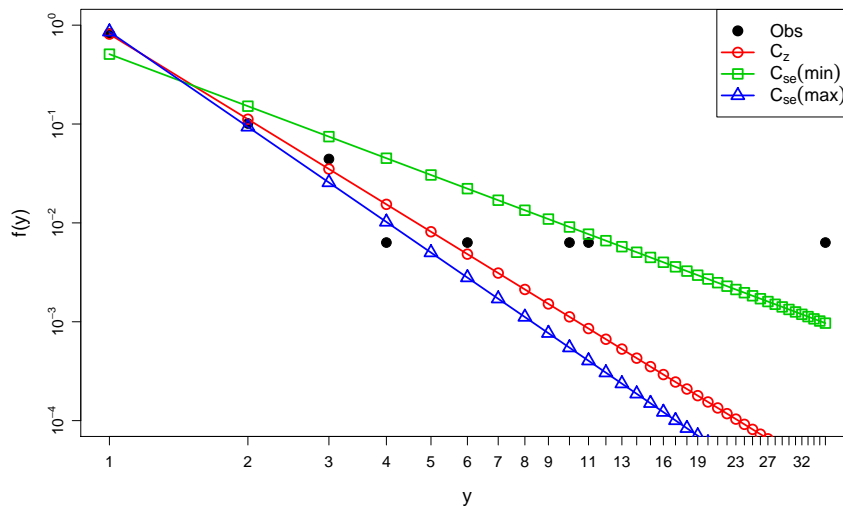


Fig 5: Estimated count probabilities for the constant parameter model (C_z with $s = 2.86$) and for the self-exciting count model (C_{se}) at the baseline state ($s = 3.19$) and most excited state ($s = 1.75$).

4.3. *Testing Model Predictions.* Based on the training period, models are selected and fit for both the hurdle and count components. To evaluate how well this can predict future terrorist activity, the models are applied to the terrorist incidents in the testing period (2001-2007) using the following procedure.

1. At day t predict the probability of an event day and conditional number of attacks on day $t + 1$, based on the current history \mathcal{H}_t .

TABLE 4

The hurdle component log probability score, G using the constant baseline only model (BL_1) as the reference model. The AIC scores are from the training period (see Table 2).

Model	AIC	G	Model	AIC	G
BL ₁	1187.77	0.00	SE ₁	1075.21	26.96
BL ₂	1149.54	-19.10	SE ₂	1077.16	25.47
BL ₃	1151.25	20.69	SE ₃	1078.71	24.96
BL ₄	1171.22	-1.39	SE ₄	1077.03	26.22
BL ₅	1136.80	-18.21	SE ₅	1078.76	24.56
BL ₆	1138.60	21.63	SE ₆	1079.60	24.57

2. Observe Y_{t+1} and re-estimate the model parameters based on the updated history \mathcal{H}_{t+1} .
3. Repeat steps 1 and 2 through the testing period to construct a final model based on all data between 1994-2007.

This update and forecast scheme results in a set of daily predictions $\hat{p}_t = \Pr(E_t = 1 | \mathcal{H}_{t-1})$ and $\hat{f}_t(y) = \Pr(Y_t = y | E_t = 1, \mathcal{H}_{t-1})$ conditioned on the past history of the process.

The predictive capability of the models are evaluated by comparing the log probability gain, or predictive log-likelihood ratio, of each model (Daley and Vere-Jones, 2004). This score is a relative measure of the improvement in predicative capability for a model compared to a reference or null model. The interpretation is that values greater than 0 show a relative improvement in predictions over the reference model, with the highest values indicating the preferred model.

For the hurdle component, the log probability gain is

$$G = \sum_{t \in \mathcal{T}_2} E_t \log \frac{\hat{p}_t}{\hat{\pi}_t} + (1 - E_t) \log \frac{1 - \hat{p}_t}{1 - \hat{\pi}_t}$$

where \hat{p}_t comes from the model of interest and $\hat{\pi}_t$ from the constant baseline only model (BL_1), and $\mathcal{T}_2 = [01 \text{ Jan } 2001, 31 \text{ Dec } 2007]$ covers the testing period. Table 4 compares the log probability gains G and the AIC scores from the training period, and shows that the results from the test period were similar to those in the training period, supporting the use of AIC for model selection.

The constant baseline self-exciting model (SE_1) has the largest log probability gain and all self-exciting models out-perform all of the baseline only models. The best baseline only model for the testing period was the full model with seasonality and a quadratic trend (BL_6). This is in contrast to

TABLE 5
The count component log probability score, G^c using the constant parameter zeta model (C_z) as the reference model.

	C_s	C_{se}	C_{si}
G^c	0.00	8.07	-0.27

the the training period results where best baseline only model had a linear trend (BL₅). The linear trend model failed to perform as well during the testing period as the general attack rate peaked around the transition point and started to decreasing during the remainder of the testing period creating a significant quadratic trend over the combined observation period. Alternatively, the self-exciting models can adapt rapidly to sudden changes in attack rate yielding better overall predictive ability.

The count component uses the log probability gain score

$$G^c = \sum_{i:t_i \in \mathcal{T}_2} \log \frac{\hat{f}(y_i)}{\hat{h}(y_i)}$$

where \hat{f} is the model of interest and \hat{h} is the reference model. Table 5 shows the values of G^c using the constant parameter zeta model (C_z) as the reference model. These results show the self-exciting zeta model (C_{se}) as the preferred model for predicting the counts. As with the hurdle component, the predictive performance of the count models followed the AIC scores from the training period. However, the model with self-exciting component does predict substantially better than the constant component model - more so than would be expected from the AIC scores in the training period (see Table 3). This may be due to changes in the attack behavior during the testing period. Figure 1 shows that while there were a few large counts during 2001, there were very few multi-attack days after. This change in the pattern of multiple attacks accounts for the improved predictive capabilities of the self-exciting model over the constant parameter model. The self-exciting count model is able to quickly adapt to changing conditions and consequently make better predictions.

In summary, the daily counts of Indonesian terrorist attacks were modeled with a self-exciting hurdle model. The separation of the model into two components allowed both the attack rate and attack characteristics to be more faithfully represented. It was found that including a self-exciting process benefited both the hurdle and count components by allowing the models to quickly adapt to changing attack behavior. It is this property as

1 a predictive tool that makes the self-exciting hurdle model especially useful 1
 2 in an applied setting where prediction is of primary interest. 2
 3

4 **5. Conclusion.** In practice, policy makers are faced with limited re- 4
 5 sources, financially, materially and in personnel. The allocation of these re- 5
 6 sources depends on an understanding of the future risk of terrorist activity. 6
 7 As an example, in the wake of the attacks of September 11, 2001, policy mak- 7
 8 ers placed additional security in the form of National Guard troops in all US 8
 9 commercial airports. This was done in hopes of re-establishing public trust 9
 10 in the safety of air travel while the intelligence community was assessing the 10
 11 risk of subsequent attacks ([Price and Forrest, 2009](#)). The obvious question 11
 12 in that case was when to shift resources from increased airport security to 12
 13 investing in long-term strategies to improve intelligence capabilities. As this 13
 14 case illustrates, only by clearly understanding the risk of terrorist activity 14
 15 can policy makers make informed decisions about allocating resources. 15

16 This paper presents a two-component self-exciting model for the analysis 16
 17 and prediction of future terrorist activity. It was found that the best model 17
 18 used a constant baseline with self-exciting term for the hurdle component 18
 19 and a Riemann zeta distribution with shot noise driven parameter for the 19
 20 count component. The improvement offered by the self-exciting term pro- 20
 21 vides support for the contagion theory and suggests a significant short term 21
 22 increase in terrorism risk after an attack. The use of a power law distribution 22
 23 for the number of same day attacks corresponds with the power law behav- 23
 24 ior of the number killed in terrorist attacks ([Clauset, Young and Gleditsch, 2007](#)). 24
 25 This is to be expected as multiple coordinated attacks would tend to 25
 26 indicate better planned, and hence more severe, terrorist activities. 26

27 The self-exciting hurdle model adheres to the theoretical concept for a 27
 28 contagion effect to terrorism as manifested in the clustering of data while 28
 29 providing good fit and predictive capabilities without relying on exogenous 29
 30 variables. The model provides a simple structure and interpretation of the 30
 31 parameters useful for understanding the dynamic nature of the terrorist ac- 31
 32 tivity. This provides an appropriate starting point for exploring additional 32
 33 covariate effects, including analysis concerning the effectiveness of counter- 33
 34 terrorism activities, geography, political, or economical factors. As the attack 34
 35 clustering may be partially attributable to a stochastic baseline driven by 35
 36 such exogenous processes ([Holden, 1986](#)), further analysis could include co- 36
 37 variates in the baseline model (9) to test the impact on the self-exciting 37
 38 component. 38

39 By focusing on the timing of terrorist attacks on the islands of Indonesia, 39
 40 other aspects of the attacks, such as location, attack type, and group re- 40
 41

sponsible, was not considered. Including such additional information could lead to more detailed models that better explain the nuances of the terrorist activity. For example, the self-exciting component could include a spatial proximity term using the models of Mohler et al. (2011) or multiple self-exciting terms could be included that represent how attacks from one terrorist group influence the subsequent attacks of other groups.

The utility of the models presented here and their ease of implementation and interpretation make them a potentially useful tool in security related fields. The results show that the risk of terrorist activity can vary greatly over short periods of time, thus policy responses in terms of resource allocation, security, and counter-terrorism responses should reflect this as well as addressing the more long-term trends in risk. For example, understanding the short-term variations in risk would allow a more effective deployment and assessment of additional airport security measures in the wake of a terrorist attack, while a grasp of the longer term trends could help guide the development and assessment of more strategic counter-terrorism resources, such as increasing the number of foreign language experts available for translation or developing effective de-radicalization programs to prevent the growth of terrorist groups.

References.

- BADDELEY, A., MØLLER, J. and WAAGEPETERSEN, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54** 329–350.
- BARROS, C. P. (2003). An intervention analysis of terrorism: The Spanish ETA case. *Defence and Peace Economics* **6** 401–412.
- BELASCO, A. (2010). The Cost of Iraq, Afghanistan, and Other Global War on Terror Operations Since 9/11 Technical Report, Congressional Research Services Report RL33110.
- CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review* **51** 661–703.
- CLAUSET, A., YOUNG, M. and GLEDITSCH, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution* **51** 58–87.
- DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes I*, 2nd ed. Springer-Verlag, New York, NY.
- DALEY, D. J. and VERE-JONES, D. (2004). Scoring probability forecasts for point processes: the entropy score and information gain. *Journal of Applied Probability* **41** 297–312.
- DIGGLE, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35** 87–101.
- DIGGLE, P. (1985). A kernel method for smoothing point process data. *Journal of The Royal Statistical Society, Series C, Applied Statistics* **34** 138–147.
- DIXON, P. M. (2002). Ripley’s K function. In *Encyclopedia of Econometrics*, (A. H. El-Shaarawi and W. W. Piegorsch, eds.) **2** 1796–1803. Wiley, Chichester.

- 1 DUGAN, L., LAFREE, G. and PIQUERO, A. (2005). Testing a rational choice model of 1
2 airline hijackings. *Criminology* **43** 1031–1066. 2
- 3 ENDERS, W. and SANDLER, T. (1993). The effectiveness of antiterrorism policies: A vector- 3
4 autoregression-intervention analysis. *The American Political Science Review* **4** 829–844. 4
- 5 ENDERS, W. and SANDLER, T. (2000). Is transnational terrorism becoming more threat- 5
6 ening? *Journal of Conflict Resolution* **44** 307–332. 6
- 7 ENDERS, W. and SANDLER, T. (2002). Patterns of transnational terrorism, 1970-1999: 7
8 Alternative time-series estimates. *International Studies Quarterly* **2** 145–165. 8
- 9 ENDERS, W. and SANDLER, T. (2006). *The Political Economy of Terrorism*. Cambridge 9
10 University Press. 10
- 11 GOLDSTEIN, M. L., MORRIS, S. A. and YEN, G. G. (2004). Problems with fitting to the 11
12 power-law distribution. *The European Physical Journal B* **41** 255–258. 12
- 13 HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point pro- 13
14 cesses. *Biometrika* **58** 83-90. 14
- 15 HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting 15
16 process. *Journal of Applied Probability* **11** 493-503. 16
- 17 HEARD, N. A., WESTON, D. J., PLATANIOTI, K. and HAND, D. J. (2010). Bayesian 17
18 anomaly detection methods for social networks. *Annals of Applied Statistics* **4** 645-662. 18
- 19 HEILBRON, D. C. (1994). Zero-altered and other regression models for count data with 19
20 added zeros. *Biometrical Journal* **36** 531–547. 20
- 21 HOLDEN, R. T. (1986). The contagiousness of aircraft hijacking. *The American Journal* 21
22 *of Sociology* **91** 874–904. 22
- 23 HOLDEN, R. T. (1987). Time series analysis of a contagious process. *Journal of the Amer-* 23
24 *ican Statistical Association* **82** 1019-1028. 24
- 25 JONES, A. M. (1994). Health, addiction, social interaction and the decision to quit smok- 25
26 ing. *Journal of Health Economics* **13** 93-110. 26
- 27 LAFREE, G. and DUGAN, L. (2007). Introducing the Global Terrorism Database. *Terror-* 27
28 *ism and Political Violence* **19** 181-204. 28
- 29 LAFREE, G., DUGAN, L. and KORTE, R. (2009). The Impact of British counter terrorist 29
30 strategies on political violence in Northern Ireland: comparing deterrence and backlash 30
31 models. *Criminology* **47** 17–45. 31
- 32 LAFREE, G., MORRIS, N. A. and DUGAN, L. (2010). Cross-national patterns of terrorism: 32
33 comparing trajectories for total, attributed and fatal attacks, 1970-2006. *British Journal* 33
34 *of Criminology* **50** 622–649. 34
- 35 LAMBERT, D. (1992). Zero-inflated Poisson Regression, With an Application to Defects 35
36 in Manufacturing. *Technometrics* **34** 1-14. 36
- 37 LI, R. P. and THOMPSON, W. R. (1975). The “Coups Contagion” Hypothesis. *The Journal* 37
38 *of Conflict Resolution* **19** 63-88. 38
- 39 LUM C., K. L. W. and SHERLEY, A. J. (2006). Are counter-terrorism strategies effective? 39
40 The results of the Campbell systematic review on counter-terrorism evaluation research. 40
41 *Journal of Experimental Criminology* **2** 489–516. 41
- MARSCHALL, M. J., RUHIL, V. S. ANIRUDH and SHAH, P. R. (2010). The New Rad-
ical Calculus: Electoral Institutions and Black Representation in Local Legislatures.
American Journal of Political Science **54** 107-124.
- MIDLARSKY, M. I. (1978). Analyzing diffusion and contagion effects: The urban disorders
of the 1960s. *The American Political Science Review* **72** 996-1008.
- MIDLARSKY, M. I., CRENSHAW, M. and YOSHIDA, F. (1980). Why Violence Spreads: The
Contagion of International Terrorism. *International Studies Quarterly* **24** 341-365.
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and
TITA, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the Amer-*

1 *ican Statistical Association* **106** 100-109. 1

2 MULLAHY, J. (1986). Specification and testing of some modified count data models. *Jour-* 2

3 *nal of Econometrics* **33** 341–365. 3

4 MYERS, D. J. (2000). The Diffusion of Collective Violence: Infectiousness, Susceptibility, 4

5 and Mass Media Networks. *American Journal of Sociology* **106** 178-208. 5

6 NAGIN, D. (2005). *Group-based Modeling of Development*. Harvard University Press, Cam- 6

7 bridge, MA. 7

8 OZAKI, T. (1979). Maximum likelihood estimation of Hawkes’ self-exciting point processes. 8

9 *Annals of the Institute of Statistical Mathematics* **31** 145–155. 9

10 PERL, R. (2007). Combating Terrorism: The Challenge of Measuring Effectiveness Tech- 10

11 nical Report, Congressional Research Services Report RL33160. 11

12 PRICE, J. C. and FORREST, J. S. (2009). *Practical Aviation Security: Predicting and* 12

13 *Preventing Future Threats*. Butterworth-Heinemann, Oxford, UK. 13

14 R DEVELOPMENT CORE TEAM, (2011). R: A Language and Environment for Statistical 14

15 Computing R Foundation for Statistical Computing, Vienna, Austria. 15

16 RICE, J. (1977). On Generalized Shot Noise. *Advances in Applied Probability* **9** 553-565. 16

17 SEAL, H. L. (1952). The maximum likelihood fitting of the discrete Pareto law. *Journal* 17

18 *of the Institute of Actuaries* **78** 115-121. 18

19 TOWNSLEY, M., JOHNSON, S. D. and RATCLIFFE, J. H. (2008). The time dynamics of 19

20 insurgent activity in Iraq. *Security Journal* **21** 139-146. 20

21 VEEN, A. and SCHOENBERG, F. P. (2006). Assessing Spatial Point Process Models Using 21

22 Weighted K -functions: Analysis of California Earthquakes. In *Case Studies in Spatial* 22

23 *Point Process Modeling*, (A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan, 23

24 eds.). *Lecture Notes in Statistics* **185** 293-306. Springer New York. 24

25 WELSH, A., CUNNINGHAM, R. B., DONNELLY, C. F. and LINDENMAYER, D. B. (1996). 25

26 Methods for analysing data with extra zero: ZIP regression models with application for 26

27 surveys of rare species. *Ecological Modelling* **88** 297-308. 27

28 28

29 29

30 30

31 31

32 32

33 33

34 34

35 35

36 36

37 37

38 38

39 39

40 40

41 41