New Findings from Terrorism Data: Dirichlet Process Random Effects Models for Latent Groups

Minjung Kyung* University of Florida

Jeff Gill† Washington University

George Casella‡ University of Florida

May 1, 2011

Abstract

Data obtained describing terrorist events are particularly difficult to analyze, due to the many problems associated with the both the data collection process, the inherent variability in the data itself, and the usually poor level of measurement coming from observing political actors that seek *not* to provide reliable data on their activities. Thus, there is a need for sophisticated modeling to obtain reasonable inferences from these data. Here we develop a logistic random effects specification using a Dirichlet process to model the random effects. We first look at how such a model can best be implemented, and then we use the model to analyze terrorism data. We see that the richer Dirichlet process random effects model, as compared to a normal random effects model, is able to remove more of the underlying variability from the data, uncovering latent information that would not otherwise have been revealed.

AMS 2000 subject classifications: Primary 62F99; secondary 62P25; secondary 62G99.

Keywords and phrases: terrorists, empirical studies of terrorism, logistic regression, generalized linear mixed models, hierarchical models, Gibbs sampling, Metropolis-Hastings Algorithm.

*Assistant Professor, Center for Applied Statistics, Washington University, One Brookings Dr., Seigle Hall, St. Louis, MO. Supported by National Science Foundation Grants DMS-0631632 and SES-0631588.Email: mkyung@artsci.wustl.edu.

†Professor, Center for Applied Statistics, Washington University, One Brookings Dr., Seigle Hall, St. Louis, MO. Email: jgill@wustl.edu.

‡Distinguished Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Supported by National Science Foundation Grants DMS-0631632 and SES-0631588. Email: casella@ufl.edu.

1 Introduction

The analysis of data on terrorists and terrorist attacks is problematic. These data are either observed public events, which omit planned but failed or canceled attacks, or classified information at government agencies that are not available to general researchers. Also, most of the datasets focus purely on *incidents*: an observed violent attack along with covariates such as responsible group, target characteristics, as well as the extent of casualties and damage. This selection along the outcome of interest, observable violent actions, effectively makes the associated analytical work a *case-control study*.

Analyzing terrorism incidents data has challenges that go significantly beyond those with regular events-based data or survey research data. Social science data can be extremely difficult to assess if the creators of the dataset are uncooperative subjects. The key problem is that humans in covert, especially dedicated terrorist networks, for strategic reasons work to conceal not just their identities and intentions, but also their interactions with others. As a result of these collection issues, the data contain confounding effects, overlapping explanatory variables, high measurement error, and unmeasured clustering forces. Confounding effects and overlapping explanations exist because the observed variables are often surrogates or summaries of the actual causal factors involved. The level of measurement here is typically from highly non-granular nominal variables, and sometimes only purely qualitative information in the form of textual description. These problems could be considered missing data issues, but the general statistical tools for handling missing data (eg. Little and Rubin 2002) are unhelpful since the pattern of missingness here is controlled by actors who routinely seek to hide information.

Despite these seemingly insurmountable challenges, terrorism remains an extremely important problem because it affects personal safety, internal government policies, public perception, and relations between nations. Normally, researchers faced with such problems would simply abandon the data at hand and develop strategies for collecting new, and more reliable, information. This simply is not feasible. Collecting data on terrorist activities beyond journalistic accounts is expensive, reliant on governments, and sometimes even physically dangerous to the researcher. While confounding, correlation, and measurement error in explanatory variables can often be handled with careful model specification, the most daunting problem remains the latent clustering that occurs because terrorist actors are imitative of each other, communicate between groups, and sometimes rely on the same resources.

We address the latent clustering problem with an improved implementation of nonparametric Bayesian generalized linear models using a logistic link function and Dirichlet process random effects. We have found that using Dirichlet process priors for the random effects can capture latent information in the data that would not otherwise be modeled (Gill and

Casella 2009). This provides a means of incorporating subclustering information in the data on terrorist events that would not otherwise be revealed in a standard context, even if it cannot be directly parameterized in the model, thus improving model fit. The term "subcluster" is used in this context instead of "cluster" since the modeling approach does not penalize for over-fragmentation of the real clusters hidden in the data. Here we use this tool to study suicide terrorist attacks in the Middle-East and Northern Africa from 1998 to 2004.

Our approach is based on a new Gibbs sampling method that is developed in Kyung *et al.* (2010b). In the work here, we generalize the newly developed Gibbs sampling based on a mixture representation of the logistic distribution in Kyung *et al.* , and we find that the Gibbs sampler appears to mix more rapidly than the commonly used alternative slice sampler of Damien *et al.* (1999). We show that the slice sampler typically has higher autocorrelation and poorer mixing than the approach taken here.

Our results support a theorized but not fully explored notion about terrorist groups. There are groups or clusters of terrorists who work in similar fashion, either because they communicate or because they emulate each other. Information from the MCMC sampling scheme produces coefficient estimates that account for this latent trait and are therefore conditioned on it. Our findings demonstrate that there are reliable coefficient estimates for attack characteristics that are shared across groups in the Middle East and Northern Africa over the period 1998 to 2004. This includes a number of sensationalist attacks against large government buildings, as well many more routine smaller events.

2 Background On Terrorism Datasets

There are now several major databases on terrorist incidents stored at academic institutions like the University of Maryland (START) and at government agencies like the U.S. Homeland Security Agency. Event data on terrorist hostage incidents are provided by International Terrorism: Attributes of Terrorist Events (ITERATE), which was originally assembled by Mickolus (1982) and later updated (2006). ITERATE records transnational terrorist incidents and therefore ignores so-called "domestic terrorism." Conversely the dataset *Political Violence in the United States, 1819–1968* (Levy, Graham and Gurr 1969) lists incidents and covariates for political violence resulting in injury or death for about 150 years. The International Policy Institute for Counter-Terrorism in Herzlia, Israel provides a dataset of terrorist attacks in Israel. The U.S. Department of Homeland Security supports the National Memorial Institute for the Prevention of Terrorism (MIPT) Terrorism Knowledge Base (TKB), which provides online a listing of terrorism incidents with information on the terrorists and an emphasis on legal information. Another online listing is the The Global Terrorism Database (GTD) which includes information on global terrorist events starting from 1970. Researchers have used these primarily for creating summary statistics and basic tabular analyses. Standard statistical modeling has yielded some insights into the determinants and timing of terrorists incidents (Enders 2007, Enders and Sandler 1995, and Li and Schaub 2004, for example), but with limited results.

Game theoretic approaches have been used to circumvent the data quality problem. The way governments and terrorist (factions) strategize has been studied, for instance, to explain why extremist groups often increase terrorist activity after a government makes concessions to moderate factions. These results have yielded important insights about terrorism in Ireland, Spain, and other countries (Bueno de Mesquita, 2005; see also such works as Bueno de Mesquita and Dickson 2007, Arce and Sandler 2007, Sandler and Arce 2003, Kydd and Walter 2002, and Siqueira and Sandler 2006), but they are not in general motivated by conventional data analysis.

Another approach is to build models of networks of terrorist and terrorist organizations. Despite the convention that terrorists should not be treated as unitary actors (Chai 1993, Crenshaw 1981), the study of terrorist organizations as networks is less developed. Social network analysts have discovered that covert organizations tend to be cellular and distributed rather than hierarchical (Carley 2004, Krebs 2002, Rothenberg 2002), and the government has supported research to model these as standard networks (National Research Council 2003). Carley (2003, 2006 uses "meta-matrices" that capture not just the ties between terrorists but also their knowledge and tasks from semiautomatic parsing of signal traffic (Tsvetovat and Carley 2006). This work yields, among other things, lessons about how best to "destabilize" terrorist networks (Tsvetovat and Carley 2005, Moon and Carley 2007).

None of these research tracks has been very successful in building standard causal regression models preferred in the social sciences. This is because the data are, in general, poorly measured and highly nongranular. Specifically, we observe roughly measured categorical variables produced by governments as they record and react to attacks on their soil or against their citizens elsewhere. A huge part of the problem is that the primary actors under study are deliberately trying to prevent such data from being collected in an accurate and useful manner. So unlike many missing data or poorly measured data problems, the key agents in the data generation process are not denying observation in benign ways, they are attempting to deny observation in willful and strategic ways. Key information, that is almost always missing from such data, is the intentions of terrorist, strategic alternatives that they face, and actions that failed to achieve any result. All of these problems above mean that the analyst has a doubly difficult task in creating meaningful inferential models.

3 Data on Suicide Attacks

The data we use here come from the Global Terrorism Database II (LaFree & Dugan 2008), restricted to events in the Middle East and Northern Africa from 1998 to 2004. After removing almost totally incomplete cases, this provides 1041 violent attacks by terrorist groups, 154 (15%) of which were suicide attacks where at least one of the individual assailants was killed by design. Our outcome variable of interest is therefore the dichotomous observation of a suicide attack or not. Suicide attacks pose a substantially higher challenge for governments since the assailant has great control over placement and timing and also does not need to plan his or her escape (Pape 2006). Thus they may inflict severe damage to otherwise-safe targets.

Terrorism data are notoriously difficult to model from a regression context, so we first pick a rich set of explanatory variables. Time is important during this era. It is also complex. Our approach is to designate the start of the period of study as an indicator variable so that estimated positive or negative coefficients are revealing about a beginning trend. There were 273 terrorist attacks worldwide in 1998 with a recorded high of 741 killed along with 5952 injured (*Patterns of Global Terrorism 1998*, U.S. State Department). Of these 273, 103 were in our geographic area of study. This particular year was also notable for the incredibly destructive simultaneous bombings of the U.S. Embassies in Nairobi, Kenya (291 killed, roughly 5000 injured), and Dar es Salaam, Tanzania (10 killed, 77 injured) in August.

Relatedly, we also consider a variable, `MULT.INCIDENT`, which indicates whether the attack is part of a coordinated multi-site event. This was the case in a surprisingly large number of cases, 136 (13.1%). Including this explanatory variable in the model is a means of testing the relationship between connected events at different sites and the use of suicide attackers.

Sometimes there are multiple parties involved in a regional conflict, such as the wars in Iraq and Afghanistan. This can make it more difficult to assess responsibilities. Strategic actors sometimes make false claims about responsibility (both for and against) as a means of concealing actions and intentions. Normally, but not always, a suicide attack can be traced to a specific group since there is identifying biological information left at the incident site. Also, in some settings, the family is forthcoming about the identity of the attacker. We measure this effect with the variable `MULT.PARTY`, where (also) 136 out of 1041 cases are coded as one.

There can also be terrorist incidents in which there is substantial uncertainty about the identity of any single attacking group. The variable `SUSP.UNCONFIRM` is coded as one (209/1041) if government officials express notable doubt about attributing responsibility. As with the involvement with multiple parties, we expect this to be less common for suicide attacks (25/209 which is 12% of such cases).

One key issue is the linkage between the success of the attack and the suicide nature of the attack. Obviously, there is some linkage or groups would not persist in organizing suicide attacks over time. Since this is not in question, we ask the question from the perspective of the attacked party, who's law enforcement, military, and intelligence services seek to understand and prevent further attacks. By using the variable SUCCESSFUL (true in 966 of the events) on the right-hand-side of the model specification, we are asking: given that it is a successful attack, how likely is it that a suicide assailant was used?

An important issue is the type of attack used. The data are presented in a quasi-ordered fashion with the categories (counts): Armed Assault (385), Assassination (47), Bombing/Explosion (514), Facility/Infrastructure Attack (50), Hijacking (7), Barricade Incident (3), and Kidnapping (35). The model results were highly robust to different recodings of this variable since armed assaults and bombings dominate. We therefore left the original ATTACK.TYPE coding in place such that for suicide attackers a positive estimated coefficient shows a preference for bombing strategies and a negative estimated coefficient shows a preferences for armed assault. We also consider a variable, WEAPON.TYPE, that is coded one for the use of: explosives, dynamite, or general bombs to help differentiate the suicide bombings from

	Suicide Versus Bomb	
	Not Bomb	**Bomb**
Not Suicide	720	661
Suicide	5	224

non-suicide bombings. This relationship is unbalanced since the bulk of attacks are not from suicide attackers. However, suicide attackers are typically more effective in creating casualties and the psychology of suicide attackers is more effective in creating terror within populations. The key relationship here is the heavy use of bombs by suicide attackers. Note that our subsequent analysis uses the coding described in ATTACK.TYPE, not the dichotomization in this table.

Of course the targets that terrorists pick are important to understand. The primary distinction, which we focus on here, is civilian (832 cases, zero coding) versus military (209 cases, one coding) targets. As noted terrorists tend to prefer civilian targets since they are usually "softer" (less guarded), more plentiful, and the perceived randomness of death and destruction has a greater effect on the morale of the citizens. This explanatory variable is labeled TARGET.TYPE.

In addition to being successful, we also want to incorporate the extent of human damage that the terrorist attack inflicts. Again, we look at this not from the terrorist perspective but from the government perspective by including NUM.FATAL (3424 total) and NUM.INJUR (8123 total) on the right-hand-side of the model. So our substantive question is: for higher (lower) levels of casualties, what is the likelihood, controlling for other factors, that the terrorists used a suicide attacker? Another measure of damage is the negative psychological/social impact that terrorist attacks inflict upon national populations. In fact, this is usually the most widespread effect that such events produce. The collectors of these data subjectively coded this into `PSYCHOSOCIAL` with ascending levels: none, minor, moderate, and major (summary below). We also look at a variable for property damage, which can vary substantially. The data are provided with yes/no values along with "minor" in-between in the variable `PROPERTY.DAMAGE`.

3.1 Visualizing the Suicide Data

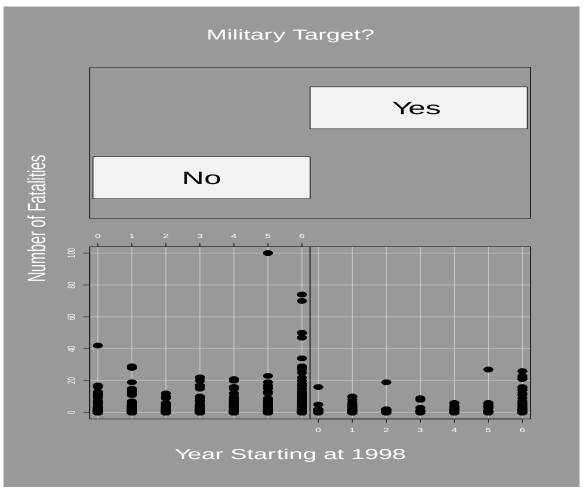
Consider the four coplots in the four frames of in Figure 1, where two variables are plotted in each, at levels of a different binary third variable, with different mixes of variables.

Notice the similarities and differences in the distribution of fatalities between the two upper frames. The annual pattern of fatalities for not using a bomb and attacking civilian targets are very close, except for some high death events at the end of the series in the first plot, which are attributable to bombs used on non-military targets. Thus the relationship is subtle and it spans four variables: year, military target, bomb used, and number of fatalities. *This confounding of effects leads to collinearity and is typical of terrorism data.* Substantively, the key is that military targets are more hardened towards terrorist attacks, even if they are possibly more desirable political targets by some groups. Civilian attacks also dominate in the data, 867 to 174. Since terrorism data are difficult to obtain, it is tempting to use as many variables as one can reasonably obtain on the right-hand-side of regression models. Yet the observed collinearity problem makes this problematic. So we would like to use `TARGET.TYPE` *and* `WEAPON.TYPE` as explanatory variables since they do reflect a graphically observable difference in effect. However, this is deleterious to model in conventional ways, which is why we develop a Dirichlet process for random effects model we accounts for unobserved clumping in the data that come from these overlapping effects.

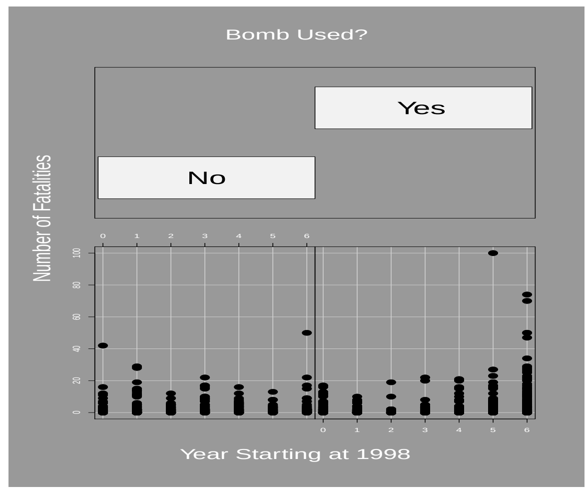
The real underlying goal for terrorists is not actually the results of the physical violence enacted. Rather it is undermining citizens' confidence in their government's ability to protect them. Thus gory attacks against soft targets that receive widespread media coverage are considerable victories for the planners of such attacks. To explore this strategic consideration we focus on the relationship between escalating negative psychological/social impact and the number injured. Injured persons are a fundamentally different kind of casualty effect from fatalities because they wander around the scene of the attack, command medical attention, agree to media interviews, and spend the rest of their life discussing the event with friends

Figure 1: Coplots For Middle-East and Northern Africa Suicide Attacks, 1998-2004. The top two frames compare the year starting with 1998 in the study (x-axis) against the number of fatalities for these years (y-axis). The comparison is split by whether the target was civilian or military (upper-left frame) and whether the a bomb was used (upper-right frame). The bottom two frames compare the levels of psychosocial effect (x-axis) against the number injured. The comparison is split whether there were multiple associated incidents for the single attack (lower-left frame) and whether or not the attacked government has suspicions about the attackers identity.

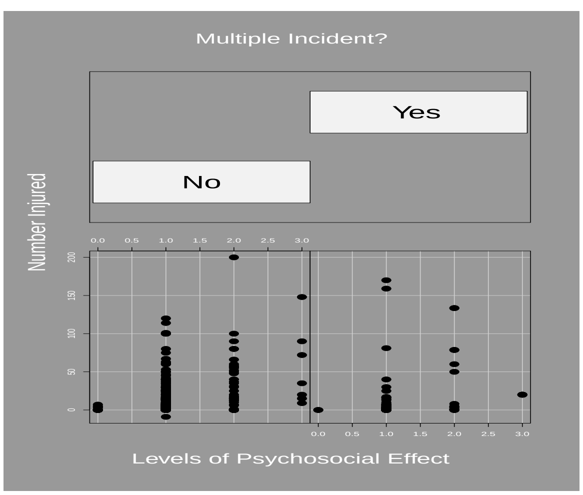
Control: `TARGET.TYPE`



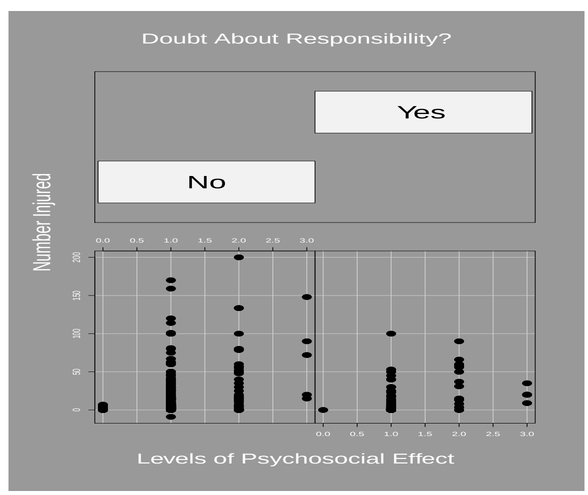
Control: `WEAPON.TYPE`



Control: `MULT.INCIDENT`



Control: `SUSP.UNCONFIRM`



and relatives. Thus while fatalities provide horrific statistics, injuries take on a longer psychological life amongst citizens of the attacked nation. Therefore the lower two panels of

Figure 1 plot the intensity of the psychological effect (x-axis) against the magnitude of the number of injuries (y-axis).

In the lower-left panel of Figure 1 it is easy to see a distinctly different pattern between `PSYCHOSOCIAL` and `NUM.INJUR` for levels of the third dichotomizing variable: whether or not the attack contained multiple associated events or not (`MULT.INCIDENT`). The summary table at right shows that the bulk of the psychological/social impacts are minor, and these are concentrated in the "No" part of the lower-left panel with low levels of injuries. In contrast, the

Summary for `PSYCHOSOCIAL`

Level	none	minor	moderate	major
Count	18	946	66	11

measures of psychological/social impact for multiple incidents are almost all minor and moderate, but show higher average injuries. The third dichotomizing variable in the lower-right panel of Figure 1 is whether or not the government has notable doubts about the identity of the attacking party. Here we see that injuries are generally higher when there is certainty about the attacking group, but also with greater variance. Hard to identify groups may inflict less damage because they are less bold in their planning as a means of reducing the government's detective ability. These two panels show the same challenge as two panels above: there are interesting, and perhaps important, features in the control variables, but the high level of collinearity makes these difficult or impossible to include in the same model as covariates.

This exploratory graphical analysis shows that data on terrorist attacks provide some special challenges. First, the level of measurement is usually low, with mostly categorical, and some purely qualitative variables. Second, as shown in Figure 1, there are collinear relationships in combinations of variables at specific values that can easily be averaged over, and therefore missed, by standard regression-style models. Third, correlation between these variables is almost certainly due to additional factors. Finally, these datasets are almost always missing key variables about the *intentions*, *strategies*, and even *failings* of the terrorist groups. If these groups are heterogeneous in such respects, and in their effectiveness, then we would expect subtle distinctions as seen in Figure 1. So it is with this appreciation for the difficulties inherent in the data-analytic understanding of terrorism, that move us to describing a Bayesian nonparametric setup that is intended to make finer grouping distinctions through Dirichlet priors on random effects to capture latent variability

4 Generalized Linear Models and Random Effects

We need a model structure that accounts for the data problems discussed in the last section, particularly the issue of latent groupings based on unobserved relationships between cases. While it is impossible to provide a reliable coefficient estimate of the magnitude of these effects in the general regression modeling sense since there are no overt indicators, we can develop a model that *accounts for* this phenomenon and thus fits such data better than alternatives that ignore latent groupings. In this section we start with conventional nonlinear regression modeling and build up to our full nonparametric Bayes specification for this purpose.

4.1 Generalized Linear Models

Generalized linear models (GLMs) have enjoyed considerable attention over the years, providing a flexible framework for modeling discrete responses using a variety of error structures. If we have observations that are discrete or categorical, $\mathbf{y} = (\mathbf{y_1}, \ldots, \mathbf{y_n})$, such data can often be assumed to be independent and from a distribution in the exponential family, where the likelihood function has components of the model such as the form of the link function and the type of error structures that result. The classic book by McCullagh and Nelder (1989) describes these models in detail; see also the developments in Dey, Ghosh, and Mallick (2000) or Fahrmeir and Tutz (2001).

4.2 Generalized Linear Mixed Models

A generalized linear *mixed* model (GLMM) is an extension of a GLM that allows random effects, and can give us flexibility in developing a more suitable model when the observations are correlated, or where there may be other underlying phenomena that contribute to the resulting variability. Thus, the GLMM can be specified to accommodate outcome variables conditional on mixtures of possibly correlated random and fixed effects (Breslow and Clayton 1993, Buonaccorsi 1996, Wang, *et al.* 1998, Wolfinger and O'Connell, 1993). Details of such models, covering both statistical inferences and computational methods, can be found in the texts by McCulloch and Searle (2001) and Jiang (2007).

4.3 Dirichlet Process Random Effects

Dirichlet process mixture models were introduced by Ferguson (1973) and Antoniak (1974), and have been the subject of much research since then, most of which described the theoretical character of the process, and some estimation strategies (notably Blackwell and

McQueen). In current applications, a popular strategy is to use Dirichlet process priors in hierarchical models. Work by Escobar and West (1995), MacEachern and Müller (1998), Neal (2000) and Teh *et al.* (2006) developed models and MCMC schemes to fit such models.

Recently, Dorazio, *et al.* (2007) used a GLMDM with a log link for spatial heterogeneity in animal abundance. They proposed an empirical Bayesian approach with the Dirichlet process, instead of the regular assumption of normally distributed random effects, because they argued that for some species, the sources of heterogeneity in abundance is poorly understood or unobservable. They noted that the Dirichlet process prior is robust to errors in model specification and allows spatial heterogeneity in abundance to be specified in a data-adaptive way. Gill and Casella (2009) suggested a GLMDM with an ordered probit link to model political science data, specifically modeling the stress, from public service, of Senate-confirmed political appointees as a reason for their short tenure. For the analysis, a semi-parametric Bayesian approach was adopted, using the Dirichlet process for the random effect.

More recently Kyung, *et al.* (2010a) developed algorithms for estimation of the precision parameter and new MCMC algorithms for a linear mixed Dirichlet random effects models. Also, they showed how to extend the results to a generalized Dirichlet process mixed model with a probit link function, and used a new parameterization of the hierarchical model to derive a Gibbs sampler that more fully exploits the structure of the model and mixes very well. They were also able to establish that the proposed sampler is an improvement, in terms of operator norm and efficiency, over other commonly used algorithms. Kyung, *et al.* (2010b) extended the available sampling schemes to handle other link functions, including logistic and loglinear. Here we apply their logistic sampling scheme to fit our terrorism data.

5 Logistic Mixed Dirichlet Process Models

Let $\mathbf{X}_i$ be covariates associated with the i^{th} observation, $\boldsymbol{\beta}$ be the coefficient vector, and ψ_i be a random effect accounting for subject-specific deviation from the underlying model. Assume that the $Y_i|\boldsymbol{\psi}$ are conditionally independent, each with a density from the exponential family, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$. Then, based on the GLMM notation in McCulloch and Searle (2001), the GLMDM can be expressed as follows. Start with the generalized linear model described above,

$$\begin{aligned} Y_i|\gamma &\overset{\text{ind}}{\sim} f_{Y_i|\gamma}(y_i|\gamma), \quad i = 1, \ldots, n \\ f_{Y_i|\gamma}(y_i|\gamma) &= \exp\left[\{y_i\gamma_i - b(\gamma_i)\}/\xi^2 - c(y_i, \xi)\right]. \end{aligned} \tag{1}$$

where y_i is discrete valued. Here, we know that $\text{E}[Y_i|\gamma] = \mu_i = \partial b(\gamma_i)/\partial\gamma_i$. Using a link function $g(\cdot)$, we can express the transformed mean of Y_i, $\text{E}[Y_i|\gamma]$, as a linear function, and we add a random effect to create the mixed model:

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta} + \psi_i. \tag{2}$$

Here, for the Dirichlet process mixture models, we assume further that

$$\psi_i \sim G, \qquad\qquad G \sim \mathcal{DP}(mG_0), \tag{3}$$

where $\mathcal{DP}$ is the Dirichlet process with base measure G_0 and precision parameter m. The base measure functions as an expected distribution in this setup, and the precision parameter determines the variabilty around this base measure, sometimes called the "cylinder" around it as a visual image. The base measure can even be specified as a normal distribution meaning that m determines the variability of normal distributions around this choice. With this model, we relax the normal assumption, and we provide a richer model to capture more variabilities in the random effects.

Blackwell and MacQueen (1973) prove that for $\psi_1, \ldots, \psi_n$ iid from $G \sim \mathcal{DP}$, the joint distribution of $\boldsymbol{\psi}$ is a product of successive conditional distributions of the form:

$$\psi_i|\psi_1, \ldots, \psi_{i-1}, m \sim \frac{m}{i-1+m}\, g_0(\psi_i) + \frac{1}{i-1+m}\sum_{l=1}^{i-1} \delta(\psi_l = \psi_i) \tag{4}$$

where $\delta()$ denotes the Dirac delta function and $g_0()$ is the density function of base measure. Thus, with n observations from a Dirichlet process with precision parameter m, the marginal distribution of a partition $\{n_1, n_2, \ldots, n_k\}$, where $\sum_j n_j = n, n_j \geq 1$, is given by

$$\pi(n_1, n_2, \ldots, n_k) = \frac{\Gamma(m)}{\Gamma(m+n)} m^k \sum_{\substack{n_j : \sum_j n_j = n \\ n_j \geq 1}} \prod_{j=1}^{k} \Gamma(n_j), \tag{5}$$

which is a normalized probability distribution on the set of all partitions of n observations.

We now define a *partition* C to be a grouping of the sample of size n into k groups, $k = 1, \ldots, n$, and we call these subclusters since the grouping is done nonparametrically rather than on substantive criteria. That is, the partition assigns different parameters across groups and the same parameters within groups; cases are iid only if they are assigned to the same subcluster. Furthermore, the use of "subcluster" is important to distinguish this grouping from substantive clusters since there is no penalty term in the model fit for increasing their number.

Applying Lo (1984) Lemma 2 and Liu (1996) Theorem 1 to formula (4), we can calculate the likelihood function, which by definition is integrated over the random effects, as

$$L(\theta \mid \mathbf{y}) \;\; = \;\; \frac{\Gamma(m)}{\Gamma(m+n)} \sum_{k=1}^{n} m^k \sum_{C:|C|=k} \prod_{j=1}^{k} \Gamma(n_j) \int f(\mathbf{y}_{(j)} \mid \theta,\; \psi_j) \mathrm{d}dG_0(\psi_j),$$

where C defines the subclusters, $\mathbf{y}_{(j)}$ is the vector of y_is that are in subcluster j, and ψ_j is the common parameter for that subcluster. There are $\mathcal{S}_{n,k}$ different subclusters C, the Stirling Number of the Second Kind (Abramowitz and Stegun 1972, 824-825).

The partition C can be represented by an $n \times k$ matrix $\mathbf{A}$ defined by

$$\mathbf{A} = (a_1, a_2, \ldots, a_n)'$$

where each a_i is a $1 \times k$ vector of all zeros except for a 1 in the position indicating which group the observation is from. Thus, $\mathbf{A}$ represents a partition of the sample of size n into k groups, with the column sums giving the subcluster sizes. Note that both the dimension k, and the placement of the 1s, are random, representing the subclustering process. Based on the representation in McCullaugh and Yang (2006), if the partition C has subclusters $\{S_1, \ldots, S_k\}$, then if $i \in S_j$, $\psi_i = \eta_j$ and the random effect can be rewritten as

$$\boldsymbol{\psi} = A\boldsymbol{\eta}, \tag{6}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$ and $\eta_j \overset{iid}{\sim} G_0$ for $j = 1, \ldots, k$.

Usually, in a regular GLMM, the random effects accounting for subject-specific deviation from the underlying model, ψ's, are assumed to be distributed as $N\left(0, \sigma^2_\psi\right)$. In this paper, we assume that $\psi_i \sim \mathcal{DP}(m, N(0, \tau^2))$, $i = 1, \ldots, n$, independent. Thus, for $\boldsymbol{\eta}$, we get $\boldsymbol{\eta} \sim N_k(0, \tau^2 I)$. These ψ's are random effects to capture the issue of latent groupings. Because of the properties of the Dirichlet process that we discussed above, we can develop a rich model with the random effects for the unobserved relationships.

In this particular paper, we only consider models for the binary responses with a logit link function. So we model

$$Y_i \sim \text{ Bernoulli}(p_i), \quad i = 1, \ldots, n$$

where y_i is 1 or 0, and $p_i = \mathrm{E}(Y_i)$ is the probability of a success for the i^{th} observation, and the sampling distribution is

$$f(\mathbf{y}|\mathbf{A}) = \int \prod_{i=1}^{n} \left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)}\right]^{y_i} \left[\frac{1}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)}\right]^{1-y_i} \mathrm{d}G_0(\boldsymbol{\eta}), \tag{7}$$

which typically can only be evaluated numerically. Here the general link function is

$$p_i \ = g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) = (1 + \exp(-\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{A}\boldsymbol{\eta})_i))^{-1}$$

This model can now be fit using a variation of the MCMC scheme of Kyung, *et al.* (2010a), where we use a mixture representation of the logistic distribution (Andrews and Mallow 1974). Details are in Appendix A

6 Simulation Study

We evaluate our sampler through a simulation study. We need to generate outcomes from Bernoulli distributions with random effects that follow the Dirichlet process. To do this we fix K, the true number of subclusters (which is unknown in actual circumstances), then we set the parameter m according to the relation

$$K = \sum_{i=1}^{n} \frac{m}{m+i-1}, \tag{8}$$

where we note that even if $\hat{m}$ is quite variable, there is less variability in $\hat{K} = \sum_{i=1}^{n} \frac{\hat{m}}{\hat{m}+i-1}$. When we integrate over the Dirichlet process, as done algorithmically according to Blackwell and McQueen [1973], the right-hand-side of (8) is the expected number of clusters, given the prior distribution on m. Neal (2000, p.252) shows this as the probability in the limit of a unique table seating, conditional on the previous table seatings, which makes intuitive sense since this expectation depends on individuals sitting at unique tables to start a new (sub)cluster in the algorithm.

Using the GLMDM with the logistic link function, we set the parameters: $n = 100$, $K = 40$, $\tau^2 = 1$, and $\boldsymbol{\beta} = (1, 2, 3)$. Our Dirichlet process for the random effect has precision parameter m and base distribution $G_0 = N(0, \tau^2)$. Setting $K = 40$, yields $m = 24.21$. Independently generated X_1 and X_2 from $N(0, 1)$ used as the fixed design matrix to generate the binary outcome Y. Then the Gibbs sampler was iterated 200 times to get values of m, A, $\boldsymbol{\beta}$, τ^2, η. This procedure was repeated 1000 times saving the last 500 draws as simulations from the posterior.

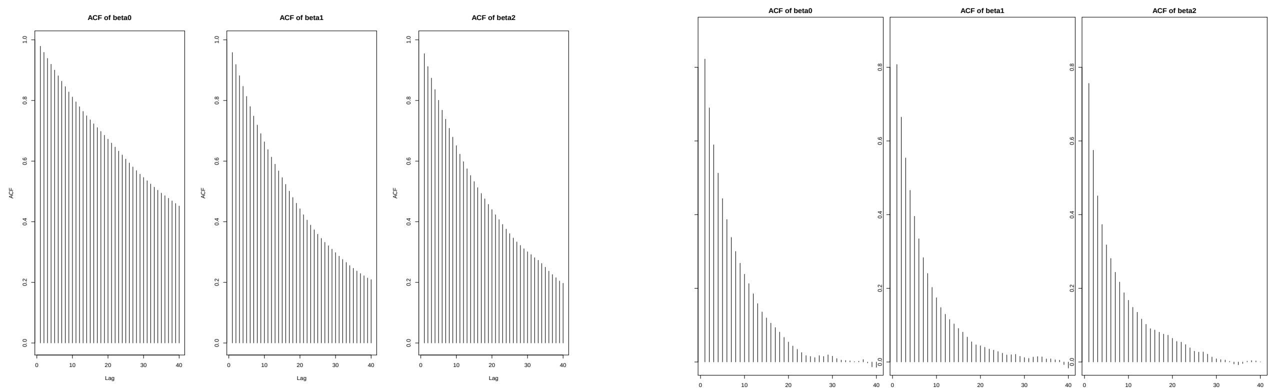
We compare the slice sampler (**Slice**) to the Gibbs sampler with the K-S distribution normal scale mixture (**K-S Mixture**) with the prior distribution of $\boldsymbol{\beta}$ from $\boldsymbol{\beta}|\sigma^2 \sim N(\mu\mathbf{1}, d^*\sigma^2 I)$ and a flat prior on μ, $\mu \sim \pi(\mu) \propto c$. For the estimation of K, we use the posterior mean of m, $\hat{m}$ and calculate $\hat{K}$ by using equation (8). The starting points of $\boldsymbol{\beta}$ come from the maximum likelihood (ML) estimates using iteratively reweighted least squares. All summaries in the tables are posterior means and standard deviations calculated from the empirical draws of the chain in its apparent converged (stationary) distribution.

The numerical summary of this process is given in Table 1. The estimate of $\boldsymbol{\beta}$ with **K-S Mixture** is closer to the true value than those with **Slice**, with smaller standard deviation. To evaluate the convergence of $\boldsymbol{\beta}$, we consider the autocorrelation function (ACF) plots that are given in Figure 2. The Gibbs sampler of β from **Slice** exhibits strong autocorrelation, implying poor mixing. The estimates of K were 43.4722 with standard error 4.0477 from **Slice** and 43.5494 with standard error 3.9708 from **K-S Mixture**. Obviously these turned out to be good estimates of the true $K = 40$.

Table 1: Estimation of the coefficients of the GLMDM with logistic link function and the estimate of K, with true values $K = 40$ and $\boldsymbol{\beta} = (1, 2, 3)$. Standard errors are in parentheses.

Estimation Method	β_0	β_1	β_2	K
Slice	1.4870(0.3139)	1.1512(0.3919)	2.6212(0.4632)	43. 4722(4.0477)
K-S Mixture	1.3410(0.1841)	2.1565(0.1999)	2.7205(0.2245)	43.5494(3.9708)

Figure 2: ACF Plots of $\boldsymbol{\beta}$ for the GLMDM with logistic link. The left panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the slice sampler, and the right panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the K-S/normal mixture sampler.



Slice Sampler K-S Mixture Representation

7 Application of the Model to Data on Suicide Attacks

We apply a GLMDM specification with a logit link function, as described, and for the model parameters with the prior distributions in (13), where $\mathbf{B} = (\mu, 0, \dots, 0)'$, $d^* = 5$ and (a, b) are fixed as (3,2). We ran the Markov chain for 50,000 iterations disposing of the first half. There is no evidence of non-convergence in these runs using standard diagnostic tools (eg. Geweke, Gelman & Rubin, Heidelberger & Welch, and graphical tools; see Gill [2007, Chapter 12]). Table 2 provides results from two approaches: a standard Bayesian random effects logit model (BRELM) with flat priors, and a Dirichlet random effects model, with: posterior means (COEF), posterior standard errors (SE), and 95% highest posterior density (HPD) intervals.

Notice from Table 2 that while there are no changes in sign or statistical reliability for the estimated *coefficients* (posterior means) between the two models, the magnitudes of the effects are uniformly smaller with the enhanced model and four of the BRELM coefficient

Table 2: Global Suicide Attack Data–coefficients, standard errors, and HPD intervals from a Bayesian random effects logit model (BRELM), and from the generalized linear mixed Dirichlet model (GLMDM) using a logit link.

	BRELM				GLMDM Logit			
Coefficient	COEF	SE	95% HPD		COEF	SE	95% HPD	
`Intercept`	-6.457	4.232	-21.605	-3.407	-4.105	0.559	-5.276	-3.079
`YEAR - 1998`	0.303	0.228	0.135	1.137	0.195	0.039	0.121	0.273
`MULT.INCIDENT`	-0.802	0.488	-2.222	-0.142	-0.585	0.221	-1.028	-0.162
`MULTI.PARTY`	-0.945	0.690	-3.289	-0.225	-0.626	0.229	-1.088	-0.189
`SUSP.UNCONFIRM`	-0.109	0.344	-0.928	0.472	-0.061	0.198	-0.455	0.331
`SUCCESSFUL`	-1.035	0.705	-3.308	-0.262	-0.695	0.245	-1.172	-0.210
`ATTACK.TYPE`	0.122	0.135	-0.122	0.466	0.098	0.073	-0.046	0.240
`WEAPON.TYPE`	2.714	1.673	1.346	7.769	1.725	0.320	1.162	2.422
`TARGET.TYPE`	-0.073	0.330	-0.749	0.527	-0.038	0.185	-0.434	0.323
`NUM.FATAL`	-0.019	0.025	-0.085	0.017	-0.013	0.012	-0.036	0.009
`NUM.INJUR`	0.030	0.030	0.010	0.126	0.017	0.004	0.008	0.025
`PSYCHOSOCIAL`	0.824	0.633	0.216	3.044	0.555	0.192	0.188	0.944
`PROPERTY.DAMAGE`	0.439	0.305	0.122	1.406	0.297	0.094	0.114	0.483

estimates are almost twice as large as the corresponding GLMDM coefficient estimates. This indicates that there is extra information in the data detected by the Dirichlet random effect that tends to dampen the size of the effect of these explanatory variables on explaining incidences of terrorist attacks. Specifically, running the standard random effects model would find an *exaggerated* relationship between these explanatory variables and the outcome. In other work (Gill and Casella 2009), we observed that the GLMDM model provided a theoretically important subset of coefficient estimates that were greater in magnitude, so the nonparametric information does not always diminish the resulting effect sizes.

These results are interesting substantively. The year 1998 does stand out relatively speaking (until 2001 in the United States of course, but that case is not in our geographic focus). This is a confirmatory finding since we know that 1998 was exceptional from empirical and journalistic sources. It also appears that multiple coordinated incidents are less associated with suicide attacks (again 9/11/2001 being a notable exception) since `MULT.INCIDENT` is negative and statistically reliable. This is interesting and it suggests that planners of simultaneous linked terrorist events find it harder to manage multiple suicidal agents. `MULTI.PARTY` in the table tells a similar story; coordinated groups of attackers are less likely to plan suicide attacks. This may be because together they have more traditional military resources at hand.

Interestingly, if the attack is successful it is less likely to be a suicide attack. When the delivery vehicle is a person, there are more variables to worry about, in particular whether the individual is sufficiently practiced and indoctrinated. We know from qualitative accounts that fervent nationalism and religious extremism are critical components of this process. Unfortunately for the planners of such attacks, the efficacy of these means is almost certainly negatively correlated with subject age and intelligence.

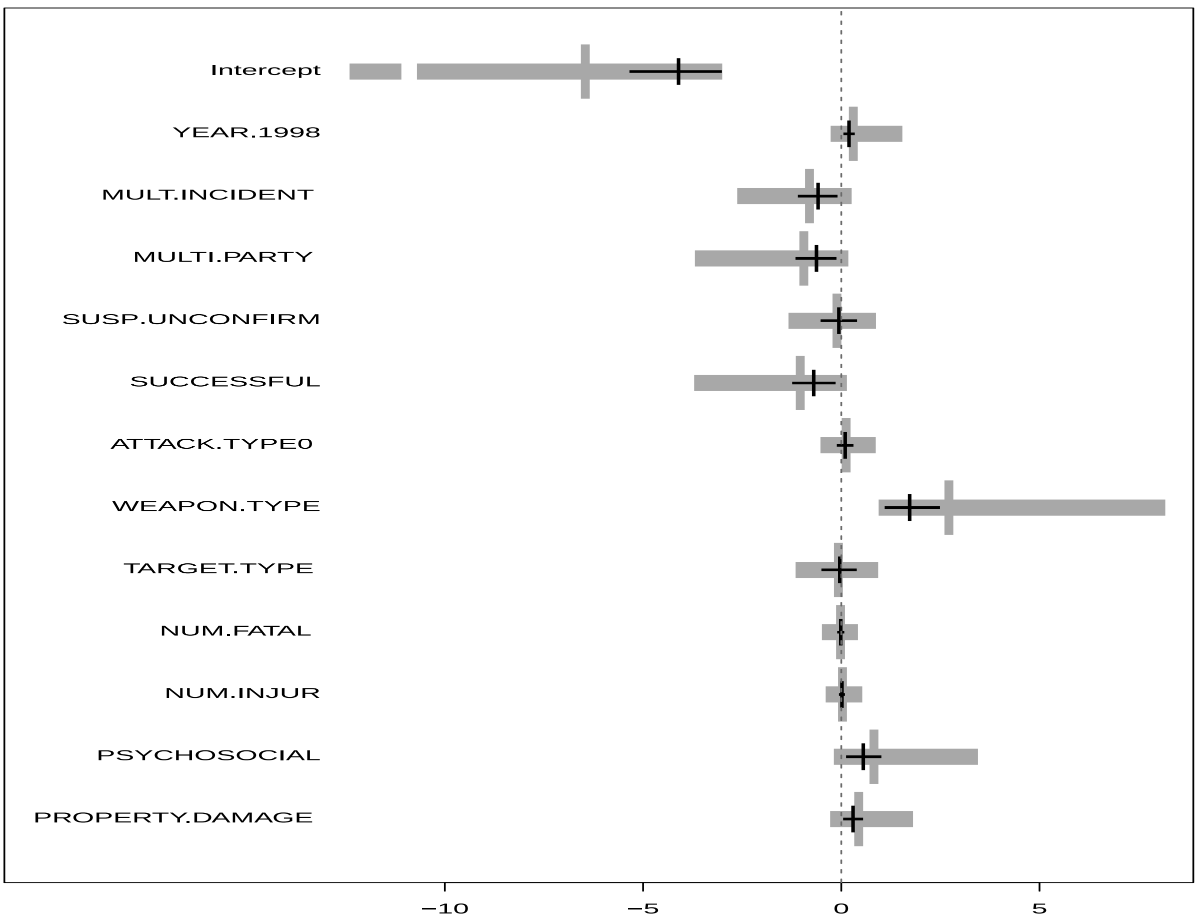
With regard to attack type and weapon type, there is strong evidence that those plotting bomb attacks (generally speaking) are more likely to consider suicide agents to deliver the weapon. Note that the coefficients for `WEAPON.TYPE` is more statistically reliable than the coefficient for `ATTACK.TYPE`, although they agree on direction.

Increased property damage is positively associated with a suicide rather than non-suicide attack. This appears to be related to the terrorists' preference for softer, civilian targets where an equivalent-sized device will produce more damage than against better protected government targets.

In addition to `ATTACK.TYPE`, three other estimated coefficients do not appear to be reliable at standard thresholds in the resulting model: `TARGET.TYPE`, `SUSP.UNCONFIRM`, and `NUM.FATAL` (their 95% HPD intervals cover zero). Since 20% of the target types were military versus civilian, it is somewhat surprising not to find a reliable effect conditioning on the rest of the model. More interestingly, greater fatalities at the event site do not suggest a greater probability of the damage being produced by a suicide attacker. The apparent, but unproven, lack of a relationship here suggests that suicide attacks are more driven by other intended consequences like widespread injuries and the psychological/social effect, both of which show up here as coefficient estimates that have substantial statistical support. This finding, while surprising to some, is consistent with a large proportion of the literature on the intentions of terrorists. Terrorism is most fundamentally aimed at diminishing a national population's confidence in the ability of the government to defend them at home.

Finally, we also note that the GLMDM-logit model has uniformly smaller HPD intervals than the standard GLMM-logit, BRELM. This is consistent with the findings in Kyung, *et al.* (2009) for models with other link functions. The comparison is shown dramatically in Figure 3, where we see that the HPD intervals for the standard random effects logit model (gray) are substantially wider than those for the new method (black). Thus, the richer random effects model is able to remove more extraneous variability, providing tighter HPD intervals. This is evidence that the Dirichlet procedure is capturing additional latent information. So the nonparametric Bayes approach taken here provides more *conservative* estimates of effect sizes (smaller absolute coefficient estimates) *and* gives more accuracy for

Figure 3: 95% HPD Intervals: Standard Logit (Grey) vs. GLMDM Logit (Black). The HPD intervals for the standard random effects logit model are given in gray, and those for the new method are given in black. The vertical bars in this figure denote the posterior means from the two models for each coefficient (same colors).



these estimates (smaller HPD intervals). Thus we have substantially improved the state of data analysis in an area where researchers struggle.

8 Discussion

In this paper we employ a new methodology to solve an old problem. Terrorism has existed since humans first built weapons, and the academic study of terrorism increased dramatically after September 11, 2001. In fact, a non-field specific `jstor.org` search with the word "terrorism" provides article 26,426 citations ranging from the year 1848 to the year 2010. The first citation for the year 2002 occurs at number 16,804, meaning that 36% of the 162 year history occurs after the 154th year (the last 5%). While this is not a complete analysis, it clearly indicates a strong up-swing after 2001.

Unfortunately, the data routinely present formidable challenges to analysis with conventional statistical tools, since such data are always more messy and interrelated than in other applications. The key underlying issue is that a set of diversely organized covert operatives do not cooperate with data collection efforts and the resulting product is often poorly measured.

Our methodological approach uses a variant of nonparametric Bayesian generalized linear models with a logistic link function and Dirichlet process random effects. We developed and verified the properties of new Markov chain Monte Carlo estimation tools for this problem in previous work, noticing that they incorporate latent information such that models fit better in the presence of unexplained heterogeneity in the data. This turns out to be ideal for the empirical analysis of terrorism, which has diverse and subversive operators generating the heterogeneous data. There is evidence from Table 2 that the GLMDM-logit model's superior properties over the analogous standard Bayesian random effects logit model (BRELM) comes from fitting on the latent scale since that is the only fundamental distinction between the two. The inherent binning to subclusters, that is part of our sampler, reaches this improvement by recognizing and modeling this underlying heterogeneity.

What did we learn about Middle East/Northern African terrorism? In general multiple groups working together do not focus on using suicide attackers. Instead, they appear to work in more quasi-military fashion with standard weapons. Surprisingly, the more damaging attacks were more likely to be from non-suicide assailants. This runs counter to media reports, which tend to focus on the goriness of someone blowing themselves up in a place crowded with civilians. Further supporting this goal is the observation that the coefficient for negative psychological/social impact is positive and reliable indicating that this is a predetermination by the plotters of the suicide attack. While damage to property may not be a prime objective, the greater the extent of this damage, the more we would expect to observe that it is from a suicide attack. We also notice a complex relationship between the type of weapon used with the type of assailant, the type of attack, and the success rate. These data may not support clear distinctions of strategy with regard to weapons, except for perhaps bombs which are an integral part of almost all suicide attacks. In summary, the existence of a set of statistically reliable coefficient estimates in the presence of substantial heterogeneity in the studied actors shows that the GLMDM model is able to uncover common patterns by modeling the unexplained variance as a latent grouping.

The important question in an empirical study of terrorism *is not* whether we have explained the complete set of motivations for how and why these deadly attacks are developed, *it is* whether we have added anything to our current knowledge. This is because terrorism is such an important and vexing political problem. In this statistical application we coaxed some revealing information out of the data that increases understanding on the correlates of suicide attacks in the Middle East and Northern Africa. Some of these findings are confirmatory, such as target preference and inter-group cooperation, while others, like the success and damage levels inflicted by suicide attacks relative to non-suicide attacks, is surprising.

The results described above contain prescriptive advice for governmental organizations seeking to reduce the number and effectiveness of intentionally violent events. While successful attacks are less associated with suicide strategies, suicide attackers generally inflict greater property damage, additional injuries, and more importantly, more psychological/social impact on the populace. Therefore countries in this region are justified in their counter-terrorism policies that seek to reduce training, remove motivation, or interdict, suicide attackers, even at the expense of less attention to non-suicide attackers. Strategic terrorism is mainly about attacking the social contract between a people and their government by eroding trust. The factors observed to have a stronger link to suicide attacks are more damaging to this relationship, and therefore deserve more attention from governments' concerned about the morale and support of their citizens.

9 References

Abramowitz, M. and Stegun, I. A. (1972). "Stirling Numbers of the Second Kind." Section 24.1.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York: Dover, 824-825.

Andrews D. F. and Mallows, C. L. (1974). "Scale Mixtures of Normal Distributions." *Journal of the Royal Statistical Society, Series B* **36**, 99-102.

Antoniak, Charles E. (1974). "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems." *Annals of Statistics* **2**, 1152-1174.

Arce, Daniel G. and Todd Sandler. (2007). "Terrorist Signaling and the Value of Intelligence." *British Journal of Political Science* **37**, 573-586.

Balakrishnan, N. (1992). *Handbook of the Logistic Distribution.* CRC Press.

Blackwell, D. and MacQueen, J. B. (1973). "Discreteness of Ferguson Selections." *Annals of Statistics* **1**, 365-358.

Breslow, N. E., and D. G. Clayton. (1993). "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* **88**, 9-25.

Buonaccorsi, J. P. (1996). "Measurement Error in the Response in the General Linear Model." *Journal of the American Statistical Association* **91**, 633-642.

Bueno de Mesquita, Ethan. (2005). "Conciliation, Counterterrorism, and Patterns of Terrorist Violence." *International Organization* **59**, 145-176

Bueno de Mesquita, Ethan and Eric S. Dickson. (2007). "The Propaganda of the Deed: Terrorism, Counterterrorism, and Mobilization." *American Journal of Political Science* **51**, 364-381.

Carley, Kathleen M. (2003). "Dynamic Network Analysis." In *Dynamic Social Network Modeling and Analysis* R. Breiger, K. Carley, and P. Pattison eds. Committee on Human Factors. National Research Council. Washington D.c. The National Academies Press.

Carley, Kathleen M. (2004). "Estimating Vulnerabilities in Large Covert Networks Using Multi-Level Data." *Proceedings of the NAACSOS Conference*, Pittsburgh, PA.

Carley, Kathleen. (2006). "A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action." *Pre Proceedings, Visualizing Network Information* IST-063/RWS-010. Royal Danish Defence College, Copenhagen.

Chai, Sun-Ki. 1993 "An Organizational Economics Theory of Antigovernment Violence." *Comparative Politics* **26**, 99-110.

Crenshaw, Martha. (1981). "The Causes of Terrorism." *Comparative Politics* **13**, 379-399.

Damien, P., Wakefield, J., and Walker, S. (1999). "Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables." *Journal of the Royal Statistical Society, Series B* **61**, 331-344.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag New York Inc.

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective.* New York: Marcel Dekker.

Dorazio, R. M., Mukherjee, B., Zhang, L., Ghosh, M., Jelks, H. L. and Jordan, F. (2007). "Modelling Unobserved Sources of Heterogeneity in Animal Abundance Using a Dirichlet Process Prior." *Biometrics*, Online Publication August 3, 2007.

Enders, Walter. (2007). "Terrorism: An Empirical Analysis." In *Handbook of Defense Economics*Volume 2. T.Sandler and K. Hartley eds. Elsevier.

Enders, Walter and Todd Sandler. (1995). "Terrorism: Theory and Applications." In *Handbook of Defense Economics* Volume 1. K. Hartley and T. Sandler eds. Elsevier.

Escobar, M. D. and West, M. (1995). "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association* **90**, 577-588.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Second Edition. New York: Springer.

Ferguson, T. S. (1973). "A Bayesian analysis of Some Nonparametric Problems." *Annals of Statistics* **1**, 209-230.

Gill, J. (2007). *Bayesian Methods for the Social and Behavioral Sciences.* New York: Chapman & Hall.

Gill, J. and Casella, G. (2009). "Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation." *Journal of the American Statistical Association* **104**, 453-464.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications.* New York: Springer-Verlag.

Krebs, Valis E. (2002). "Mapping Terrorist Cells." *Connections* **24**, 43-52.

Kydd, Andrew and Barbara F.Walter. (2002). "Sabotaging the Peace: The Politics of Extreme Violence." *International Organization* **56**, 263-296.

Kyung, M, Gill, J. and Casella, G. (2009). "Characterizing the variance improvement in linear Dirichlet random effects models." *Statistics and Probability Letters* **79** 2343-2350.

Kyung, M, Gill, J. and Casella, G. (2010a). "Sampling Schemes for Generalized Linear Dirichlet Process Random Effects Models." *Statistical Methods and Applications*, forthcoming 2011.

Kyung, M, Gill, J,, and Casella G. (2010b). "Estimation in Dirichlet Random Effects Models." *Annals of Statistics* **38**, 979-1009.

LaFree, Gary, and Laura Dugan. (2008). Global Terrorism Database II, 1998-2004 [Computer file]. ICPSR22600-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Levy, Sheldon, Graham, Hugh Davis, and Gurr, Ted Robert. (1969). "A 150-year study of political violence in the United States." In Violence in America: Historical and Comparative Perspectives (Official Report), pp.81-91. Washington: National Commission on the Causes and Prevention of Violence.

Li, Quan and Drew Schaub. (2004). "Economic Globalization and Transnational Terrorist Incidents." *Journal of Conflict Resolution* **48**, 230-258.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Second Edition. New York: John Wiley & Sons.

Liu, J. S. (1996). "Nonparametric Hierarchical Bayes Via Sequential Imputations." *Annals of Statistics* **24**, 911-930.

Lo, A. Y. (1984). "On A Class of Bayesian Nonparametric Estimates: I. Density Estimates." *Annals of Statistics* **12**, 351-357.

MacEachern, S. N. and Müller, P. (1998). "Estimating Mixture of Dirichlet Process Model." *Journal of Computational and Graphical Statistics* **7**, 223-238.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Edition. New York: Chapman & Hall.

McCullagh, P. and Yang, J. (2006). "Stochastic Classification Models." *International Congress of Mathematicians* **III** 669-686.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* New York: John Wiley & Sons.

Mickolus, Edward F. (1982). International terrorism: Attributes of terrorist events, 1968-1977 (ITERATE 2). Inter-university Consortium for Political and Social Research (Ann Arbor, Mich.) ISBN 089138927X.

Mickolus, Edward. (2006). International terrorism: Attributes of terrorist events, 1968-2002. Dunn Loring, VA: Vinyard Software.

Moon, Il-Chul and Kathleen M. Carley. (2007). "Modeling and Simulation of Terrorist Networks in Social and Geo-spatial Dimensions." *IEEE Intelligent Systems: Special Issue on Social Computing*, **22**, 40-49.

Neal, R. M. (2000). "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics* **9**, 249-265.

Pape, Robert A. (2006). *Dying to Win. Why Suicide Terrorists Do It.* London: Gibson Square Books.

Rothenberg, Robert. (2002). "From Whole Cloth: Making up the Terrorist Network." *Connections* **23**, 36-42.

Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Sandler, Todd and Daniel Arce (2003). "Terrorism and Game Theory." *Simulation & Gaming* **34**, 319-37.

Siqueira, Kevin and Todd Sandler. (2006). "Terrorist Versus the Government: Strategic Interaction, Support, and Sponsorship." *Journal of Conflict Resolution* **50**, 878-898.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006). "Hierarchical Dirichlet Processes." *Journal of the American Statistical Association* **101**, 1566-1581.

Tsvetovat, Maksim and Kathleen M. Carley. (2005). "Structural Knowledge and Success of Anti-terrorist Activity:The Downside of Structural Equivalence." *Journal of Social Structure* **6**.

Tsvetovat, Maksim and Kathleen M. Carley, (2006). "Improving Effectiveness of Communications Sampling of Covert Networks." *E-Social Science Conference* Manchester, UK.

Wang, N., X. Lin, R. G. Gutierrez, and R. J. Carroll. (1998). "Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models." *Journal of the American Statistical Association* **93**, 249-261.

West, M. (1987). "On Scale Mixtures of Normal Distributions." *Biometrika* **74**, 646-648.

Wolfinger, R. and M. O'Connell. (1993). "Generalized Linear Mixed Models: A Pseudolikelihood Approach." *Journal of Statistical Computation and Simulation* **48**, 233-243.

A Appendix: Fitting the Logistic/Dirichlet Model

A.1 The Logistic Distribution as a Mixture of Normals

The logistic distribution can be represented as a mixture of normals; see Andrews and Mallows (1974) or West (1987), which leads us to a very efficient algorithm that can be put into a Gibbs sampler. If Z has a standard normal distribution, and $V/2$ has the asymptotic distribution of the Kolmogorov distance statistic and is independent of Z, then $Y = Z/V$ is logistic. From the identities in Andrews and Mallows (1974) (see also Theorem 10.2.1 in Balakrishnan 1992), for

$$f_X(x) = 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2 x e^{-2\alpha^2 x^2} \quad x \geq 0, \tag{9}$$

the Kolmogorov-Smirnov (K-S) density function (see Devroye 1986), we can write

$$f_Y(y) \;=\; \int_0^{\infty}\left[\frac{1}{\sqrt{2\pi}}\frac{1}{2x}\exp\left\{-\frac{1}{2}\left(\frac{y}{2x}\right)^2\right\}\right] f_X(x)\mathrm{d}x = \frac{e^{-y}}{(1+e^{-y})^2}, \tag{10}$$

where the density in square brackets is $N(0, 1/4x^2)$ and the result is the density function of a logistic distribution with mean 0 and variance $\frac{\pi^2}{3}$. Therefore, $Y \sim \mathcal{L}\left(0, \frac{\pi^2}{3}\right)$, where $\mathcal{L}()$ denotes the logistic distribution.

Starting from the logistic distribution, we can model binary responses with a logit link function through a latent variable W_i such that

$$W_i = X_i\boldsymbol{\beta} + \psi_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{L}\left(0, \frac{\pi^2}{3}\sigma^2\right), \tag{11}$$

with $y_i = 1 \quad \text{if } W_i > 0$ and $y_i = 0 \quad \text{if } W_i \le 0$, for $i = 1, \ldots, n$. It can be shown that Y_i are independent Bernoulli random variables with $p_i = [1+\exp(-\mathbf{X}_i\boldsymbol{\beta}-(\mathbf{A}\boldsymbol{\eta})_i)]^{-1}$, the probability of success, and without loss of generality we fix $\sigma = 1$. With the mixture representation (10) an equivalent representation uses a normal latent variable, resulting in the likelihood function

$$\begin{aligned} L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}|\mathbf{A}, \mathbf{y}, \sigma^2) &= \prod_{i=1}^{n} \{I(U_i > 0)I(y_i = 1) + I(U_i \le 0)I(y_i = 0)\} \\ &\quad \times \int_0^\infty \left(\frac{1}{2\pi\sigma^2(2\xi)^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2(2\xi)^2}|\mathbf{U}-\mathbf{X}\boldsymbol{\beta}-\mathbf{A}\boldsymbol{\eta}|^2} \\ &\quad \times 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}\mathrm{d}\xi \left(\frac{1}{2\pi\tau^2}\right)^{k/2} e^{-\frac{1}{2\tau^2}|\boldsymbol{\eta}|^2}, \end{aligned} \tag{12}$$

where $\mathbf{U} = (U_1, \ldots, U_n)$, and U_i is the truncated normal variable such that

$$U_i = X_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2(2\xi)^2\right),$$

and $y_i = 1$ if $U_i > 0$ and $y_i = 0$ if $U_i \le 0$, for $i = 1, \ldots, n$.

A.2 Sampling Schemes

An overview of the general sampling scheme is as follows. We identify three groups of parameters:

(*i*) m, the precision parameter of the Dirichlet process,

(*ii*) $\mathbf{A}$, the indicator matrix of the partition defining the subclusters, and

(*iii*) $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$, the model parameters.

In the Gibbs sampler we iterate between these three groups until convergence:

1. Conditional on m and A, generate $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)|\mathbf{A}, m$;
2. Conditional on $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$ and m, generate $\mathbf{A}$, a new partition matrix;
3. Conditional on $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$ and $\mathbf{A}$, generate m, the new precision parameter.

The model parameters are given the following prior distributions,

$$\begin{aligned} \beta|\mu, \sigma^2 &\sim N(\mathbf{B}, d^*\sigma^2 I) \\ \mu &\sim \pi(\mu) \propto c \quad \text{a flat prior with constant } c \\ \tau^2 &\sim \text{Inverted Gamma}(a, b), \end{aligned} \tag{13}$$

where $\mathbf{B} = (\mu, 0, 0)'$, $d^* > 1$ and (a, b) are fixed such that the inverse gamma is diffuse ($a = 1$, b very small). We can either fix σ^2 or put a prior on it and estimate it in the hierarchical model with priors; here we will fix a value for σ^2. For the base measure of the Dirichlet process, we assume a normal distribution with mean 0 and variance τ^2, $N(0, \tau^2)$.

A.2.1 Generating the Logistic Parameters

Given ξ, for fixed m and $\mathbf{A}$, a Gibbs sampler of $(\mu, \boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U})$ is

$$\begin{aligned}
\boldsymbol{\eta}|\mu, \boldsymbol{\beta}, \tau^2, \mathbf{U}, \mathbf{A}, \mathbf{y}, \sigma^2 &\sim N_k\left(\frac{1}{\sigma^2(2\xi)^2}\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2(2\xi)^2}\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'(\mathbf{U} - \mathbf{X}\boldsymbol{\beta}),\right. \\
&\qquad\qquad \left.\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2(2\xi)^2}\mathbf{A}'\mathbf{A}\right)^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
\mu|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{A}, \mathbf{y}, \sigma^2 &\sim N\left(\beta_0, d^*\sigma^2\right) \\
\boldsymbol{\beta}|\mu, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{A}, \mathbf{y}, \sigma^2 &\sim N_p\left(\left(\frac{1}{d^*}I + \frac{1}{(2\xi)^2}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{d^*}\mathbf{B} + \frac{1}{(2\xi)^2}X'(\mathbf{U} - \mathbf{A}\boldsymbol{\eta})\right),\right. \\
&\qquad\qquad \left.\sigma^2\left(\frac{1}{d^*}I + \frac{1}{(2\xi)^2}\mathbf{X}'\mathbf{X}\right)^{-1}\right)
\end{aligned}$$

$$\tau^2|\mu, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{A}, \mathbf{y}, \sigma^2 \sim \text{Inverted Gamma}\left(\frac{k}{2} + a, \frac{1}{2}|\boldsymbol{\eta}|^2 + b\right)$$

$$U_i|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{A}, y_i, \sigma^2 \sim \begin{cases} N\left(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i, \sigma^2(2\xi)^2\right) I(U_i > 0) & \text{if } y_i = 1 \\ N\left(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i, \sigma^2(2\xi)^2\right) I(U_i \le 0) & \text{if } y_i = 0 \end{cases}$$

where $\mathbf{B} = (\mu, 0, 0)'$ and μ is from posterior density. Then we update ξ from

$$\xi|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{A}, \mathbf{y} \sim \left(\frac{1}{(2\xi)^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2(2\xi)^2}|\mathbf{U}-\mathbf{X}\boldsymbol{\beta}-\mathbf{A}\boldsymbol{\eta}|^2} 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}.$$

The conditional posterior density of ξ is the product of a inverted gamma with parameters $\frac{\alpha}{2} - 1$ and $-\frac{1}{8\sigma^2}|\mathbf{U} - -\mathbf{X}\boldsymbol{\beta} - \mathbf{A}\boldsymbol{\eta}|^2$, and the infinite sum of the sequence $(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}$. To generate samples from this target density, we consider the alternating series method that is proposed by Devroye (1986). Based on his notation, we take

$$\begin{aligned}
ch(\xi) &= 8\left(\frac{1}{\xi^2}\right)^{n/2} e^{-\frac{1}{8\sigma^2\xi^2}|\mathbf{U}-\mathbf{X}\boldsymbol{\beta}-\mathbf{A}\boldsymbol{\eta}|^2}\xi e^{-2\xi^2} \\
a_n(\xi) &= (\alpha+1)^2 e^{-2\xi^2\{(\alpha+1)^2-1\}}
\end{aligned}$$

Here, we need to generate sample from $h(\xi)$, and we use accept-reject sampling with candidate $g(\xi^*) = 2e^{-2\xi^*}$, the exponential distribution with $\lambda = 2$, where $\xi^* = \xi^2$. Then we follow Devroye's method.

A.2.2 Dirichlet Process Parameters

In generating the Dirichlet process parameters we first update the partition matrix $\mathbf{A}$ and then the precision parameter m. We use a Metropolis-Hastings (M-H) algorithm with a candidate taken from a multinomial/Dirichlet. This produces a Gibbs sampler that converges faster than the popular "stick-breaking" algorithm. (See Kyung *et al.* 2010a for details).

Given the model parameters at iteration t, $\theta^{(t)} = (\boldsymbol{\eta}^{(t)}, \boldsymbol{\beta}^{(t)}, \tau^{2(t)})$, we generate

$$\begin{aligned} \mathbf{q}^{(t+1)} &= (q_1^{(t+1)}, \ldots q_n^{(t+1)}) \sim \text{Dirichlet}(n_1^{(t)} + 1, \ldots, n_k^{(t)} + 1, 1, \ldots, 1) \\ \mathbf{A}^{(t+1)} &\sim P(\mathbf{A}')f(\mathbf{y}|\theta^{(t+1)}, \mathbf{A}') \binom{n}{n_1 \ \cdots \ n_{k'}} \prod_{j=1}^{k'} [q_j^{(t+1)}]^{n'_j} \end{aligned} \tag{14}$$

where $n_j > 0$, $n_1 + \cdots + n_{k'} = n$.

Generating $\mathbf{q}^{(t+1)}$ is easy. To generate $\mathbf{A}^{(t+1)}$ we first generate a candidate that is an $n \times n$ matrix where each row is an independent multinomial with probabilities $\mathbf{q}^{(t+1)}$, and the effective dimension of the matrix, the size of the partition, $k^{(t+1)}$, are the non-zero column sums. Deleting the columns with column sum zero is a marginalization of the multinomial distribution. The probability of the candidate is given by

$$P\left(\mathbf{A}^{(t+1)}\right) = \frac{\Gamma(n)}{\Gamma(n - k^{(t+1)} + 1)} \frac{\Gamma(n_{k^{(t+1)}}^{(t+1)} + n - k^{(t+1)} + 1)}{\Gamma(2n)} \prod_{j=1}^{k^{(t+1)}-1} \Gamma(n_j^{(t+1)} + 1)$$

and a Metropolis-Hastings step is then done.

A.2.3 Generating the Precision Parameter

Rather than estimate m conventionally, a better strategy is to include m in the Gibbs sampler, as the marginal maximum likelihood estimate can be very unstable (Kyung, *et al.* 2010a). Conditional on $(\theta, \mathbf{A}, \mathbf{y})$, and k, with a prior $g(m)$ we get the posterior density

$$\pi(m \mid \theta, \mathbf{A}, \mathbf{y}) = \frac{\frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k}{\int_0^\infty \frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k \mathrm{d}m}, \tag{15}$$

where $\int \pi(m \mid \theta, \mathbf{A}, \mathbf{y}) \, dm < \infty$ as long as the exponent of m is positive. Note how far removed m is from the data, as the posterior only depends on the number of groups k. We consider a gamma distribution as a prior, $g(m) = m^{a-1}e^{-m/b}/\Gamma(a)b^a$, and generate m using an M-H algorithm with another gamma density as a candidate.