University of California, San Diego
May, 2011

New Findings from Terrorism Data:

Dirichlet Process Random Effects Models for Latent Groups

Minjung Kyung
Department of Statistics
University of Florida

Jeff Gill
Center for Applied Statistics
Washington University

George Casella
Department of Statistics
University of Florida

Supported by NSF Grants: SES-0958982 & SES-0959054.

Introduction
Terrorism Data

- The analysis of data on terrorists and terrorist attacks is difficult.

- Typical data are
  - Observed public events
  - Not including failed attacks
- Classified government information
- Terrorists seek to strategically hide information

Introduction
Terrorism Data

- Data collection can even be physically dangerous for the researcher

- Terrorism is an important problem
  - It affects personal safety
  - Internal government policies
  - Public perception
  - Relations between nations

Introduction
Overview of the Talk

▶ Background about terrorism data sets	Problems with the data
▶ Logistic Random Effects Models	An Introduction to Modelling Random Effects
▶ Fitting the Models	Markov Chain Monte Carlo
▶ Analysis of a Terrorism Data Set	What the Covariates Explain
▶ Conclusions	What We Learned

Background On Terrorism Data
Types of Data Available

- Most of the datasets focus on *incidents*
- Data from an observed violent attack and covariates such as
  - Responsible group
  - Target characteristics
  - The extent of casualties and damage.
- Humans in terrorist networks conceal their identities and intentions
- Therefore there is a lack of informative covariates

Background On Terrorism Data
Major Databases

- University of Maryland (START)
- US Homeland Security Agency
- International Terrorism: Attributes of Terrorist Events (ITERATE)
  - Records transnational terrorist incidents
- International Policy Institute for Counter-Terrorism in Herzlia, Israel
  - Detailed online database of terrorist attacks in Israel
- The Global Terrorism Database (GTD)
  - Information on global terrorist events starting from 1970
  - We used this one

Background On Terrorism Data
Previous Findings

- Extremist groups often $\uparrow$ terrorist activity after government concessions
  - Anecdotal evidence rather than statistical data analysis

- Statistical models try to forecast the occurrence of terrorists incidents
  - Limited results

- Networks of terrorist and terrorist organizations
  - Tend to be cellular and independent
  - Rather than hierarchical and connected

Background On Terrorism Data
Data Problems

- Not much success in building standard regression models
  - The data are, in general, poorly measured
  - Categorical variables with large variability

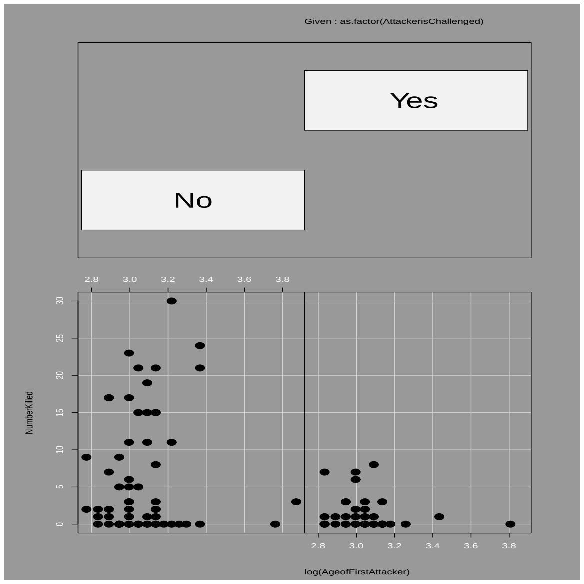
- Huge Problem: The terrorists under study
  - Are deliberately trying to prevent accurate data from being collected

- The statistician has a difficult task in creating meaningful models.

Background On Terrorism Data

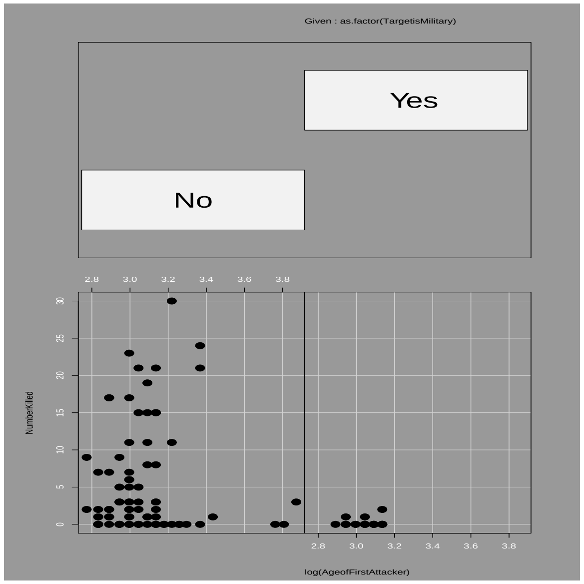
Data Quality Example: Attacks in Israel

Attacker is Challenged



- Y-axis = Number of Casualties

Target is Military

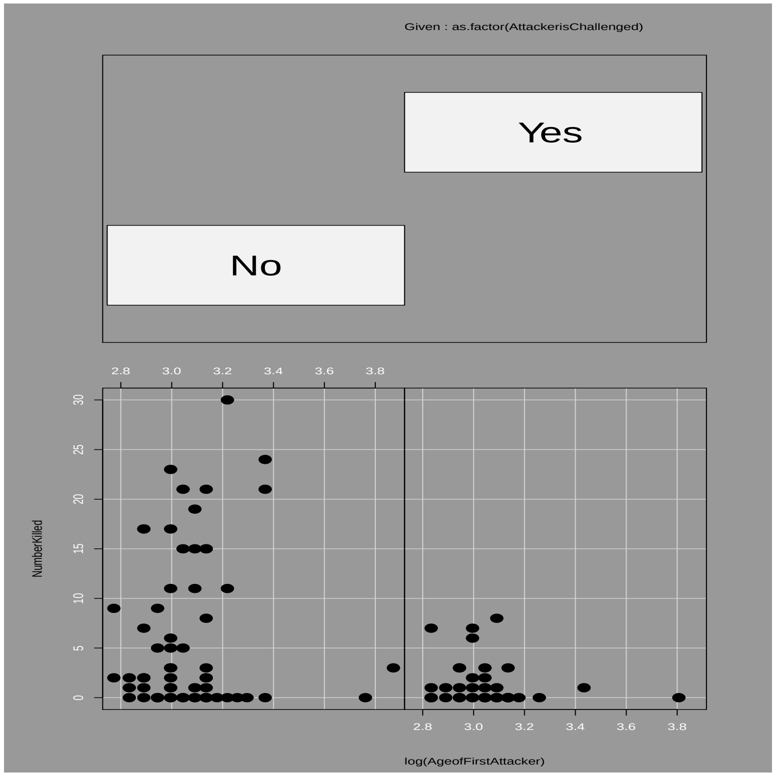


- X-axis = Age of Attacker

- Consider some details

Background On Terrorism Data
Attacks in Israel – Attacker is Challenged

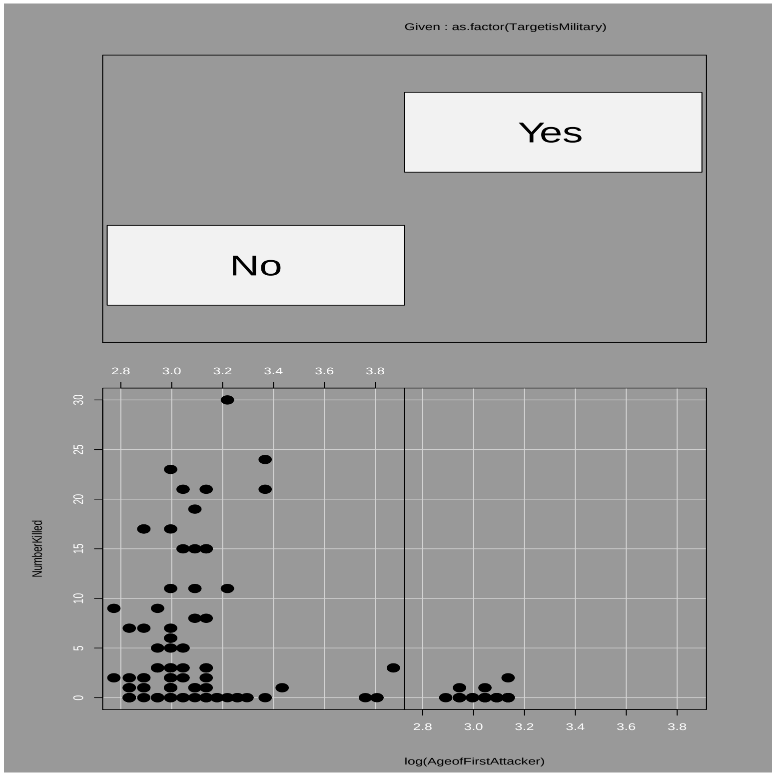
Attacker is Challenged



- Difference in the distribution of fatalities between the plots.
- The attack is less deadly if the attacker is challenged

Background On Terrorism Data
Attacks in Israel – Target is Military

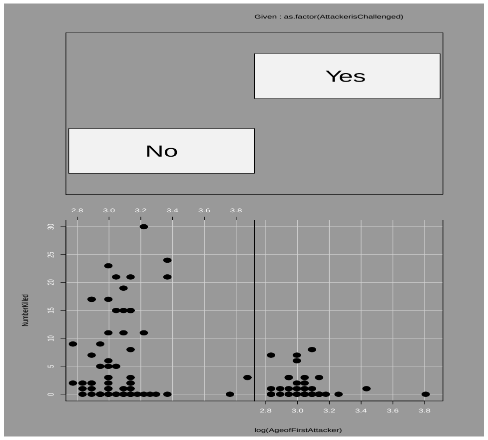
Target is Military



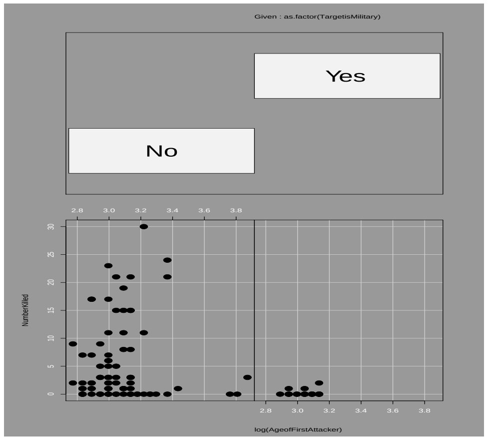
- Much higher level of fatalities for non-military attacks
- These terrorist groups prefer civilian targets.

Background On Terrorism Data
Attacks in Israel – Confounding

Attacker is Challenged



Target is Military



- There is confounding
- Suicide bombers attacking civilian targets are rarely challenged
- So there is little to distinguish between these two plots.

Background On Terrorism Data
Challenges from the Data

- Data on terrorist attacks have special challenges
  - Coarse measurements; many categorical and qualitative variables.
  - Important variables missing: *intentions* and *strategies* of the terrorists
- Assume: Observed events resemble events that failed or were cancelled
- These difficulties in the data-analytic understanding of terrorism
  - Lead us to a Bayesian nonparametric setup
  - Use a rich error structure with Dirichlet process priors
  - Attempt to capture latent variability

———But First———
Here is the Big Picture

▶ Usual Random Effects Model

$$\mathbf{Y}|\psi \sim N(X\beta + \psi, \sigma^2 I), \quad \psi_i \sim N(0, \tau^2)$$

▷ Subject-specific random effect

▶ Dirichlet Process Random Effects Model

$$\mathbf{Y}|\psi \sim N(X\beta + \psi, \sigma^2 I), \quad \psi_i \sim \mathcal{DP}(m, N(0, \tau^2))$$

▶ Results in

▷ Fewer Assumptions

▷ Better Estimates

▷ Shorter Credible Intervals

A Dirichlet Process Random Effects Model
Estimating the Dirichlet Process Parameters

- A general random effects Dirichlet Process model can be written

$$(Y_1, \ldots, Y_n) \sim f(y_1, \ldots, y_n \mid \theta, \psi_1, \ldots, \psi_n) = \prod_i f(y_i|\theta, \psi_i)$$

  - $\psi_1, \ldots, \psi_n$ iid from $G \sim \mathcal{DP}$
  - $\mathcal{DP}$ is the Dirichlet Process
    - Base measure ϕ_0 and precision parameter m
  - The vector θ contains all model parameters

- Blackwell and MacQueen (1973) proved

$$\psi_i|\psi_1, \ldots, \psi_{i-1} \sim \frac{m}{i-1+m}\,\phi_0(\psi_i) + \frac{1}{i-1+m}\sum_{l=1}^{i-1} \delta(\psi_l = \psi_i)$$

  - Where δ denotes the Dirac delta function.

Some Distributional Structure

- Freedman (1963), Ferguson (1973, 1974) and Antoniak (1974)
  - ▷ Dirichlet process prior for nonparametric G
  - ▷ Random probability measure on the space of all measures.

- Notation
  - ▷ G_0, a base distribution (finite non-null measure)
  - ▷ $m > 0$, a precision parameter (finite and non-negative scalar)
    - ▷ Gives spread of distributions around G_0,
  - ▷ Prior specification $G \sim \mathcal{DP}(m, G_0) \in \mathcal{P}$.

- For *any* finite partition of the parameter space, $\{B_1, \ldots, B_K\}$,

$$(G(B_1), \ldots, G(B_K)) \sim \mathcal{D}\left(mG_0(B_1), \ldots, mG_0(B_K)\right),$$

A Mixed Dirichlet Process Random Effects Model
Likelihood Function

- The likelihood function is integrated over the random effects

$$L(\theta \mid \mathbf{y}) = \int f(y_1, \ldots, y_n \mid \theta, \psi_1, \ldots, \psi_n) \pi(\psi_1, \ldots, \psi_n) \, d\psi_1 \cdots d\psi_n$$

- From Lo (1984 Annals) Lemma 2 and Liu (1996 Annals)

$$L(\theta \mid \mathbf{y}) = \frac{\Gamma(m)}{\Gamma(m+n)} \sum_{k=1}^{n} m^k \left[\sum_{C:|C|=k} \prod_{j=1}^{k} \Gamma(n_j) \int f(\mathbf{y}_{(j)} \mid \theta, \ \psi_j) \phi_0(\psi_j) \, d\psi_j \right],$$

  - The partition C defines the subclusters
  - $\mathbf{y}_{(j)}$ is the vector of y_is in subcluster j
  - ψ_j is the common parameter for that subcluster

How Is This Nonparametric?

- These models stipulate uncertainty at the level of distribution functions
  - ▷ Allows for infinite dimensional alternatives
  - ▷ Thus a nonparametric approach
- If $\{f(y|\boldsymbol{\psi}){:}\boldsymbol{\psi} \in (\boldsymbol{\Psi} \subset \Re^d)\}$ is a parametric family of distributions
  - ▷ Construct the family of distributions $\mathcal{F} = \{F_G : G \in \mathcal{P}\}$:

$$f(y|G) = \int f(y|\boldsymbol{\psi})dG(\boldsymbol{\psi}).$$

- Now $\mathcal{F}$ becomes a nonparametric family of mixtures.
- G remains random because it comes from a definable measure
  - ▷ Dirichlet process

Logistic Regression with Random Effects
Setup

- We begin with the model

$$Y_i \sim \text{ Bernoulli}(p(\mathbf{X}_i)), \quad i = 1, \ldots, n$$

where

  - $y_i = \begin{cases} 1 & \text{if the attack is a suicide attack} \\ 0 & \text{if the attack is not a suicide attack} \end{cases}$
  - $p(\mathbf{X}_i) = \mathrm{E}(Y_i|\mathbf{X}_i)$ is the probability of a success
  - $\mathbf{X}_i$ = covariates associated with the i^{th} observation

- Extra variation is modeled with a random effect

$$\text{logit}(p(\mathbf{X}_i)) = \frac{\log(p(\mathbf{X}_i))}{1 - \log(p(\mathbf{X}_i)} = \mathbf{X}_i\boldsymbol{\beta} + \phi_i,$$

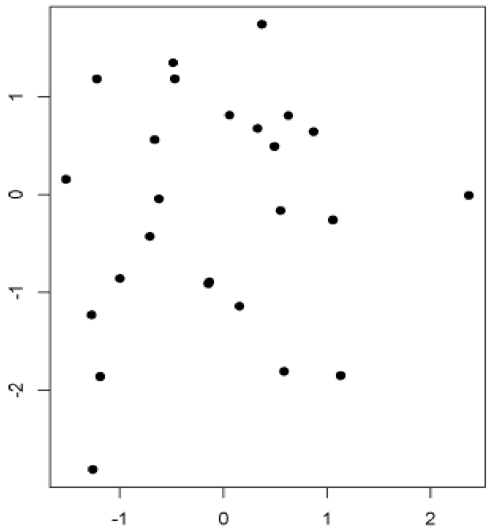
where ϕ_i is a random variable to model extra unexplained variation.

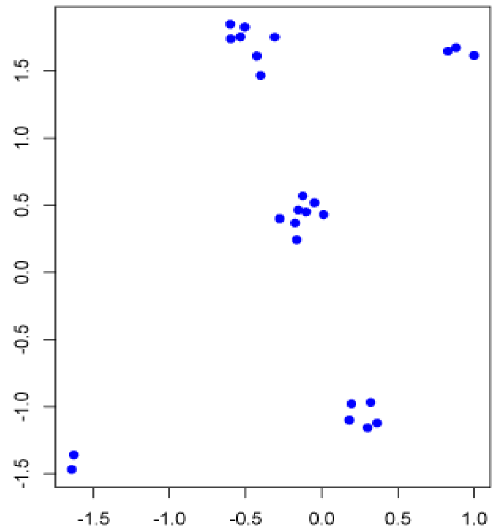
Logistic Regression with Random Effects
Choices for Random Effect Models

- The typical random effect model

$$\mathrm{logit}(p(\mathbf{X}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \phi_i,$$

  - Will often model ϕ_i with a normal distribution
  - We use the alternative ψ_i from a Dirichlet process





- Notice the clustering
- Models extra variability

Logistic Regression with Random Effects
The Full Hierarchical Model

- Observe $Y_i = 0$ or 1 depending on whether the attack was a suicide attack

$$Y_i \sim \text{Bernoulli}(p(\mathbf{X}_i)), \quad i = 1, \ldots, n$$

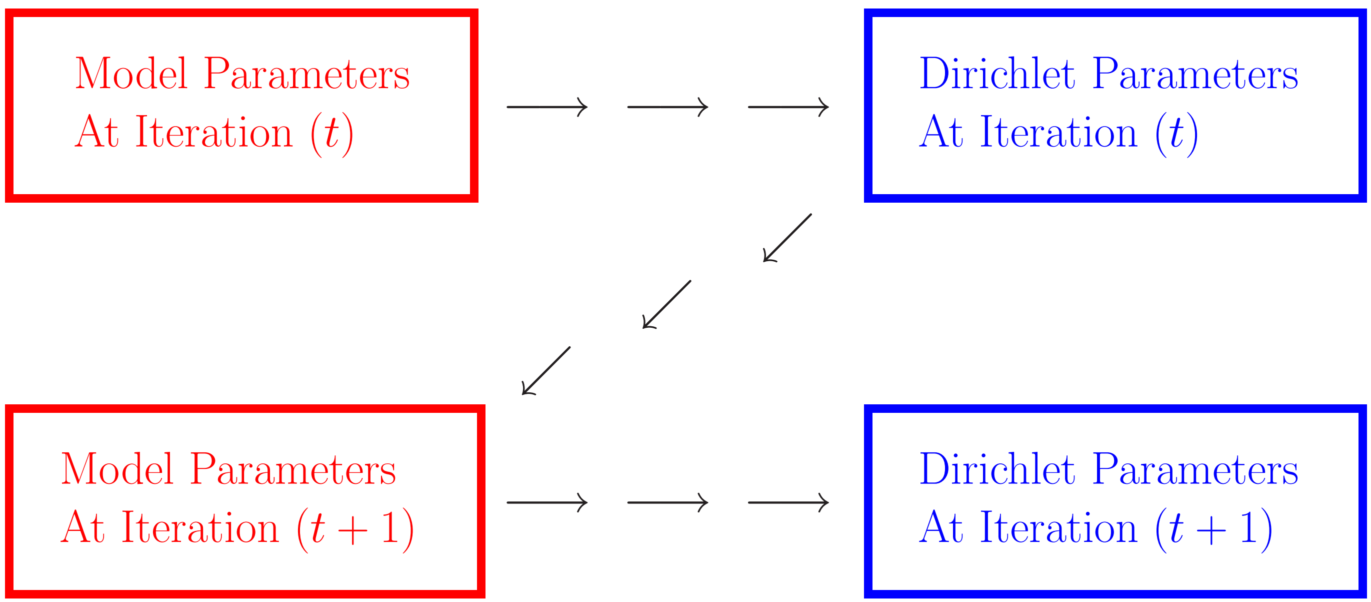
$$\text{logit}(p(\mathbf{X}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \psi_i,$$

- $\boldsymbol{\beta} \sim N\left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 I\right)$
  - ▷ $\mu \propto 1$ a flat prior
  - ▷ σ^2 is fixed
- Model Parameters

- $\psi_i \sim G, \quad G \sim \mathcal{DP}(mG_0),$
  - ▷ $G_0 = \text{Normal}(0, \tau^2)$
  - ▷ $\tau^2 \sim$ Inverted Gamma
  - ▷ $m \sim$ Gamma
- Dirichlet Parameters

Fitting the Model
Markov Chain Monte Carlo

- Use a Gibbs Sampler, a Markov Chain Monte Carlo Algorithm.
  - Estimates the posterior distribution of the parameters
  - Gives point estimates and confidence intervals
- Iterates between Model Parameters and Dirichlet Parameters.



Fitting the Logistic Parameters
Mixture Representation

- Logistic is a Mixture of Normals

  - Kolmogorov-Smirnov density:

$$f_{KS}(x) = 8 \sum_{\alpha=1}^{\infty} (-1)^{\alpha+1} \alpha^2 x e^{-2\alpha^2 x^2} \quad x \geq 0$$

  - Mixture of normals is logistic (Andrews and Mallows 1974)

$$\int_0^{\infty} \frac{1}{2x\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y}{2x}\right)^2\right\} f_{KS}(x)\, dx = \frac{e^{-y}}{(1+e^{-y})^2}$$

- Easy to simulate (Devroye's (1986) Accept-Reject Algorithm)
- Outperforms Slice Sampler

Fitting the Dirichlet Parameters
Matrix Representation of Partitions

- $\boldsymbol{\psi} \sim \mathcal{DP}$
  - $\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\eta}$, $\eta \sim N_k(0, \sigma^2 I)$
- $\mathbf{A}_{n\times k}$ random with
  - Rows: a_i is a $1 \times k$ vector of all zeros except for a 1 in its subcluster
  - Columns: Column sums are the number of observations in the groups
- To Generate $\mathbf{A}$

$$\begin{aligned} \mathbf{q}_{n\times 1} &\sim \text{Dirichlet Distribution} \\ a_i &\sim \text{Multinomial}, \quad i = 1, \dots n \end{aligned} \qquad \mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

- Eliminate columns with all zeros (Kyung *et al.* 2010)

Analysis of the Terrorism Data
Background

- The data come from the Global Terrorism Database II
  - Events in the Middle East and Northern Africa from 1998 to 2004
- 1998: 273 attacks worldwide, record high of 741 killed, 5952 injured.
  - 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.
- Categorization of Attack Types

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

- Outcome variable: Suicide attack/Not. ← Case-Control
- Suicide attacks pose a substantially higher challenge for governments
  - The assailant has great control over placement and timing
  - Does not need to plan his or her escape (Pape 2006).

Analysis of the Terrorism Data

Some Covariates Used in the Analysis

`MULT.INCIDENT`	Indicates if the attack is part of a coordinated multi-site event
`SUCCESSFUL`	The perceived success rated by the party attacked
`WEAPON.TYPE`	Type of Weapon: Bomb or Other Weapon

Analysis of the Terrorism Data

Other Covariates Used in the Analysis

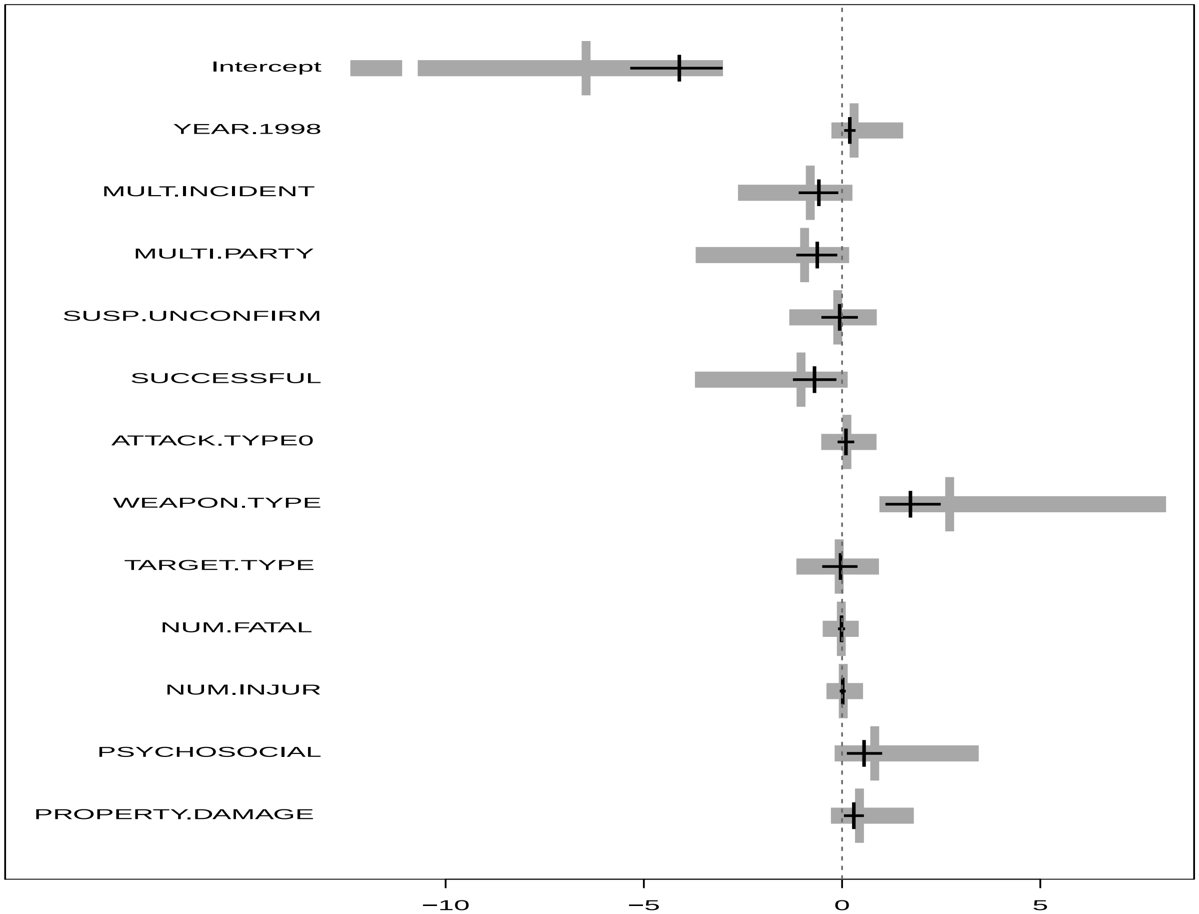
`NUM.INJUR`	Extent of human damage from the terrorist attack.
`PROPERTY.DAMAGE`	Amount of property damage.
`PSYCHOSOCIAL`	The negative psychological/social impact; ascending levels: none, minor, moderate, and major.

But First, We Are Statisticians After All
Model Results, Suicide Attacks

	Standard Bayes Model				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
PROPERTY.DAMAGE	0.439	0.305	0.122	1.406	0.297	0.094	0.114	0.483
PSYCHOSOCIAL	0.824	0.633	0.216	3.044	0.555	0.192	0.188	0.944

- Standard errors are smaller with $\mathcal{DP}$ random effects

Model Results, Suicide Attacks
Grey=Standard, Black=$\mathcal{DP}$



- And the credible intervals tend to be shorter

Analysis of the Terrorism Data
Estimates and Confidence Intervals

Coefficient	Coefficient	Std. Error	95% CI	
MULT.INCIDENT	-0.585	0.221	-1.028	-0.162
SUCCESSFUL	-0.695	0.245	-1.172	-0.210
WEAPON.TYPE	1.725	0.320	1.162	2.422
NUM.INJUR	0.017	0.004	0.008	0.025
PROPERTY.DAMAGE	0.297	0.094	0.114	0.483
PSYCHOSOCIAL	0.555	0.192	0.188	0.944

- Significant Coefficients

Analysis of the Terrorism Data
Results

`MULT.INCIDENT` -0.585^*	Multiple coordinated incidents are less associated with suicide attacks (9/11/2001 an exception)

▶ Planners of simultaneous terrorist events find it difficult to arrange multiple suicidal terrorists.

`SUCCESSFUL` -0.695	Successful attacks are less likely to be from suicides

▶ With suicide attacks, variables such as fervent nationalism and religious extremism, experience, age, intelligence, are important

`WEAPON.TYPE` 1.725	Bomb attacks are more likely to be from suicide terrorists.

Analysis of the Terrorism Data

Results – Continued

`NUM.INJUR` 0.017	More injuries at the event site suggest a greater probability of a suicide attack.
`PROPERTY.DAMAGE` 0.297	Increased property damage is positively associated with a suicide attack.

- This shows the terrorists preference for civilian targets, which will have more damage than better protected military targets.

`PSYCHOSOCIAL` 0.555	A goal of suicide attacks are consequences such as the psychological/social effect.

- A fundamental goal of terrorism is to reduce the people's confidence in the ability of their government to defend them

Conclusions
What Did We Learn From the Model?

- Multiple groups working together do not typically use suicide attackers
- They work in a more military manner with standard weapons

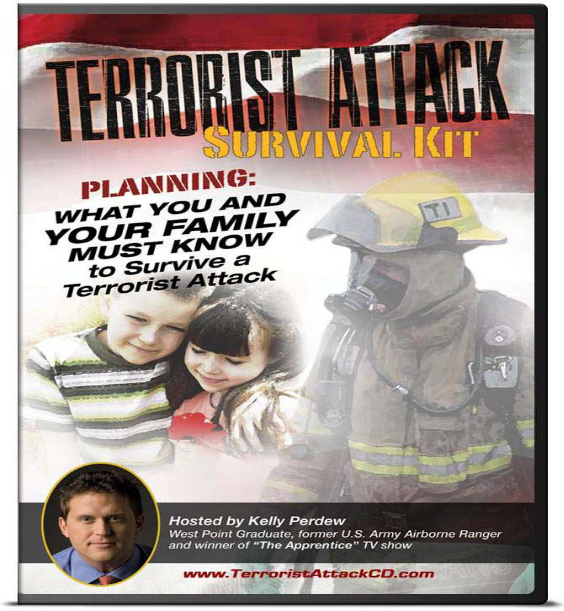
- Increased property damage from suicide attacks.
- Increased human injuries from suicide attacks.
- Suicide attackers prefer civilian targets
- Fewer fatalities from suicide attacks.

Conclusions
From This Model to the Next Step in the Statistical Analysis

- Advantages of the Dirichlet model

- Usual Model
  - Cannot remove enough error variability
  - Over-estimates effects of the covariates

- Dirichlet Model
  - Removes addional error variability
  - Does not over-estimate covariate effects



- We need more explanatory power
  - More covariates
  - More government data
  - Meta-analysis

- These findings may help governments reduce effectiveness of terrorist events.

Conclusions
What Actions are Suggested from the Data Analysis?

- Information on
  - Target/Weapon preferences
  - Multiple/Single attacks

Help focus intelligence gathering

- Plotters of suicide attacks want
  - negative psychological/social impact

- Better education of the population
- Increase availability of counseling

Conclusions

What Actions are Can We Hope For?

- A challenge to the terrorist
  - Reduces success

- Increase Police/Military Presence
- Increase Population Awareness

- We hope for more stories like this

Thank You for Your Attention

George Casella

casella@ufl.edu

Findings So Far for Dirichlet Process Random Effects in GLMs

- Gill and Casella(2009). "Nonparametric Priors For Ordinal Bayesian Social Science Models: Specification and Estimation." *JASA*, 104, 453-464
  DPP on RE can uncover latent clustering.

- Kyung *et al.*(2009) "Characterizing the Variance Improvement in Linear Dirichlet Random Effects Models." *Stat. Prob. Letters*, 79, 2343-2350
  DPP on RE can produce lower SE for regression parameters on average.

- Kyung, Gill and Casella(2010) "Estimation in Dirichlet Random Effects Models." *Annals of Statistics*, 38, 979-1009
  Estimation of the precision parameter; improved Gibbs sampler.

- Kyung *et al.* (2011) "Sampling Schemes for Generalized Linear Dirichlet Process Random Effects Models." *Stat. Methods & Applications, to appear*.
  Slice sampling worse than KS mixture representation or MH algorithm.

- Kyung *et al.* (2011) "New Findings from Terrorism Data: Dirichlet Process Random Effects Models for Latent Groups." *JRSSC, to appear*.
  Logistic model, uncovering latent information with difficult data.