Random walk negative binomial: a model for persistent count series.

Andreas Beger (andreas.beger@duke.edu)

Duke UNIVERSITY

1. Introduction

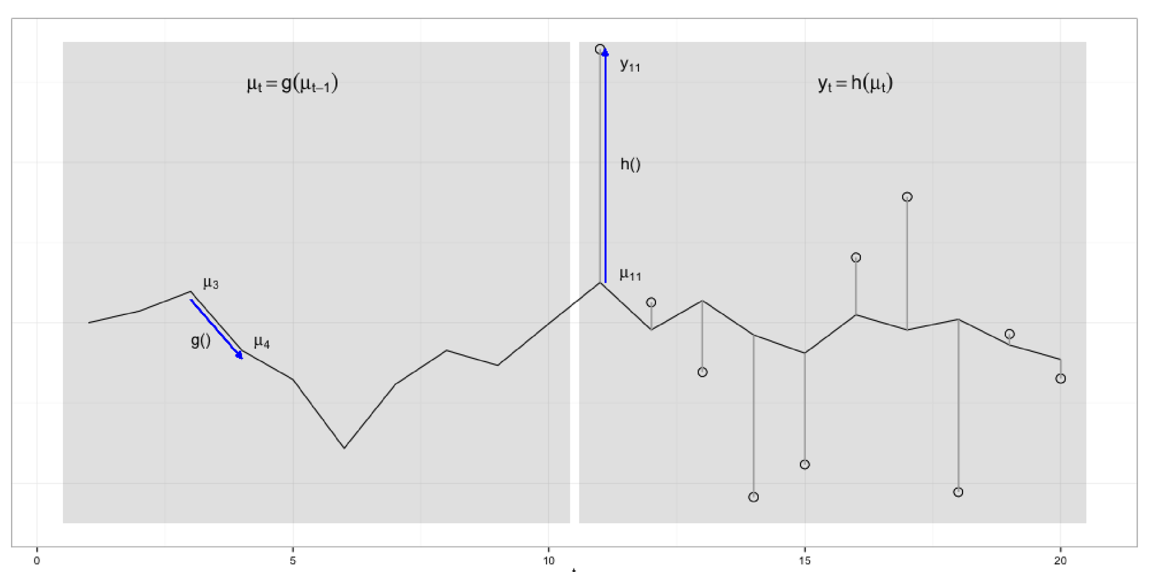
Time-series event counts are common in political science, e.g. violent events or deaths over time, legislation passed or vetoed, annual counts of interstate wars or militarized interstate disputes, etc., and this poster presents a model designed for such persistent counts. The two sections below introduce state-space modeling in general and the random walk negative binomial model used here. The next box on the top row shows some simulated data created with this process to illustrate the flexibility of the data-generating process. Data for an empirical application to Iraqi civilian deaths are presented in the large box on the top right, and results are in the large box in the bottom row. Highlights and references are on the bottom right.

2. State-space models

State space models are a general class of time-series models that separate observations Y from a latent system state μ:

$$\begin{aligned} y_t &\sim f(\mu_t) \\ \mu_t &\sim g(\mu_{t-1}) \end{aligned}$$

where the functions $f()$ and $g()$ govern the observation and state transition processes respectively. This framework is flexible and can accommodate many non-stationary, trending, and seasonal types of data.



3. Random walk negative binomial – NB I(1)

To model time-series event counts we can use a negative binomial density for the observation function and gamma for the state transition function, in what essentially is a random-walk negative binomial or NB I(1) model:

$$\begin{aligned} y_t &\sim NegBin(mean = \mu_t \times e^{\delta x}, precision = \theta_{obs}) \\ mu_t &\sim Gamma(mean = \mu_{t-1} \times e^{\gamma x}, rate = \beta_{proc}) \end{aligned}$$

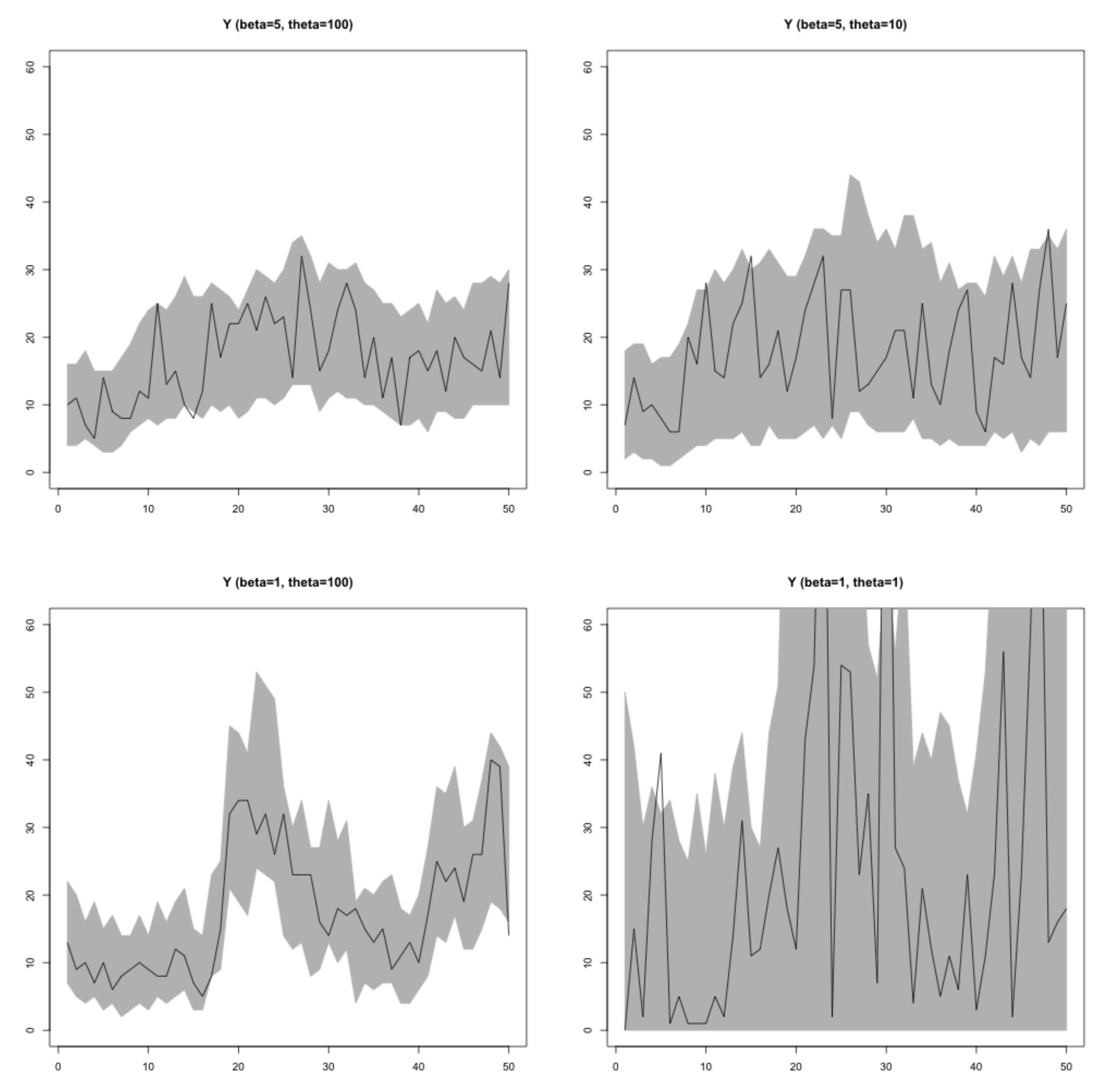
This model has several distinctive features:

- Separate observation and state transition errors, which better accommodates data with significant error sources.
- Short-term vs. long-term effects, depending on whether covariates impact observation or state transition.
- Accommodates overdispersion and persistent series.

4. DGP simulations

The parameters θ_{obs} and β_{proc} govern the precision of the observation and state transition processes, with larger values indicating higher precisions and less variance/error. Changing these parameters creates a broad range of simulated data.

Simulated data with varying state transition and observation precisions.

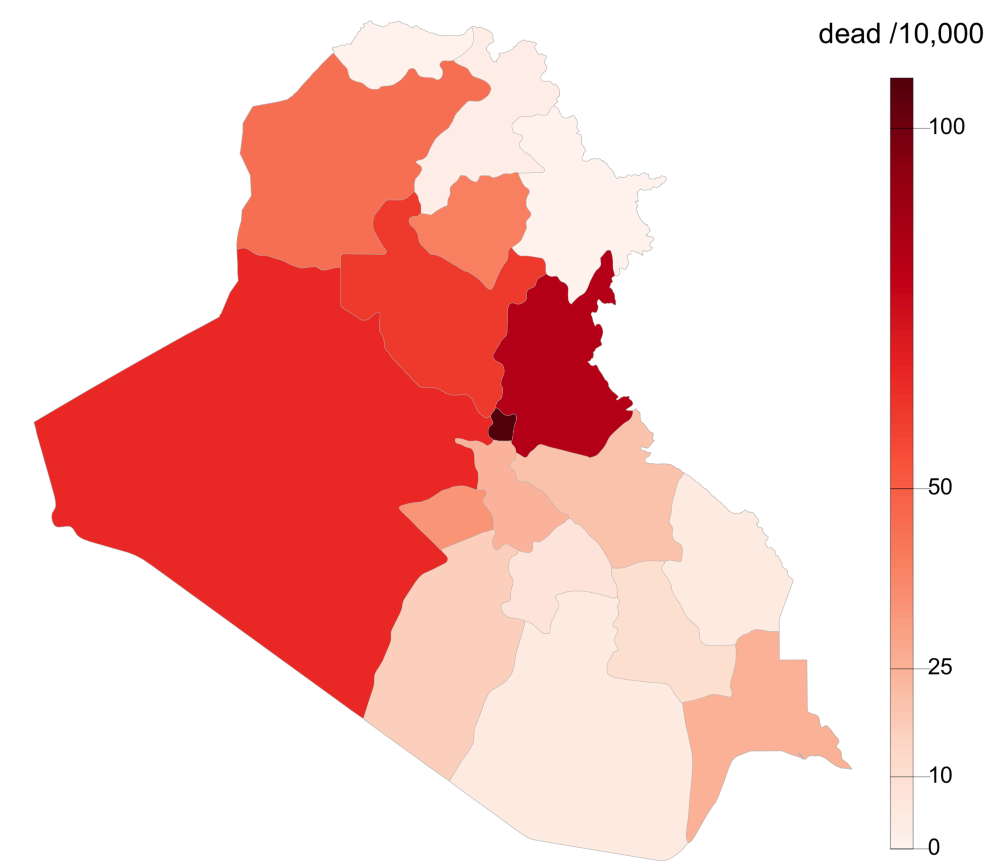


5. Data example: Iraq Body Count

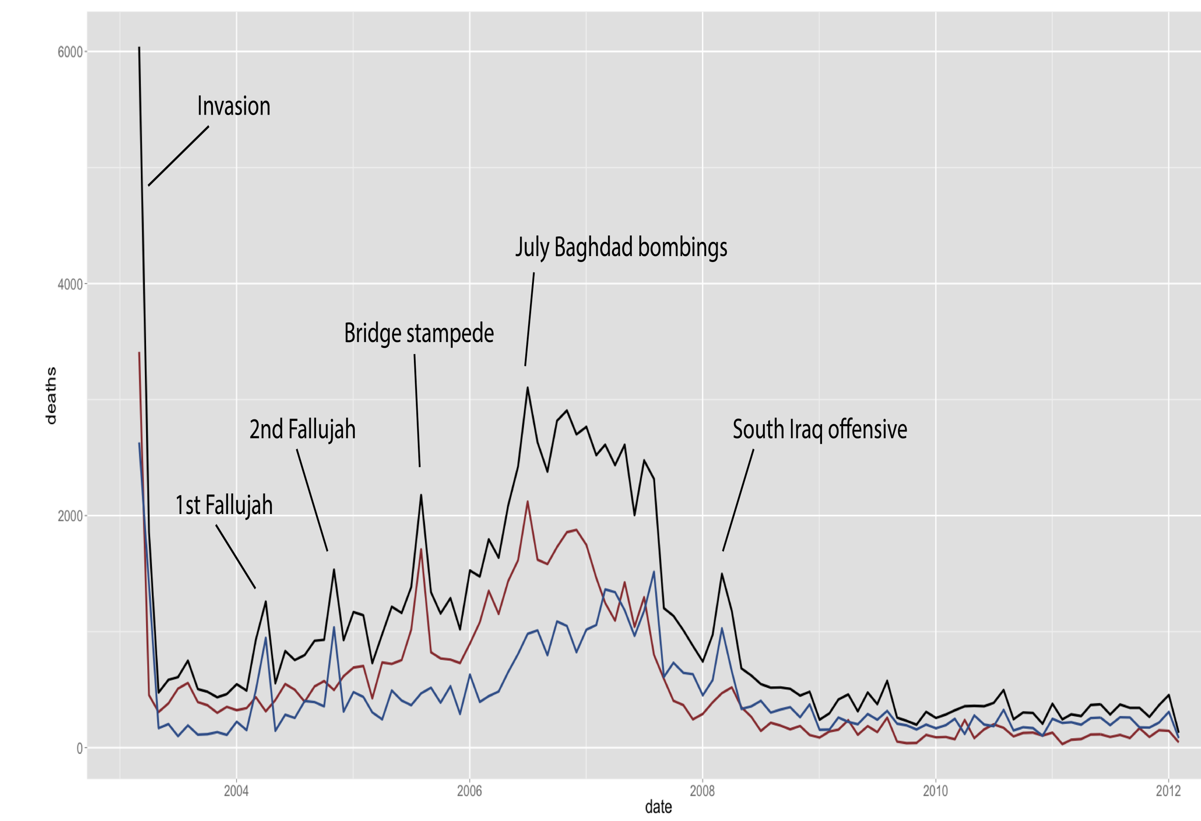
Source and process. Events consist of media reports collated by the Iraq Body Count (IBC) project. They include high and low estimates of deaths as well as descriptive general location or nearby city. Provinces are coded from location descriptions using an automated R script that classifies locations using a dictionary of ~200 city-province pairs.

Summary: 27,500 events from March 2003 to February 2012, with > 95% province coding accuracy. Civilian death totals range from $100,000 - 110,000$. Dependent variable is monthly reported deaths Iraq-wide for 108 months. Covariates consist of four binary indicators for the invasion period, major government offensives, Ramadan, and elections.

Per capita civilian deaths in Iraq from March 2003 to February 2012.



Monthly civilian deaths country-wide (black), in Baghdad (red), and in the rest of Iraq (blue).

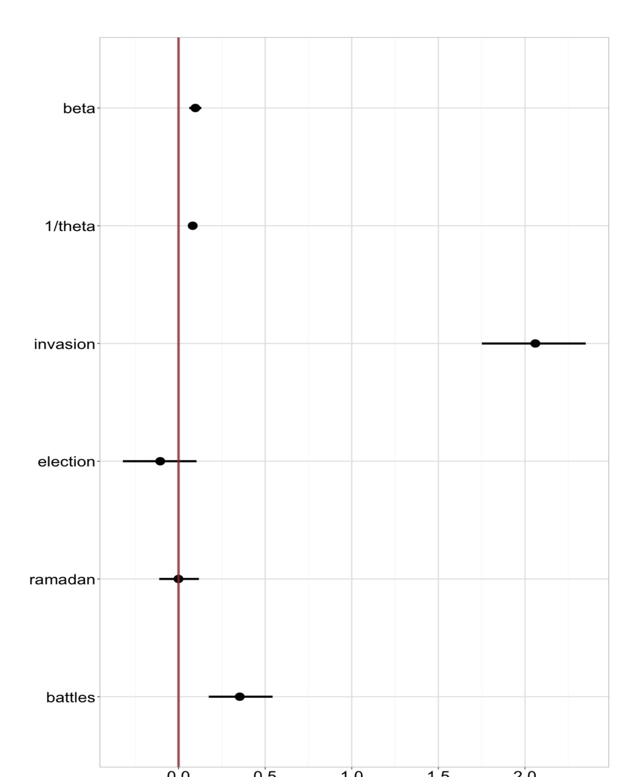


6. Model of Iraq-wide monthly deaths

How well does this random walk negative binomial model do against real data? The results here are for a model of monthly Iraqi civilian deaths which includes 4 binary covariates in the measurement equation (i.e. short impact only). Estimates are via MCMC using JAGS and R. Below are parameter estimates, and below right fitted values.

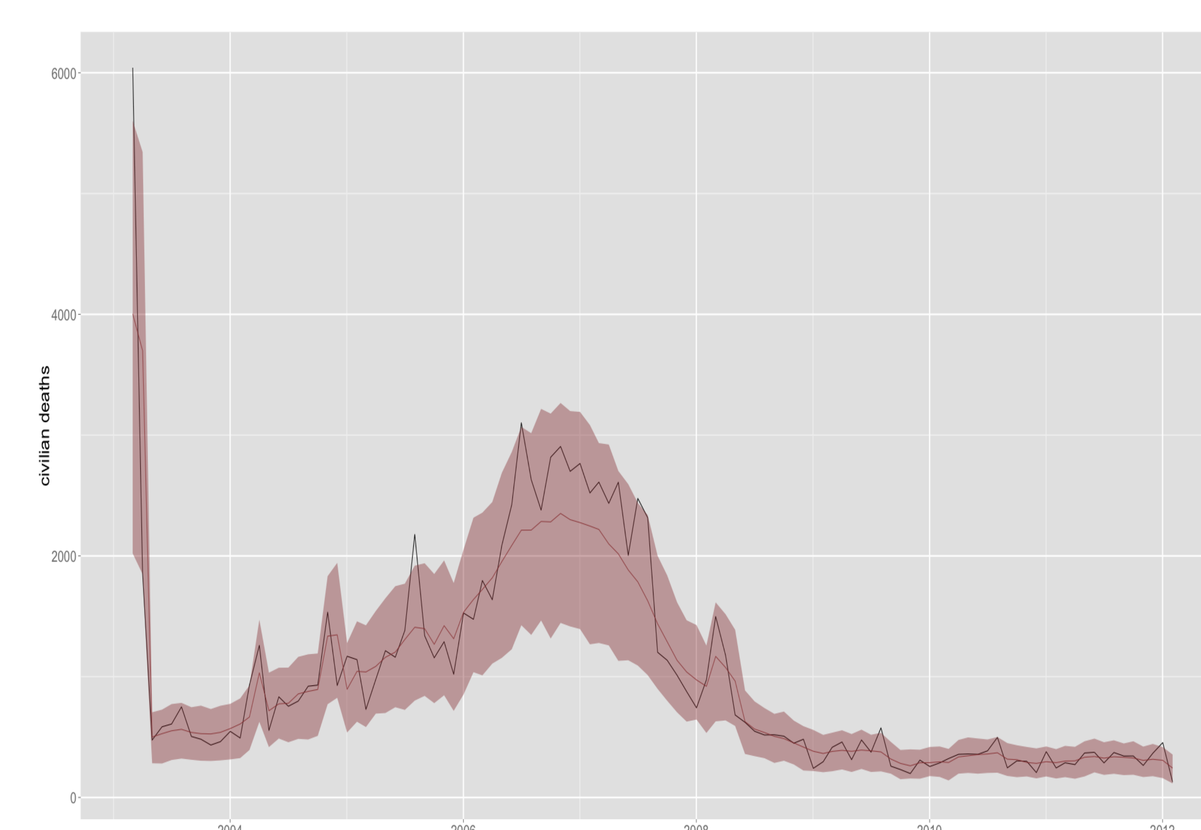
Parameter estimates for the random walk model

Beta and theta are the precision parameters. Covariate coefficients enter the observation equation, and thus have a instantaneous impact only, not on future time periods. They can be interpreted like count regression coefficients. Iraqi civilian deaths were significantly higher during the invasion period and during major government offensives, as the estimates show.



The black line in the plot below shows actual total monthly civilian deaths in Iraq. Median fitted values for y are shown by the red line, with an 80% interval corresponding to the red band. The jitter in the fitted point and interval predictions is an artifact of the estimation procedure, but even so the median fitted values present a plausible smoothing of the observed monthly deaths.

In-sample predictions for a random walk negative binomial model of monthly Iraqi civilian deaths.



7. Highlights

- Model for non-stationary, persistent time-series event counts with possible over-dispersion.
- Covariates can have separate short-term and long-term effects, depending on whether they impact the state transition or measurement processes.
- Implemented in JAGS/BUGS and tested against simulated data with known generating process.
- Example application to Iraq monthly civilian deaths, showing increased violence during invasion and government offensives, but not elections or Ramadan.
- R function to code provinces for IBC events, 95% accuracy for ~27,500 records using 200 word dictionary.
- Future directions: (1) extensions for a spatial component, trending and seasonality, (2) R interface function and post-estimation support.

8. References

1. Fukumoto, Kentaro. 2006. "A Bayesian Analysis of Time-Series Event Count Data." Working paper.
2. Beger, Andreas, Kentaro Fukumoto and Will H. Moore. 2012. "Event Count Time Series Models: A Bayesian State Space Investigation." Working paper.