Uncertain Events:
A Dynamic Latent Variable Model of Human Rights Respect and Government Killing with Binary, Ordered, and Count Outcomes

Christopher J. Fariss

Pennsylvania State University

Summary

- The counting of repressive events is difficult because state leaders have an incentive to conceal the actions of their subordinates and destroy evidence associated with abuse
- I introduce a model that generates new and more informative estimates of the number of one sided government killings.
- The model builds on an existing latent variable model of repression or respect for human rights.
- The original model includes 13 categorical variables measuring state-sponsored repression drawn from standards-based and events-based sources.
- In previous work, I demonstrated that documentary sources used to generate the standards-based data systematically changed over time, which necessitated the use of dynamic item-difficulty parameters for these items in the latent variable model.
- The extended version of the model presented here accounts for the uncertainty related to the estimation of heterogenous event data by introducing overdispersion parameters unique to each country-year observation (r_{it}), which are related to one another using hierarchical priors (δ_t and η_t).

One Sided Government Killing Data

- Data are published by the Uppsala Conflict Data Program (UCDP).
- Estimates are included in the data if at least 25 individuals (non-combatants) are killed.
- Reported estimates are generated using a variety of documentary sources in order to provide three estimates of one-sided government killing: $\{Low,\ Best,\ High\}$.
- The data arrise from the same underlying process, i.e., the latent level of repression. It is therefore useful to model all three outcomes as a function of the same underlying data generating process for each country-year observation with varying amounts of uncertainty represented by the three estimates published each country-year observation in the data.

The Model

	M_5	M_6
Latent Variable		
country-year latent variable (first year)	$\theta_{i1} \sim N(0,1)$	$\theta_{i1} \sim N(0,1)$
country-year latent variable (other years)	$\theta_{it} \sim N(\theta_{it-1}, \sigma)$	$\theta_{it} \sim N(\theta_{it-1}, \sigma)$
uncertainty of latent variable	$\sigma \sim U(0,1)$	$\sigma \sim U(0,1)$
Model Parameters (Categorical Data)		
event-based variable cut-points (constant)	$\alpha_{jk} \sim N(0,4)$	$\alpha_{jk} \sim N(0,4)$
standards-based variable cut-points (first year)	$\alpha_{1jk} \sim N(0,4)$	$\alpha_{1jk} \sim N(0,4)$
standards-based variable cut-points (other years)	$\alpha_{tjk} \sim N(\alpha_{t-1,jk}, 4)$	$\alpha_{tjk} \sim N(\alpha_{t-1,jk}, 4)$
slope	$\beta_j \sim Gamma(4,3)$	$\beta_j \sim Gamma(4,3)$
Model Parameters (Count Data)		
event-based variable cut-points (constant)	$\alpha_J \sim N(0,4)$	$\alpha_J \sim N(0,4)$
slope	$\beta_J \sim Gamma(4,3)$	$\beta_J \sim Gamma(4,3)$
population over-dispersion rate	$r_j \sim U(0,100)$	
country-year over-dispersion rate		$r_{it} = exp(\delta_i + \eta_t)$
country-random effect for over-dispersion rate		$\delta_i \sim N(0,1)$
year-random effect for over-dispersion rate		$\eta_t \sim N(0,1)$

The likelihood function for the parameters given the data and model is:

$$\mathcal{L}(\beta, \alpha, \theta, r | y, M_6) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\left[F(\alpha_{tjy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{tjy_{itj}-1} - \theta_{it}\beta_j) \right]^{(v_j)*(1-c_j)} \right] *$$
$$\left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j) \right]^{(1-v_j)*(1-c_j)} *$$
$$\left[\frac{\Gamma(r_{it} + y_{itj})}{\Gamma(r_{it})y_{itj}!} \left(\frac{r_{it}}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}} \right)^{r_{it}} \left(\frac{exp(\alpha_J + \theta_{it}\beta_J)}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}} \right)^{y_{itj}} \right]^{(1-v_j)*(c_j)}$$

Note: $F(\cdot)$ denotes the logistic cumulative distribution function. For notational convenience let $v_j = 1$ when the j indicator is one of the standards-based variables and then $v_j = 0$ when it is one of the event-based variables and let $c_j = 0$ when the j indicator is binary or ordinal and then $c_j = 1$ when the j indicator is measured as a count.

Model Comparisons

	ρ_{M_5}	ρ_{M_6}
Best	0.693 [0.620, 0.755]	0.919 [0.897, 0.937]
Low	0.746 [0.683, 0.798]	0.933 [0.914, 0.948]
High	0.710 [0.633, 0.763]	0.909 [0.883, 0.929]

Correlation coefficients between the three observed one-sided government killing count variables $\{Low, Best, High\}$ and the estimated count variable from two versions of the extended latent variable model. Model M_5 assumes one over-dispersion parameter r_j for each of the three one-sided government killing variables. Model M_6 assumes a unique over-dispersion parameter r_{it} for each of the country-year observations.

New Estimates of One Sided Government Killing
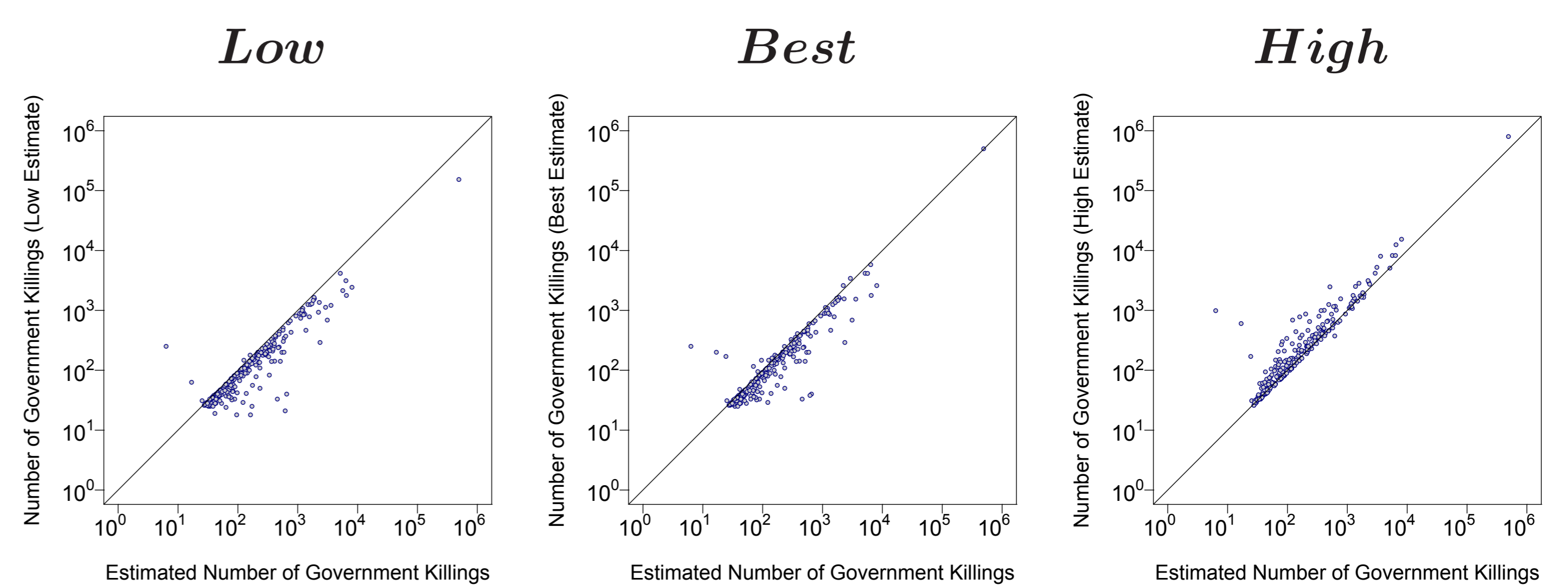
The points are the observed $\{Low, Best, High\}$ value for the count variable plotted against the predicted posterior count from the latent variable model M_6.
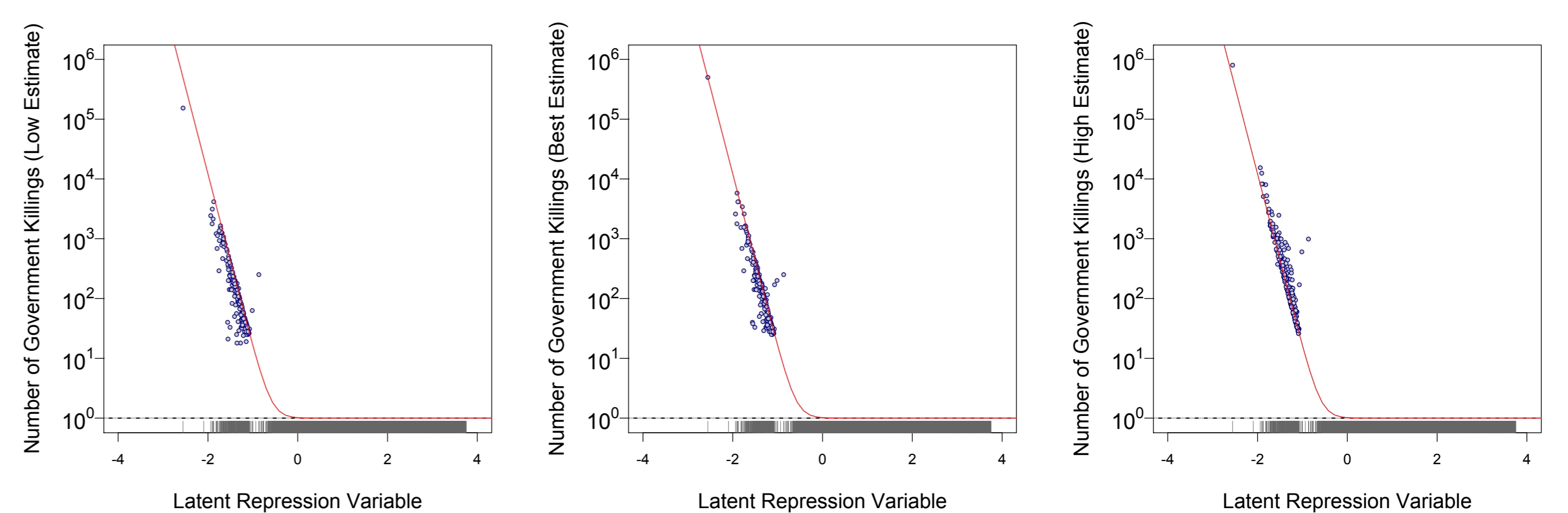
New latent variable estimates of human rights provide predictions of the expected number of government one-sided killings. The red line is the posterior expectation of the count variable. The points are the observed best value reported by the UCDP conflict dataset for the count variable plotted against the corresponding latent variable estimate. The model suggests that one-sided government killing stops at approximately the mean value of the "true" level of repression. Government-killing reaches a level great enough for the UCDP conflict coders to find sufficient evidence of such human rights violations such that the country-year observations enter the dataset at approximately one standard deviation (-1.0) below the mean value of the "true" level of repression.

Replication

Table : Replication of study using one-sided government killing data.

	Original	Replicated	Difference	$p-value$	n
Model 1	0.164 (0.060)	0.450 (0.151)	0.286	0.039	850
Model 2	0.379 (0.158)	0.418 (0.231)	0.039	0.445	890
Model 3	0.400 (0.180)	1.013 (0.346)	0.613	0.058	850
Model 4	0.684 (0.260)	0.611 (0.381)	-0.073	0.437	890
Model 5	0.569 (0.190)	1.074 (0.347)	0.505	0.101	345

The new uncensored count estimates generated from the extended latent variable model have strengthened the reported relationship between violence against civilians and UN interventions found by Hultman (2013). Analysts that use the UCDP one-sided government killing data for future research should consider using the new count estimates presented in this paper along side the three existing estimates. Inconsistent results would suggest that censoring is biasing the results between models using the original data and estimates using the new count estimates presented here.