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 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 estimate of heterogenous event data by introducing overdispersion parameters unique to each country-year observation ($r_{ij}$), which are related to one another using hierarchical priors ($\delta_i$ and $\eta_j$).

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 arise 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

<table>
<thead>
<tr>
<th>Latent Variable</th>
<th>$M_L$</th>
<th>$M_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>country-year latent variable (first year)</td>
<td>$\delta_1 \sim N(0, 1)$</td>
<td>$\delta_1 \sim N(0, 1)$</td>
</tr>
<tr>
<td>country-year latent variable (other years)</td>
<td>$\alpha_j \sim N(\alpha_j-1, 4)$, $\sigma \sim U(0, 1)$</td>
<td>$\sigma \sim U(0, 1)$</td>
</tr>
</tbody>
</table>

Model Parameters (Categorical Data)

- event-based variable cut-points (constant)
  - $\alpha_j \sim N(0, 4)$
  - $\alpha_j \sim N(\alpha_j-1, 4)$
- standards-based variable cut-points (first year) and standards-based variable cut-points (other years)
  - $\beta_j \sim Gamma(4, 3)$
- model parameters (count data)
  - event-based variable cut-points (constant)
  - $\beta_j \sim Gamma(4, 3)$
  - $r_j \sim U(0, 100)$

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

$$L(a, \alpha, \beta, \delta, M, \theta) = \prod_{i=1}^{n} \left[ F(a_{ij}, \alpha_{i1}, \beta_{i1}) - F(a_{ij}, \alpha_{i1}, \beta_{i1})^{[\exp\left(c_{ij} + \beta_{i1}\right)}] \right] \times \frac{1}{	au(r_{ij})} \times \left( \exp\left(a_{ij} + \beta_{i1}\right) + r_{ij} \right) \left( \exp\left(a_{ij} + \beta_{i1}\right) + r_{ij} \right)^{\frac{1}{r_{ij}}}.$$

Note: $F(\cdot)$ denotes the logistic cumulative distribution function. For notational convenience let $c_i = 1$ when the $j$ indicator is one of the standards-based variables and then $c_i = 0$ 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

- 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 $M_L$ assumes a single over-dispersion parameter $r_1$ for each of the three one-sided government killing variables. Model $M_M$ assumes a unique over-dispersion parameter $r_{1ij}$ 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 $M_M$.

Replication

<table>
<thead>
<tr>
<th>Model</th>
<th>Low</th>
<th>High</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.460</td>
<td>0.379</td>
<td>0.596</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.450</td>
<td>0.418</td>
<td>0.613</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.500</td>
<td>0.150</td>
<td>0.704</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.684</td>
<td>0.611</td>
<td>0.505</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.569</td>
<td>0.190</td>
<td>0.505</td>
</tr>
</tbody>
</table>

The new censored 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.