Disaggregating Battle Death Counts using Multiple Imputation

Dana Higgins

Department of Government, Harvard University

Battle Deaths Data

Current data is limited by offering either coarse estimates (entire conflict or annual) or estimates for selected conflicts (sub-Saharan Africa). In order to impute monthly counts of battle deaths, the following datasets are used:

- UCDP Battle-Related Deaths Dataset
- UCDP/PRIO Armed Conflict Dataset
- UCDP Georeferenced Event Dataset
- Global Database of Events, Language, and Tone (GDELT)

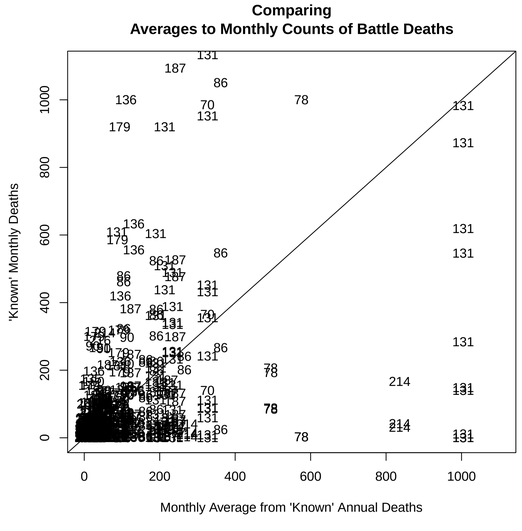
Monthly estimates are imputed for 141 interstate and internationalized intrastate conflicts from 1989 to 2012.

Motivation

- Time-varying analysis involving conflict intensity
- Measuring changing rates of violence
- Obtaining fine-grained data from grouped or aggregated measures in other areas

Current Approach

Efforts to use monthly data currently use the average, which may induce post-treatment bias and, as shown below, is generally inaccurate.

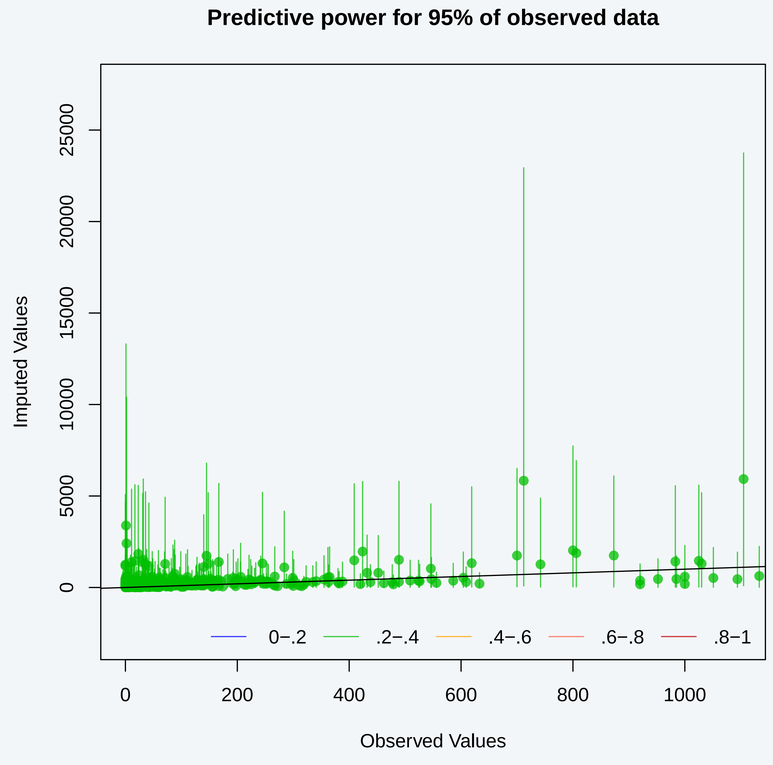


Multiple Imputation Model

Using Amelia, monthly battle counts were imputed using the following covariates: annual count, counts of monthly reports of violence, and counts of monthly reports of violence with small arms and tanks. Out of 141 conflicts, 45 had complete information, including monthly battle counts. The remaining conflicts had complete missingness in monthly battle counts and complete information otherwise.

Overimputation

To test the accuracy of the model, overimputation eliminates known monthly counts and using the imputation model to predict those values. The following graph shows 90% confidence intervals around the point estimates for 95% of the known data:



Future Expansions

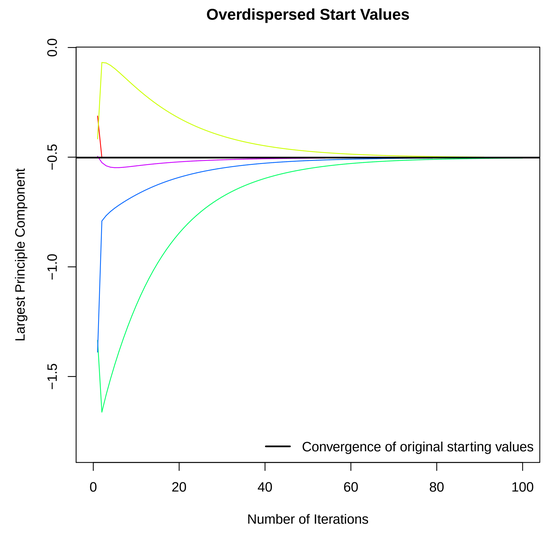
1. The model presented is minimalist in predictive covariates, resulting in reduced variation in imputed values. Additional information such as conflict characteristics (primary issue of dispute, conflict type, number of disputing parties, etc) and explanation of variations in reporting accuracy (terrain, infrastructure, year, duration of conflict, etc) will be considered to include.
2. The UCDP battle death estimates are taken as accurate for the purposes of imputation, but consistency with other estimates will be checked.
3. Annual counts will be used to impute $n-1$ of the months in a conflict-year with the nth month calculated from remaining known deaths

Assumptions

- Complete data is multivariate normal
- Data is missing at random
- sub-Saharan conflicts with known monthly battle deaths are comparable to others
- Reporting accuracy does not alter across months of a conflict

Overdispersion

To ensure the global maximum was found, overdispersion sets the starting point for the algorithm at distant points in the latent space. The following graph shows convergence to the original point, strongly suggesting it is a global not local maximum:



References

[1] Dempster, Arthur, N.M. Laird and D.B. Rubin. (1977) Maximum likelihood estimation from incomplete data via the em algorithm. Journal of the Royal Statistical Society B 39:1-38

[2] Honaker, James, Gary King, Matthew Blackwell. (2013) Amelia II: A Program for Missing Data

[3] Uppsala Conflict Data Program, Uppsala University, Battle-Related Deaths Dataset v.5 (2013), UCDP/PRIO Armed Conflict Dataset v.4 (2013), GED v1.5 (2011)

[4] Leetaru, Kalev and Philip Schrodt (2013) GDELT: Global Data on Events, Location and Tone, 1979-2012. ISA Presentation