

PLSC 504

Exercise Two

September 7, 2023

Part I

The first part of this exercise is simulation-based. Specifically, you're tasked with answering the question "What happens if my data have zero inflation, but I ignore it and fit a standard count model instead?" You're free to examine this question however you would like. *One* approach might be:

- Start by generating a response/"dependent" variable Y that follows a known count distribution (the Poisson being the most obvious choice) and has a known association with one or more predictors X . Show that the corresponding regression model "works" on those data.
- Next, introduce "excess" zeros by randomly setting some fraction of the observations π to zero, fitting the model, and examining the results.
- Repeat the above step, varying π .
- Repeat the above two steps, this time adding zeros in a way that is correlated with one or more elements of X .
- In every instance, repeat the process many times and summarize (and discuss!) your findings.

Note: This is one possible approach to this exercise. You can surely think of others as well; when you do, feel free to do them instead, if you so choose.

Part II

In a 2012 [article](#) in *International Studies Quarterly*, Matthew Krain asks (what I think is) an interesting question: Can "naming and shaming" of leaders by transnational organizations (IOs, NGOs, and/or the media) reduce the severity of ongoing genocides / politicides? Krain analyzes cross-national time-series data on genocides between 1976 and 2008; his primary dependent variable is an 11-category ordinal indicator of the severity of the genocide; his predictors include several indicators of targeting of the regime by western news media, Amnesty International, and the UNHCR, along with a series of control variables. The (short) article is available at the link above; replication materials for Krain's paper are available at the article's [dataverse](#), and the data for this assignment are available on the PLSC 504 [Github repository](#), in the "Exercises" folder.

This part of the assignment asks you to conduct a replication + extension + validation of Krain's 2012 article. More specifically:

1. Begin by replicating the results that appear in the first column (Model 1) of Krain's Table 3. Those results examine the effect of western media reporting on genocides, and will be the focus of this assignment. For purposes of what we're doing here, you need not concern yourselves with either the panel/TSCS aspects of the data nor the use of "robust" (Huber/White) standard errors.

2. Second, assess the parallel regressions assumption for this model, with a particular focus on whether the assumption holds vis-a-vis the main independent variable (western media coverage). Discuss your findings, and what (if anything) you would recommend regarding reanalysis.
3. Third, consider the effects of differential aggregation of the outcome variable. More specifically, reanalyze Krain's outcome using the following coding:
 - Fewer than 1,000 deaths → 0
 - 1,000-8,000 deaths → 1
 - 8,000-64,000 deaths → 2
 - More than 64,000 deaths → 3

(For this part of the exercise, you'll need to re-aggregate Krain's main dependent variable.) Once again, consider the parallel regressions assumption in this revised model. Discuss any differences in your results from what you found in step (2).

4. Finally, re-fit Krain's Model 1 – with the original / 11-category response variable – using linear (OLS) regression. Discuss any differences in your findings from those in parts (1) and (3), and briefly assess the utility of adopting ordered-response models in this particular case.

This assignment – write-up / discussion and computer code – is due *electronically*, as a *PDF file*, at 11:59 p.m. EST on Friday, September 15, 2023; you can submit your homework by emailing copies **both** to Dr. Zorn (zorn@psu.edu) and Mr. Morse (nam@psu.edu). This assignment is worth 50 possible points.