PLSC 504

Exercise Two

September 15, 2017

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 $\mathbf{X}$. 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 $\mathbf{X}$.
- In every instance, repeat the process many times and summarize (and discuss!) your findings.

Part II

We'll be addressing the subject of civil war, with a particular focus on Africa. The data are annual measures for 50 African countries from 1995-2007, though with missing data the total $NT \approx 400$. The main variable of interest is creatively named `count`, and is the count of "internal" armed conflicts (that is, civil wars and insurgencies) for each country in each year, as measured by the Uppsala Armed Conflict Project data. Covariates include:

- measures of `latitude` and `longitude`, with which you are free to do whatever you feel like,
- the adult `literacy` rate for each country-year,
- the number of `refugees` living in that country in that year,
- the (lagged natural log of) `GDP` and `trade`, in constant dollars, and
- the lagged `polity` score ($\in [-10, 10]$, with 10 = full democracy and -10 = full autocracy).

Your assignment for this exercise is as follows:

1. Begin by describing / summarizing the response ("dependent") variable, in whatever fashion you deem appropriate.

2. Next, specify and estimate a model of the incidence of internal conflict. Feel free to include whatever variables, in whatever form (logged, etc.) you feel are appropriate; *do not* feel obligated to use all available covariates. Interpret your results, using the approaches we discussed in class and/or whatever means you deem useful.

3. Suggest some substantive reasons why, for example, over- or underdispersion might be a problem in these data. Test for the possibility of overdispersion, using the tests we covered in class and/or a negative binomial regression. Report and discuss your results, including both those relevant to overdispersion as well as any changes in coefficients.

4. Fit, interpret, and discuss one or more zero-modified (zero-inflated and/or hurdle) count models to the data in question. In your discussion, be sure to talk about why (in substantive terms) you chose the statistical model that you did.

This exercise should be submitted *electronically, in PDF format* by 5:00 p.m. ET on Thursday, September 21, 2017. It is worth 50 possible points.