Measuring State Capacities in Latin America using an Infrastructural Approach: Earthquakes and Institutional Development, 1900-2010

Héctor Bahamonde

*PhD Candidate • Political Science Dpt. • Rutgers University

e:hector.bahamonde@rutgers.edu

w:www.hectorbahamonde.com

March 20, 2017

Abstract

ABSTRACT

Please consider downloading the last version of the paper here.

*I thank Bob Kaufman, Dan Kelemen, Doug Blair, Paul Poast for all the useful comments. I also thank the School of Arts and Sciences at Rutgers for granting me funds to collect part of the data used in this project. All errors are my own.

I. Multilevel Analyses

In this section I model the number of dead individuals caused by earthquakes.

Include summary stats here. Explain what's national and what's subnational.

The data are fitted using a Bayesian Poisson regression. The main independent variables are the proportion of national agriculture output relative to industrial output and a dummy for whether in year t the law of income tax had been implemented. I expect the yearly death tolls to be lower when the national proportion of agricultural production decreases, when the law of income taxation has been passed, and where the industry predominates at the local level. The model controls for local population, an indicator for local urban/rural, and earthquake magnitude.

Since the 'treatment,' i.e. the proportion of agricultural output relative to industrial output, and the implementation of the income tax, takes place at the national level but the outcome (death tolls associated to earthquakes) is measured at the local level, I implement a multilevel model.[1]

Particularly, I include year fixed-effects to account for unobservable/unmeasured yearly factors such as the evolution of the political system, demographic, climate and cultural changes, economic shocks (both national and international), and others. Particularly, the multilevel component of Equation 1 allows the slopes of the national proportion of agriculture relative to industry (β_{1_j}) and the earthquake's magnitude (β_{2_j}) to vary by subnational sectoral predominance indexed by j. I consider whether affected localities were predominantly *agricultural*, *industrial* or *mixed*.

The latitude where the earthquake occurred was included to control for the proximity to the Andean mountains. This variable controls for a built-in tectonic predisposition of a higher propensity of earthquakes. Longitude controls for climate and other unobserved conditions that make agricultural development more difficult. In turn, both measurements serve as good proxies for terrain ruggedness and the difficulties the state had to face to centralize political power. More formally, I fit the next equation,

See if I included this lit. already.

$$
\begin{aligned}
\text{Deaths } &\sim \text{Poisson}(\lambda_i) \\
log(\lambda_i) =& \mu + \beta_{1_j}\text{Proportion}_i + \beta_{2_j}\text{Magnitude}_i + \beta_3\text{Tax}_i + \\
& \beta_4\text{Population}_i + \beta_5\text{Urban}_i + \\
& \beta_6\text{Latitude}_i + \beta_7\text{Longitude}_i + \beta_{8_t}\text{Year}_i
\end{aligned} \tag{1}
$$

where,

[1] Gelman and Hill [2006, 237].

$$
\begin{aligned}
i_{1,\ldots I} &\text{ where I} = 91 \\
j_{1,\ldots J} &\text{ where J} = 3 \\
t_{1,\ldots T} &\text{ where T} = 59.
\end{aligned}
\tag{2}
$$

The i indicator denotes the event (i.e. earthquake), the j index expresses the type of subnational economic composition of the affected locality (agricultural, industrial, or mixed), and the t subscripts denotes the year when earthquake i happened. Finally, μ is the intercept. Since earthquakes can happen more than once per year, in my dataset $i > t$.[2] The estimated parameters β_k have uninformative normally distributed priors, while the precisions τ_p of β_{1_j}, β_{2_j} and β_{8_t} have uninformative Gamma priors, of the form,

$$
\begin{aligned}
\beta_{k,\ldots K} &\sim \mathcal{N}(0, 0.01) \text{ where K} = 8 \\
\tau_{p,\ldots P} &\sim \mathcal{G}(0.5, 0.001) \text{ where P} = 3.
\end{aligned}
\tag{3}
$$

[2]For the years in which there is just one earthquake, the 'group' variable has only one observation. This does not endangers the robustness of the model. Gelman and Hill [2006, 276] explains that it "is even acceptable to have one observation in many of the groups."

	Mean	SD	Lower	Upper	Pr.
Prop. Agr/Ind (Agr.)	5.95	4.72	-0.19	11.81	0.89
Prop. Agr/Ind (Ind)	-16.73	4.63	-22.85	-10.89	1.00
Prop. Agr/Ind (Mixed)	-61.83	23.96	-85.50	-16.74	0.99
Income Tax	-7.00	3.87	-11.68	-1.97	0.95
Eq. Magnitude (Agr.)	0.66	0.18	0.43	0.89	1.00
Eq. Magnitude (Ind.)	1.83	0.24	1.52	2.13	1.00
Eq. Magnitude (Mixed)	4.27	1.21	2.03	5.41	1.00
Urban	-0.89	1.49	-3.13	0.74	0.71

Note: 200000 iterations with 100000 iterations discarded at the beginning.
80% credible intervals in parenthesis. All R-Hat statistics below critical levels.
Standard convergence diagnostics suggest good mixing and convergence.
Year fixed effects, latitude and longitude were omitted in the table.
A total of 4 chains were run.

Table 1: *Poisson Regression: Simulated Posterior Predictions*

I. Appendix

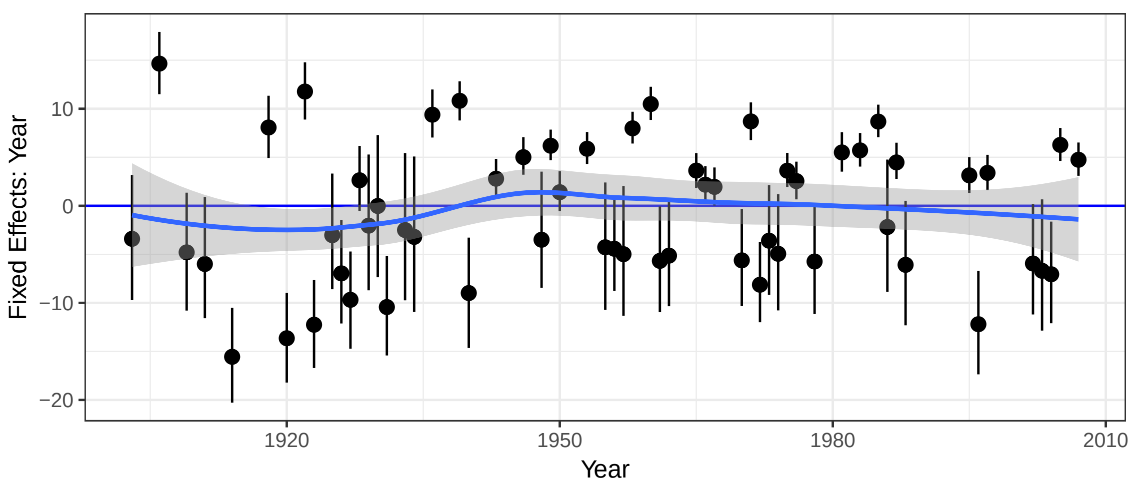

Figure A1: *Year Fixed Effects*

Figure A2: *Assessing Model Fit*

............................**Word count**: 585 ....

References

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.