Replication: ‘The Distributive Politics of Enforcement’

Beau Meche

2/21/2020

1 Abstract:

“The Distributive Politics of Enforcement” by Alisha Holland (2014) analyzes electoral behavior’s relationship with police action in opposition to low-income unlicensed street vendors in three cities in Latin America. I was mostly successful in replicating the results, with minute variance due to apparent differences between regression output between R and Stata. In my extenstion I re-regressed the models from the original paper under Bayesian modeling methods in the interest of discovering any differences likely to arise. The regression outputs themselves were quite similar and model comparisons favored similar models to the author; however upon cross-validation model analysis I found that a majority of the models were laden with ‘problematic’ values. This implies that the models showing statistically significant support for the author’s claim do not effectively model the original dataset should any one value be removed and further implies that caution should be taken with associated claims pending better modeling or more data.

2 Introduction and Overview:

This paper(Holland 2014) looks at police behavior in varying contexts around Latin America to explore what gets prosecuted and what is allowed to go unnoticed. The areas of interest are the relationships between constituency of the powers that be (at a given time) [fig. 1] and how much those individuals in the constituent groups care to have property rights enforced. Involving a great deal of original research, the author does an intriguing job of addressing this question in a way that I think could be complemented by additional Bayesian analysis, so I thought this would be a good project on a topic that I briefly studied previously and can now apply new skills to.

The author argues that in many cases, not enforcing these laws is as intentional as enforcing them and makes an early assertion of one of her findings where she found up to a five times higher enforcement rate in places where constituents were primarily non-poor (a binary distinction make in the data analysis portion of the research) [fig.1]. Holland uses many standardized poisson regression models [fig. 2], which makes sense in the setting where response variables are the topic of interest. The catch in that model is the concern for invalidation by reverse causality. To counter this point, a supply vs. demand framework is raised as one might find in your favorite introductory economics course with the y-axis showing enforcement frequency and the x-axis showing the count of offenses. The key to this argument is in assuming that the intentional non-enforcement cases would translate to an outward shift in the 'supply curve' of enforcement. This framework gives a solid basis to analyze electoral and otherwise political behavior and regional differences. The results of the data and the proposed model seem to support this framework and initial intuition, even if it is specific to the case observed. Becuase of the specificity in political frameworks inquestion here, I would have concerns about the external validity of Holland's models, but nonetheless I agree that the model results in support for the framework claiming that there is an inverse relationship between police enforcement of vending licensing and quantity of poor individuals voting in a district.

3 Literature Review:

The author of this paper has become relatively well known for her research and her findings in the realm of forbearance and electoral impacts relative to low income populations in Latin America. In the literature published after this particular paper in 2014, a majority of it has also been authored by Holland and so far as I can tell there are few contradictions to be seen. The setting of the cases observed in this research is one where vendors of various goods are 'guilty' of using public space to peddle their private enterprises. This gives the police forces the ability to, within the letter of the law, arrest or otherwise enforce the statute that they are unlicensed to the space in question and that they should vacate. The reality tangent to this is that these vendors are also voters, and they are unlikely to forget the political head of the forces inhibiting their livelihood. Holland cites Colombia's National Data Archive with respect to the aspect that these vendors often do tend to rely on thier 'unlicensed incomes' as they are usually quite low-ranking in the income distribution.

Without digressing into Holland's research or the other supporting research again, it seems intuitive that where there is a majority of voters comprised of these individuals in the lower income distribution range who may rely, or know someone who relies, on secondary income to sustain themselves you would find that the local government would be less harsh on these cases of unregistered free markets. In fact, since this paper was published other authors have found cases where the government actually incentivizes the collaboration or unionization of these unregistered individuals as a sort of subset of the economy (Hummel 2017).

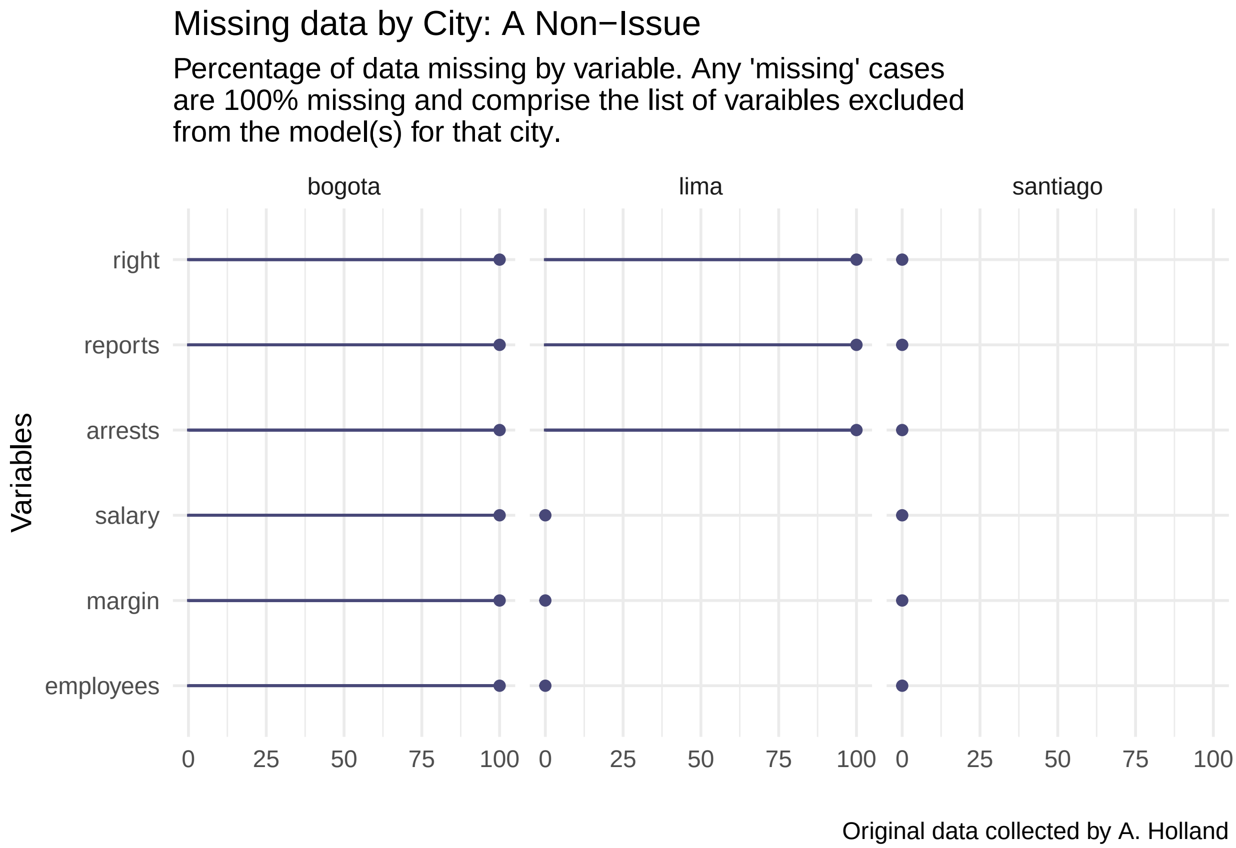
Overall this vein of inquiry has fascinating ramifications in electoral predictions and voter behavior, especially in cases where statistical analyses show decreased sensitivities to potentially exogenous enforcement factors like department budgets and other angles. Carefully constructed models slowly begin to point their fingers toward constituency as a large driver of enforcement behaviors; and while I would prefer to see a few more rigors in the process, I tend to be convinced by the findings available thus far.

4 Extension:

Fortunately, I believe that I was able to replicate the major findings from the original paper. In effort to further explore[1] the findings from Holland's work, the same poisson models that comprise 'Figure 1' were run through models built with intrinsic Bayesian methods. Because Bayesian inference and regression methods present different types of interpretation of uncertainty and offer a different analytical approach, these results have been compared to the original as a kind of sensitivity analysis with respect to statistical methodology.

The first exploration was with respect to missing data. In short, there was no place for concern for missing data. Whether this is due to the pre-cleaned data being the only repository made available, the nature of the original research lending a low rate of missing values, or other factors is unknown. It was found to be the case that there were complete records funneled into all models constructed in the original study, and the only 'missing' values were variables completely present or missing at the city level.

[1]"All analysis and output code for this replication project as well as the original paper are available here.



The second exploration was with regard to the Bayesian regression coefficients and whether they were notably different from the originals. On the whole, the coefficients seemed to be quite similar. The signs of the coefficients were certainly correct, and the actual estimates were either identical or close enough to attribute to rounding error or slight differences in statistical modeling across the transition from the original language (Stata) to the one used in this analysis (R).

The third inquiry was a test of the fit of the models. Using the leave-one-out cross validation method of analyzing the Bayesian regressions, there were between 3 and 7 observations per model in the regressions for Lima and Santiago that were deemed to be ‘problematic’ by this analysis. What this tells us is, when point x is excluded from the data and the model re-fit, how well the model would predict said point. A probelmatic value then would imply cases where this extreme-case k-fold analysis indicates a poor prediction of the ommitted point by the model at hand. In small data sets with any level of variance, it seems reasonable at first to assume that there would be a few outlying values that would not be great fits to our model. With that said, too many problematic observations could become worrisome.

The final inquiry was with regard to ‘statistical significance’ and its equivalent in Bayesian inference through rstanarm (Gabry 2020) . This portion of the analysis was done by juxtaposing the frequestist confidence intervals with the Bayesian credible intervals for the model parameters. All confidence, credible intervals were evaluated at the 95% significance level. As Bayesian analyses do not involve p-values and t-distributions, but rather posterior distributions of parameter estimates, I found it convenient to compare the intervals between the two model methodologies for a comparison of statistical significance

- Figure 2

Table 1: Replicated Models

	Dependent variable:								
	operations								arrest_T
	Bogota		Lima				Santiago		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lower	−0.053	−0.693***	−0.641***	−0.975***	−0.707***	−0.673***	−0.767***	−0.449***	0.047***
	(0.087)	(0.052)	(0.054)	(0.085)	(0.166)	(0.165)	(0.218)	(0.156)	(0.004)
Vendors	0.571***	−0.182***	−0.162**	−0.090	−0.528*	−0.683**	−0.754**	−0.496**	
	(0.068)	(0.064)	(0.064)	(0.066)	(0.290)	(0.336)	(0.364)	(0.230)	
Budget	0.038	0.060**	0.120***	0.087***	−0.149	0.096	0.138	−0.177	−0.065***
	(0.127)	(0.027)	(0.029)	(0.031)	(0.122)	(0.130)	(0.145)	(0.115)	(0.004)
Population	0.207**	0.468***	0.467***	0.368***	0.272*	0.586***	0.635***	0.406**	0.096***
	(0.087)	(0.054)	(0.056)	(0.060)	(0.165)	(0.186)	(0.203)	(0.181)	(0.002)
Margin			0.166***	−0.188**		−0.680***	−0.868***		
			(0.034)	(0.078)		(0.160)	(0.330)		
Margin*Lower				0.425***			0.292		
				(0.085)			(0.436)		
Right								2.083***	0.092***
								(0.412)	(0.005)
Reports									0.400***
									(0.002)
Constant	1.987***	2.920***	2.899***	2.876***	0.665***	0.461***	0.478***	−0.826**	8.473***
	(0.091)	(0.043)	(0.044)	(0.045)	(0.139)	(0.169)	(0.169)	(0.388)	(0.004)
Observations	19	36	36	36	34	34	34	34	34
Log Likelihood	−41.918	−231.254	−220.184	−207.689	−108.395	−97.262	−97.035	−88.994	−4,631.258
Akaike Inf. Crit.	93.837	472.509	452.369	429.378	226.791	206.524	208.069	189.987	9,274.516

Note: *p<0.1; **p<0.05; ***p<0.01

- Figure 3

With regard to missing data analysis, there were no points of concern as there were no regression analyses that incorporated missing data. below you can see the summary of data missing from the onset, but at the level of analyses, the individual cities, there were no calculations done on variables with incomplete sets of information.

5 Extension: pt.2

(a) The object of this portion is to compare modeling methods between standard poisson regressions and those done in the Bayesian method. Errors, estimate measures, and fit metrics are slightly different, so we will see whether this methodological change brings up points of concern

(1) Figure 4: Bogota and Lima:

Table 2: Bogota and Lima

	(B 1)	(L 1)	(L 2)	(L 3)
Intercept	1.970	2.917	2.898	2.872
	(0.093)	(0.042)	(0.045)	(0.046)
Lower	-0.050	-0.695	-0.640	-0.974
	(0.086)	(0.049)	(0.052)	(0.084)
Vendors	0.571	-0.182	-0.162	-0.091
	(0.067)	(0.065)	(0.064)	(0.067)
Budget	0.023	0.060	0.119	0.087
	(0.133)	(0.027)	(0.030)	(0.030)
Population	0.206	0.468	0.465	0.367
	(0.088)	(0.055)	(0.055)	(0.061)
Margin			0.164	-0.188
			(0.034)	(0.076)
Margin*Lower				0.425
				(0.083)
Num. Observations	19	36	36	36

B1, L1-3 indicate regression country / model number from the replicated regression model table

(2) Figure 5: Santiago:

Table 3: Santiago

	(1)	(2)	(3)	(4)	(5)
Intercept	0.628	0.413	0.419	-0.853	8.472
	(0.140)	(0.171)	(0.179)	(0.386)	(0.004)
Lower	-0.708	-0.671	-0.761	-0.445	0.046
	(0.168)	(0.167)	(0.215)	(0.162)	(0.004)
Vendors	-0.562	-0.724	-0.777	-0.536	
	(0.280)	(0.328)	(0.349)	(0.242)	
Budget	-0.154	0.094	0.132	-0.183	-0.065
	(0.127)	(0.133)	(0.145)	(0.116)	(0.004)
Population	0.275	0.594	0.640	0.406	0.096
	(0.165)	(0.183)	(0.202)	(0.185)	(0.002)
Margin		-0.690	-0.867		
		(0.159)	(0.324)		
Margin*Lower			0.264		
			(0.439)		
Num. Observations	34	34	34	34	34

5.1 LOO-Comparisons:

(b) This section contains leave-one-out cross-validation model comparisons between the Bayesian models from part 'a' for cities with multiple regressions in the original analysis. For those unfamiliar, the favored model is shown with zero values for 'elpd_diff' and 'se_diff'. The preferred models in the LOO-compared perspective are the models for which I provide density overlay plots below in part 'd'.

5.1.1 Lima, Peru:

```
##
## ============================================================================
##         elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## ----------------------------------------------------------------------------
## limreg2     0         0    -242.086   26.024    31.147  5.935   484.172  52.049
## limreg1  -8.703   15.284  -250.788   25.941    28.476  5.255   501.577  51.882
## ----------------------------------------------------------------------------
```

```
##
## ============================================================================
##         elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## ----------------------------------------------------------------------------
## limreg3     0         0    -238.013   28.228    41.285  10.375  476.025  56.455
```

```
## limreg2  -4.073   16.514  -242.086   26.024    31.147  5.935   484.172  52.049
## -------------------------------------------------------------------------------
```

```
##
## ===============================================================================
##         elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## -------------------------------------------------------------------------------
## limreg3     0        0    -238.013   28.228    41.285  10.375  476.025  56.455
## limreg1  -12.776  18.459  -250.788   25.941    28.476  5.255   501.577  51.882
## -------------------------------------------------------------------------------
```

5.1.2 Santiago, Chile:

```
##
## ================================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## --------------------------------------------------------------------------------
## santreg2     0        0    -122.374   25.601    33.859  10.579  244.748  51.203
## santreg1  -7.233   13.801  -129.607   32.166    29.367  10.907  259.214  64.332
## --------------------------------------------------------------------------------
```

```
##
## ================================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## --------------------------------------------------------------------------------
## santreg3     0        0    -125.302   26.110    36.828  10.767  250.605  52.220
## santreg1  -4.305   13.750  -129.607   32.166    29.367  10.907  259.214  64.332
## --------------------------------------------------------------------------------
```

```
##
## ================================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## --------------------------------------------------------------------------------
## santreg4     0        0    -112.739   23.864    31.278  9.680   225.477  47.727
```

```
## santreg1  -16.869  13.700  -129.607   32.166    29.367  10.907  259.214  64.332
## -------------------------------------------------------------------------------

##
## ===============================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## -------------------------------------------------------------------------------
## santreg2     0        0    -122.374   25.601    33.859  10.579  244.748  51.203
## santreg3  -2.928    1.656  -125.302   26.110    36.828  10.767  250.605  52.220
## -------------------------------------------------------------------------------

##
## ===============================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## -------------------------------------------------------------------------------
## santreg4     0        0    -112.739   23.864    31.278  9.680   225.477  47.727
## santreg2  -9.635   14.144  -122.374   25.601    33.859  10.579  244.748  51.203
## -------------------------------------------------------------------------------

##
## ===============================================================================
##          elpd_diff se_diff elpd_loo se_elpd_loo p_loo  se_p_loo  looic  se_looic
## -------------------------------------------------------------------------------
## santreg4     0        0    -112.739   23.864    31.278  9.680   225.477  47.727
## santreg3  -12.564  14.299  -125.302   26.110    36.828  10.767  250.605  52.220
## -------------------------------------------------------------------------------
```

5.2 Bayesian - Frequentist Fit Comparisons:

(c) In this section, the first stat line contains the model coefficients for the indicated model run with frequentist model construction and the subsequent stat lines show 95% credible intervals for the Bayesian version of the same model.

5.2.1 Bogota Model #1

```
##               2.5 % 97.5 %
## (Intercept)  1.8025  2.159
## slower      -0.2258  0.116
## svendors     0.4373  0.704
## sbudget     -0.2373  0.267
## spop         0.0339  0.376
```

```
##                2.5%  97.5%
## (Intercept)  1.7842 2.1395
## slower      -0.2273 0.1183
## svendors     0.4317 0.7027
## sbudget     -0.2604 0.2614
## spop         0.0270 0.3713
```

5.2.2 Lima Model #1

```
##                2.5 % 97.5 %
## (Intercept)  2.83397  3.003
## slower      -0.79601 -0.593
## svendors    -0.30724 -0.057
## sbudget      0.00639  0.111
## spop         0.36095  0.574
```

```
##                2.5%   97.5%
## (Intercept)  2.8304  3.0031
## slower      -0.7930 -0.5949
```

```
## svendors    -0.3117 -0.0512
## sbudget      0.0054  0.1098
## spop         0.3531  0.5740
```

5.2.3 Lima Model #2

```
##               2.5 % 97.5 %
## (Intercept)  2.8111  2.984
## slower      -0.7484 -0.536
## svendors    -0.2876 -0.038
## sbudget      0.0610  0.177
## spop         0.3573  0.575
## smargin      0.0977  0.233
```

```
##               2.5%  97.5%
## (Intercept)  2.807  2.980
## slower      -0.746 -0.534
## svendors    -0.287 -0.040
## sbudget      0.058  0.177
## spop         0.354  0.572
## smargin      0.095  0.235
```

5.2.4 Lima Model #3

```
##               2.5 %  97.5 %
## (Intercept)  2.7858  2.9627
## slower      -1.1422 -0.8078
## svendors    -0.2202  0.0395
## sbudget      0.0244  0.1459
## spop         0.2506  0.4840
## smargin     -0.3420 -0.0348
## s_int_ML     0.2587  0.5912
```

```
##               2.5%  97.5%
```

```
## (Intercept)  2.782  2.959
## slower      -1.135 -0.814
## svendors    -0.222  0.034
## sbudget      0.024  0.144
## spop         0.254  0.482
## smargin     -0.337 -0.034
## s_int_ML     0.254  0.584
```

5.2.5 Santiago Model #1

```
##              2.5 %  97.5 %
## (Intercept)  0.374  0.9226
## slower      -1.037 -0.3831
## svendors    -1.174 -0.0260
## sbudget     -0.388  0.0928
## spop        -0.049  0.6057
```

```
##               2.5%  97.5%
## (Intercept)  0.336  0.889
## slower      -1.047 -0.376
## svendors    -1.180 -0.056
## sbudget     -0.408  0.084
## spop        -0.043  0.618
```

5.2.6 Santiago Model #2

```
##               2.5 % 97.5 %
## (Intercept)  0.0995  0.767
## slower      -1.0008 -0.351
## svendors    -1.4352 -0.111
## sbudget     -0.1585  0.354
## spop         0.2326  0.974
## smargin     -1.0086 -0.381
```

```
##               2.5%  97.5%
## (Intercept)  0.064  0.720
## slower      -1.011 -0.363
## svendors    -1.448 -0.160
## sbudget     -0.151  0.345
## spop         0.256  0.962
## smargin     -1.019 -0.381
```

5.2.7 Santiago Model #3

```
##              2.5 % 97.5 %
## (Intercept)  0.116  0.783
## slower      -1.203 -0.347
## svendors    -1.580 -0.141
## sbudget     -0.144  0.428
## spop         0.254  1.064
## smargin     -1.565 -0.266
## s_int_ML    -0.551  1.162
```

```
##               2.5%  97.5%
## (Intercept)  0.055  0.730
## slower      -1.181 -0.344
## svendors    -1.547 -0.178
## sbudget     -0.160  0.413
## spop         0.255  1.046
## smargin     -1.524 -0.263
## s_int_ML    -0.573  1.074
```

5.2.8 Santiago Model #4

```
##               2.5 %  97.5 %
## (Intercept) -1.6889 -0.1466
## slower      -0.7591 -0.1451
## svendors    -1.0113 -0.0874
```

```
## sbudget     -0.4028  0.0498
## spop         0.0473  0.7606
## right        1.3408  2.9803
```

```
##               2.5%  97.5%
## (Intercept) -1.709 -0.172
## slower      -0.759 -0.138
## svendors    -1.069 -0.116
## sbudget     -0.409  0.052
## spop         0.049  0.752
## right        1.327  2.973
```

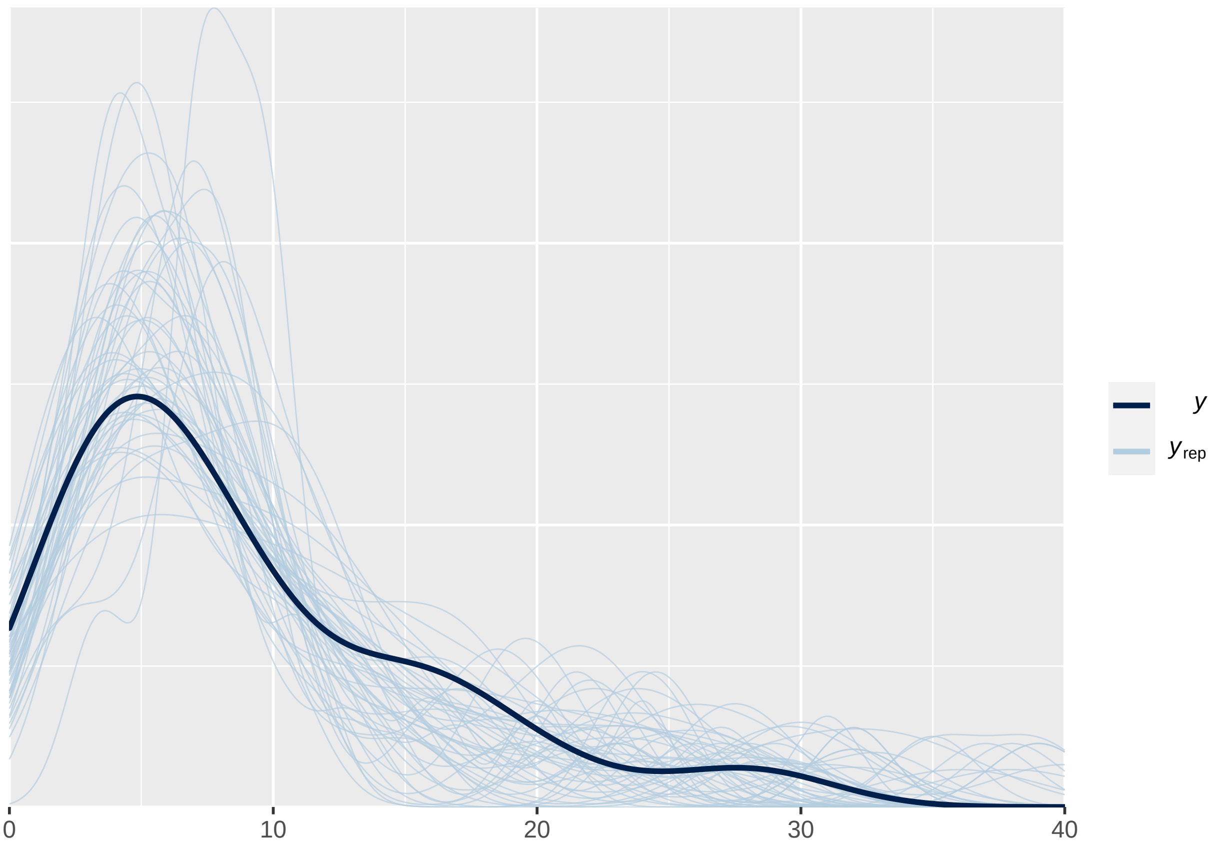
5.2.9 Santiago Model #5

```
##               2.5 %  97.5 %
## (Intercept)  8.4649  8.4802
## slower       0.0381  0.0550
## sbudget     -0.0734 -0.0563
## spop         0.0926  0.1000
## right        0.0813  0.1027
## sreports     0.3961  0.4044
```

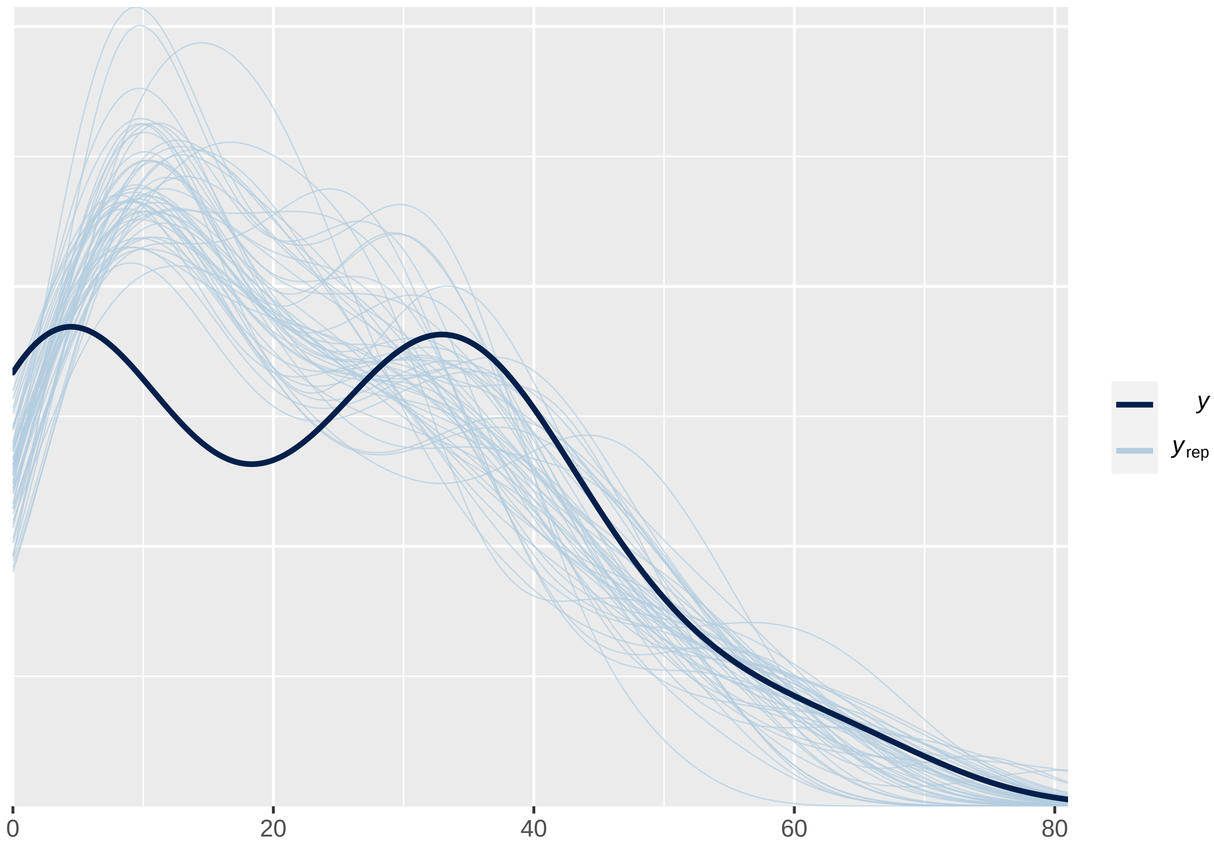
```
##               2.5%  97.5%
## (Intercept)  8.465  8.480
## slower       0.038  0.055
## sbudget     -0.073 -0.057
## spop         0.093  0.100
## right        0.081  0.103
## sreports     0.396  0.404
```

6 Model checks done on preferred competing models.

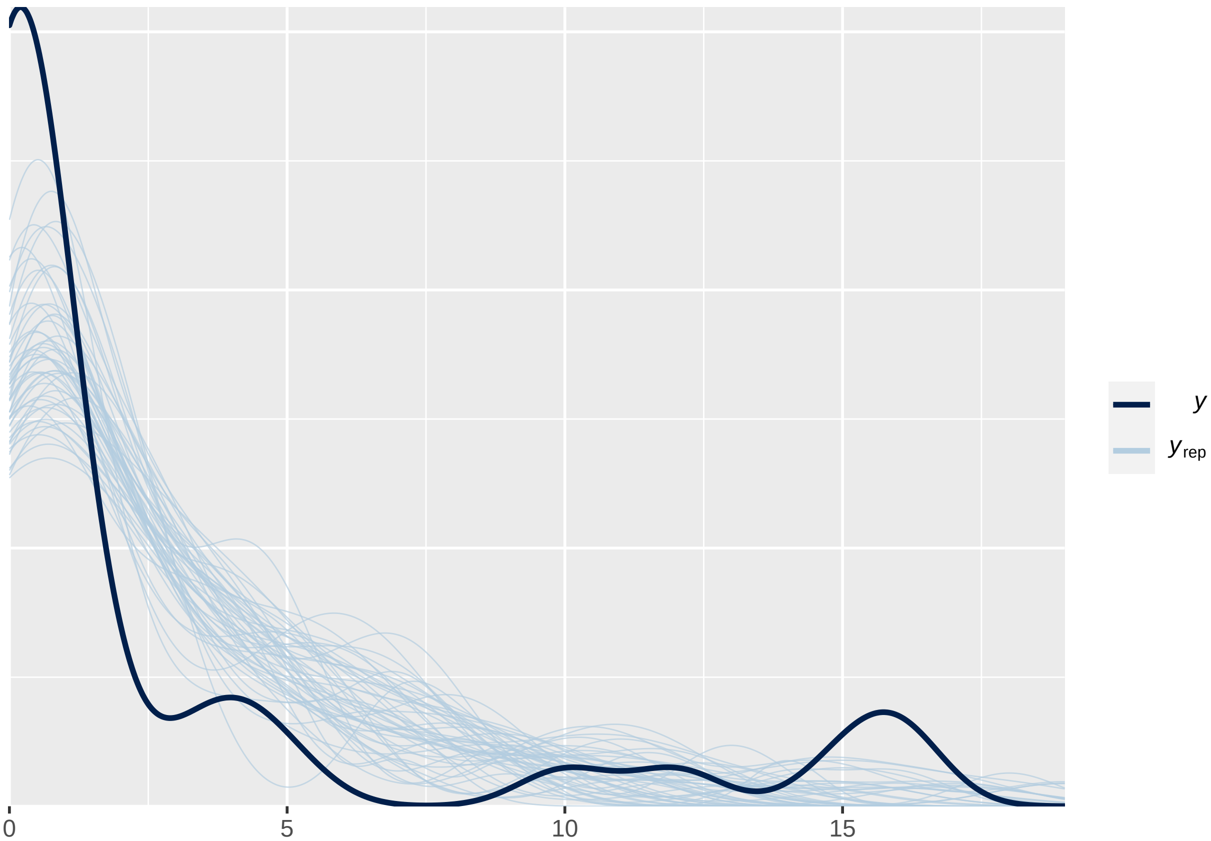
6.1 Bogota: Regression 1



6.2 Lima: Regression 3

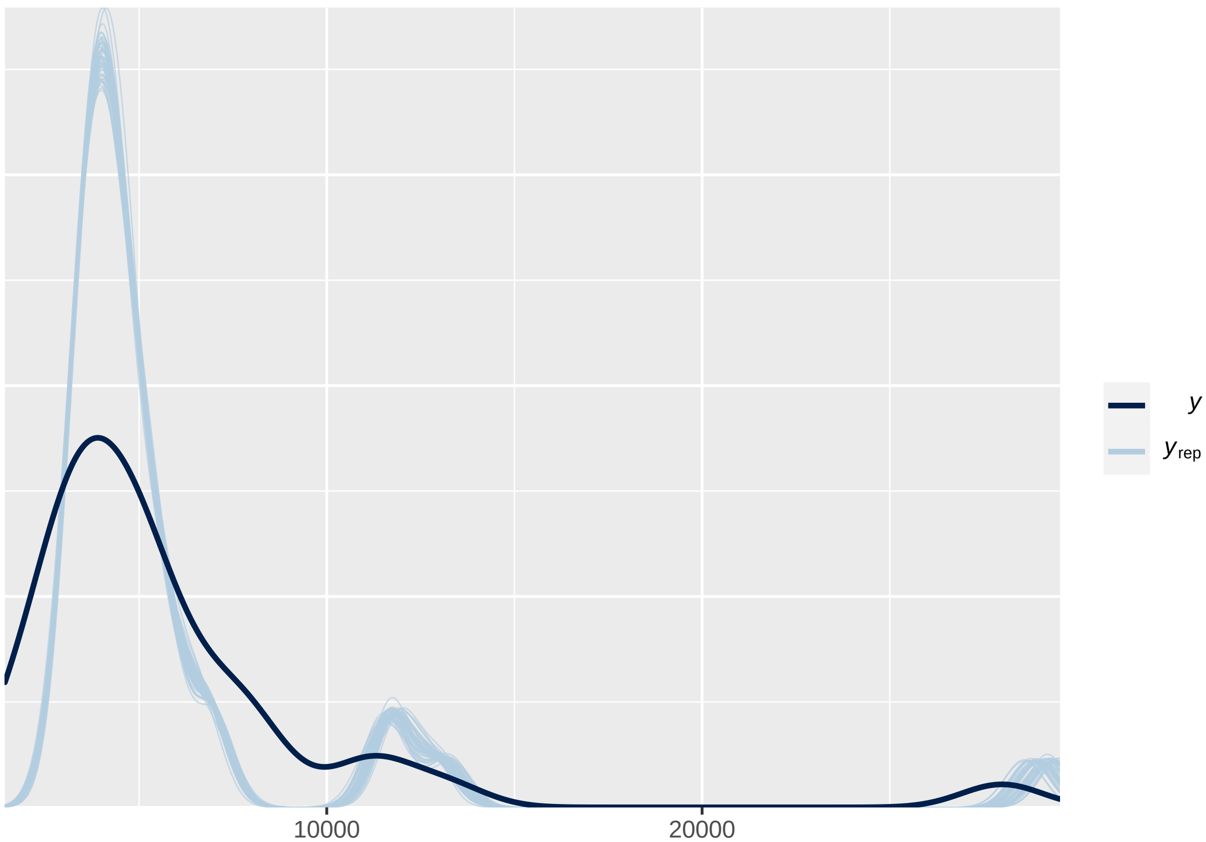


6.3 Santiago: Regression 4



6.4 Santiago: Regression 5

- Regression 5 is a separate model construction and thus was not compared to models 1-4.



Gabry, Jonah. 2020. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://cran.r-project.org/web/packages/rstanarm/index.html.

Holland, Alisha C. 2014. "The Distributive Politics of Enforcement." https://doi.org/10.1111/ajps.12125.

Hummel, Calla. 2017. "Disobedient Markets: Street Vendors, Enforcement, and State Intervention in Collective Action." *Comparative Political Studies* 50 (11): 1524–55. https://doi.org/10.1177/0010414016679177.