Women and The Environment

A Bayesian Logistic Regression Analysis

Kaylee Hodgson

Brigham Young University

December 10, 2018

Introduction

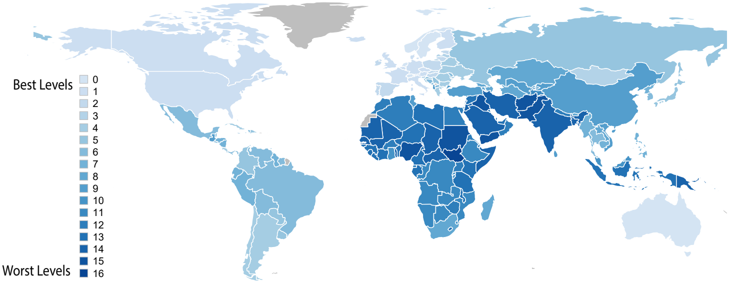
Subordination of women in the home seems to be indicative of a violent culture that does not treat its valuable assets well, including the environment.

Purpose: identify whether higher levels of gender inequality in the household are significantly tied to poor environmental wellbeing outcomes

The Syndrome Scale

Combines 11 indicators of the subordination of women in the home:

- prevalence of patrilocal marriage
- prevalence of brideprice or dowry
- prevalence and legality of polygyny
- presence of counsin marriage
- age of marriage for girls
- laws and practices surrounding women's property rights
- presence of son preferences or sex ratio alteration
- presence of inequity in family law/custom that favors males
- overall level of violence against women in society
- presence of societal sanction for femicide
- whether there is legal exoneration for rapists who offer to marry their victims



Control Variables

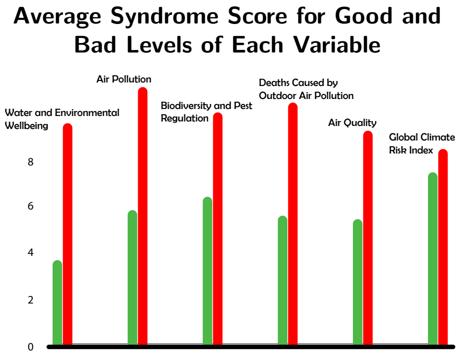
- percentage of the population that lives in urban areas (The World Bank)
- aggregated civilization identification (based on Samuel Huntington's civilizational)
- colonial heritage - dichotomous, indicates whether a country was colonized (from Valerie Hudson and Donna Lee Bowen's scale in their forthcoming book)
- percentage of land that is arable (The World Bank)
- number of unique land neighbors (Wikipedia)
- level of ethnic fractionalization (Alesina et al., 2003)
- level of religious fractionalization (Alesina et al., 2003).

Standardizing Variables for Analysis

- Standardization of independent variables following the methods put forward in Gelman et al. (2014).
- Scale all to have mean=0, standard deviation=0.5
- Dichotomous variable (colonization status): find proportion of 0's (0.13) and proportion of 1's (0.87), then redefining the 0's as 0.87 and the 1's as -0.13.
- Categorical variable (civilization): create 3 dichotomous variables which indicated 3 of the levels, then standardized each of those in the same way as colonial status

Environmental Indicators

- Water and Environmental Wellbeing
- Air Pollution
- Biodiversity and Pest Regulation
- Deaths caused by outdoor air pollution
- Air Quality
- Global Climate Risk Index



Each variable is dichotomized where 0 indicates “good” levels and 1 indicates “bad” levels

Preliminary Model

Hierarchical Bayesian Model

$$
\begin{aligned}
&\text{Level 1: } y_{im}|p_{im} \sim \text{ Bernoulli}(p_{im}), \\
&\text{logit}(p_{im}) = \beta_{m0} + \sum_{j=1}^{7} x_{imj}\beta_{mj} + \sum_{k=1}^{3} x_{imk}\alpha mk + \epsilon_{im} \\
&\text{Level 2: } \beta_0 \sim \text{ Cauchy}(0,\sigma),\ \beta_j \sim \text{ Cauchy}(0,\theta),\ \alpha_k \sim \text{ Cauchy}(\mu,\phi) \\
&\text{Level 3: } \sigma \sim \text{ Uniform}(5,15), \theta \sim \text{ Normal}^{+}(2.5,1), \mu \sim \text{ Gamma}(3,.5)
\end{aligned}
\tag{1}
$$

- Not enough data to estimate the parameters on multiple levels
- Covariance matrix of the draws is not positive definite

The Model

$$
\begin{aligned}
y_{im}|p_{im} \sim & \text{ Bernoulli}(p_{im}), \\
\text{logit}(p_{im}) = & \beta_{m0} + \sum_{j=1}^{7} x_{imj}\beta_{mj} + \sum_{k=1}^{3} x_{imk}\alpha mk + \epsilon_{im} \\
\beta_0 \sim & \text{ Cauchy}(0, \sigma),\ \beta_j \sim \text{ Cauchy}(0, \theta),\ \alpha_k \sim \text{ Cauchy}(\mu, \phi) \\
\sigma = & 5, \theta = 2.5, \mu = \phi = 2
\end{aligned}
\tag{2}
$$

$i = 1, ..., n$ corresponds with the country (or observation), $m = 1, ..., 6$ corresponds with the regression model (one model for each environmental variable), $j = 1, ..., 7$ and $k = 1, ..., 3$ corresponds with the independent variables.

Prior Justification

Cauchy distribution chosen over the normal or the t distributions because the Cauchy is more flexible with extreme values.

- β priors are non-informative
- α priors correspond with the variables the I have more prior information for: Terrain, Urbanization, and Syndrome

Posterior Distribution

The posterior distribution is given, up to proportationality:

$$p(.|y) \propto p(\beta_0, \beta_1, ..., \beta_6, \alpha_1, ..., \alpha_4, \sigma, \theta, \mu, y) \propto$$
$$\left[\prod_{i=1}^{n} \left(\frac{e^{X_i\beta}}{1+e^{X_i\beta}}\right)^{y_i} \left(\frac{1}{1+e^{X_i\beta}}\right)^{1-y_i}\right] \left[\frac{1}{\pi\sigma\left[1+(\beta_0/\sigma)^2\right]}\right] \quad (3)$$
$$\left[\prod_{j=1}^{6} \frac{1}{\pi\theta\left[1+(\beta_i/\theta)^2\right]}\right] \left[\prod_{k=1}^{4} \frac{1}{\pi\phi\left[1+((\alpha_i-\mu)/\phi)^2\right]}\right]$$

Computational Methods: Metropolis-Hastings

Markov chain Monte Carlo with multivariate proposal scheme:

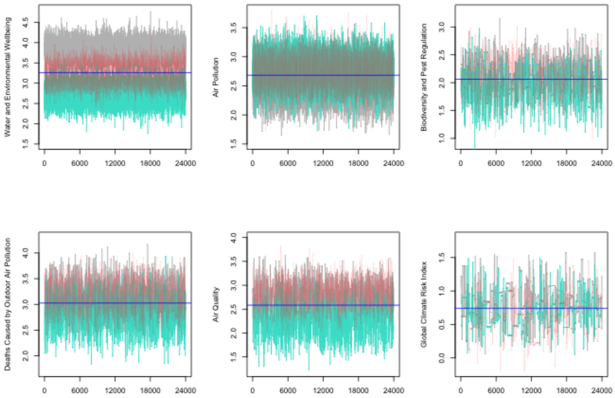
- Set initial values for the parameters.
- Use normal approximation to obtain the hessian estimate for the covariance matrix ($\mathbf{S}$) and the optimized parameter estimates for the mean vector ($\mathbf{m}$).
- For each iteration:

1. Set the proposal: $\mathbf{P} = \mathbf{m} + (\text{chol}(\mathbf{S}))'\text{rnorm}(11)$.
2. Use a Metropolis-Hastings update to determine whether to keep the last samples or update with the proposal, using the proposal density: $f(\mathbf{P}) = -0.5(\mathbf{P} - \mathbf{m})'\mathbf{S}^{-1}(\mathbf{P} - \mathbf{m})$.

Computational Methods: Metropolis-Hastings

- 3 chains, 60,000 draws each
- initial values: all 0, mean estimates from priors, frequentist estimates
- burn-in: 20,000, thinning: every fifth draw
- CPU: under 5 seconds for each chain

Model Checks



Posterior draws for Syndrome coefficient for each regression model

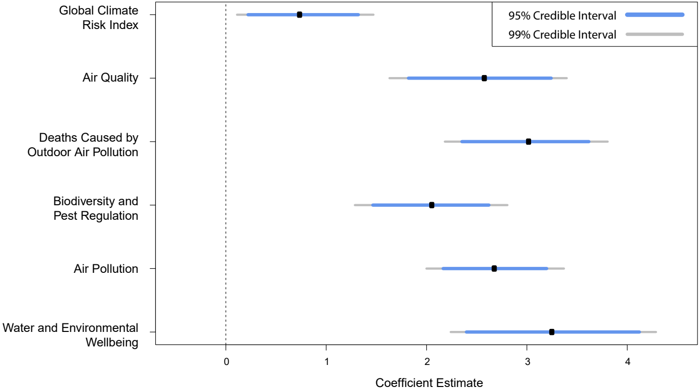
Model Checks

Diagnostics for Syndrome coefficient

	Syndrome Coefficient Diagnostics		
Model	Acceptance Rate	Effective Sample Size	$\hat{R}$
Water and Environmental Wellbeing	0.34	53.58	2.77
Air Pollution	0.34	2615.55	1.10
Biodiversity and Pest Regulation	0.04	666.72	1.12
Deaths Caused by Outdoor Air Pollution	0.10	667.66	1.31
Air Quality	0.09	357.73	1.67
Global Climate Risk Index	0.01	325.19	1.07

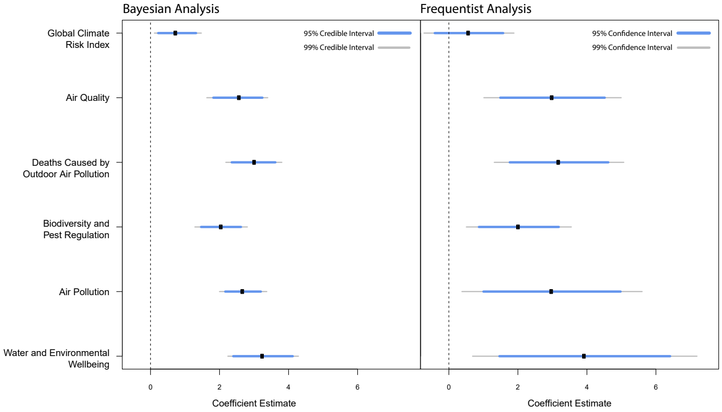
Syndrome Estimate Results

Syndrome Coefficient Estimates and Credible Intervals for each Environmental Model



Frequentist Approach

Comparison of Syndrome Estimates in Bayesian and Frequentist Analysis



Model Sensitivity

Less Informative Model

$$\beta_0 \sim \text{Cauchy}(0, \sigma),\ \beta_j \sim \text{Cauchy}(0, \sigma),\ \alpha_k \sim \text{Cauchy}(0, \sigma) \quad (4)$$

$\sigma = 5$

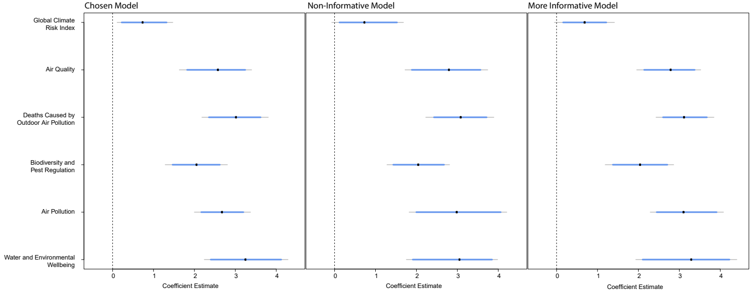
More Informative Model

$$\beta_0 \sim \text{Normal}(0, \sigma),\ \beta_j \sim \text{Normal}(0, \theta),\ \alpha_k \sim \text{Normal}(\mu, \phi) \quad (5)$$

$\sigma = 5, \theta = 2.5, \mu = \phi = 2$

Model Sensitivity Results

Comparison of Syndrome Estimates In Sensitivity Analyses



Model Sensitivity Diagnostics

Diagnostic Comparisons for Syndrome coefficient in Sensitivity Analyses:
M1=Chosen Model, M2=Less Informative Model, M3=More Informative Model

	Acceptance Rate			Effective Sample Size			R-hat		
Model	M1	M2	M3	M1	M2	M3	M1	M2	M3
Water & Env. Wellbeing	0.34	0.35	0.34	**53.58**	22.33	22.52	3.36	**1.00**	2.13
Air Pollution	0.34	0.35	0.29	**2615.55**	22.73	123.23	**1.32**	5.00	3.73
Biodiv. and Pest Reg.	0.04	0.04	0.04	**666.72**	474.21	426.77	**1.15**	1.39	1.45
Deaths from O.A.P.	0.10	0.11	0.11	667.66	429.87	**2120.84**	**1.03**	1.37	1.05
Air Quality	0.09	0.10	0.09	357.73	112.37	**577.40**	**1.01**	1.55	1.05
GCRI	0.01	0.02	0.01	325.19	199.05	**407.38**	1.42	2.24	**1.26**

Discussion and Conclusion

- Results confirm the hypothesis
- The results for Global Climate Risk Index model should be interpreted more cautiously
- Future research: collect more data for hierarchical approach, improve convergence and model fit

Overall, the results of this analysis indicate that the subordination of women in the household is a significant determinant of a country's environmental performance.