Supplemental Information for: Turning the Lights on to Keep Them in the Fold: How Governments Preempt Secession Attempts

Rob Williams*

April 8, 2021

Contents

*Postdoctoral Research Fellow, Weidenbaum Center on the Economy, Government, and Public Policy, Washington University in St. Louis, rob.williams@wustl.edu, jayrobwilliams.com.

A Descriptive Statistics

Figure A.1 presents descriptive statistics for all predictors included in the various models. Due to their skewed distributions, nightlights, population, capital distance, area, and GDP per capita are log-transformed. Figure A.1 depicts these transformed distributions.

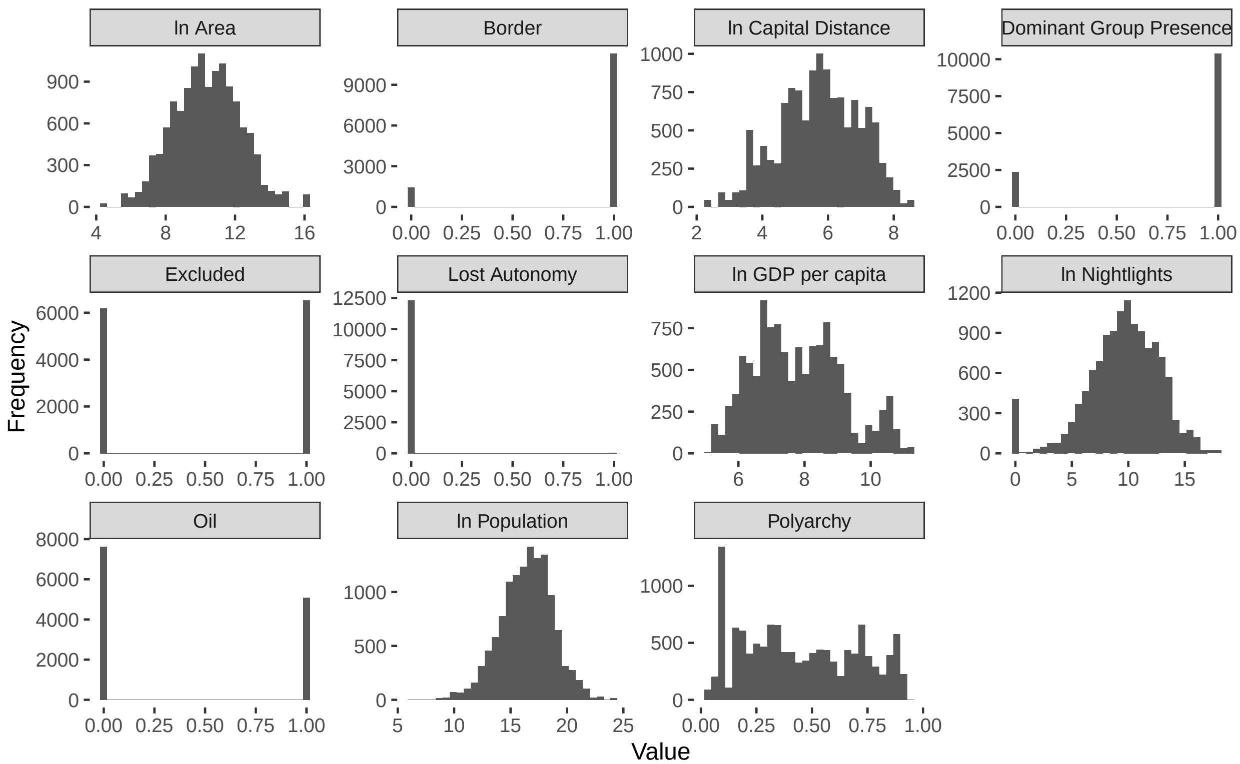
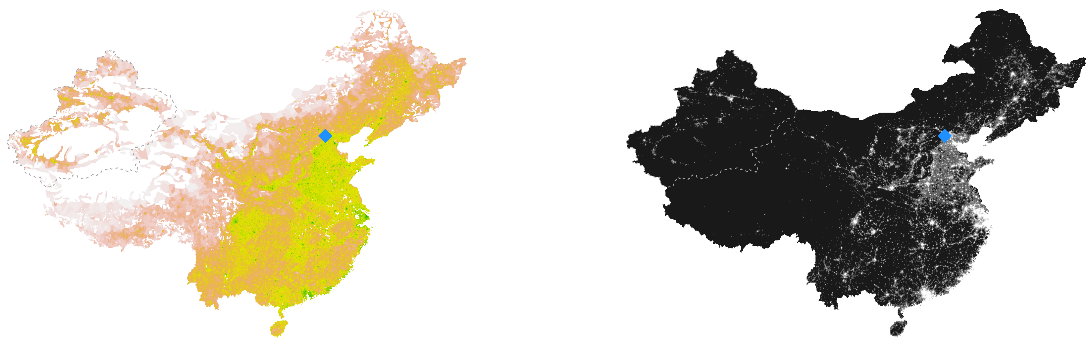


Figure A.1: Descriptive statistics for predictors included in analysis. Demographic balance, horizontal inequality, GDP, population density, nightlights, accessibility, and area are log transformed.

B Nightlights Considerations

One of the main downsides of the DMSP OLS data is that they are unable to distinguish variation within urban areas where light levels are high due to saturation from neighboring pixels (Hsu, Baugh, Ghosh, Zhizhin & Elvidge 2015). In these cases, all pixels in a saturated area receive the maximum value. This phenomenon can be clearly seen in the area around Beijing in Figure B.1b. Luckily, I am interested in variation between entire ethnic group territories, not within individual cities, so this is less problematic for my analyses.



(a) Population (b) Nightlights

Figure B.1: China in 2013. Panel (a) displays (log) population and Panel (b) displays nightlights. The gray dashed line denotes the Xinjiang Uyghur Autonomous Region, while Beijing is represented by the blue diamond.

Using the 'cookie cutter' approach (Cederman, Buhaug & Rød 2009, Cederman, Weidmann & Gleditsch 2011, Cederman, Weidmann & Bormann 2015) requires correcting for cells where multiple group territories overlap. I do this by dividing the cell value by the number of group polygons that cover it for each cell in the raster data. For example, a substantial portion of the Syrian Kurds' settlement area overlaps with areas inhabited by Sunni Arabs. Each raster cell in these areas has its nightlights value divided by 2 before aggregation to the group level, so the Kurds and the Sunnis each receive half of the cell's nightlights. While equal distribution of nightlights, and thus state capacity, between overlapping territories is a strong assumption, it introduces less bias than ignoring the problem. Doing nothing double counts the nightlights of overlapping cells, resulting in the state devoting 'extra' attention relative to the total investment in a given region. I do not carry out this adjustment for nightlights used in the placebo tests in Section 3.1 of the paper as neither second order administrative units nor PRIO GRID cells overlap one another.

Another shortcoming of these data is that the units of brightness are not inherently meaningful and are not stable over time. In addition to sensor drift within a satellite over time, values are not comparable across satellites. The maximum value in the data is 63, but that does that not mean that 63 in two years of the same satellite is equivalent, or that 63 between two satellites is equivalent. Users of the data have developed an intercalibration method to deal with these issues (Wu, He, Peng, Li & Zhong 2013). Essentially, geographic regions that do not vary over time are identified, one year of data is chosen as a reference raster, and then a model is fit using all other years to explain the invariant region in reference year. The coefficients of this model represent the difference between a given satellite-year and the reference raster. Once this model is trained, it is applied to the rest of the world, adjusting estimates for all other years so that they can be compared to the reference year. Following Wu et al. (2013), I select the Japanese prefecture of Okinawa, the American territory of Puerto Rico, and the nation of Mauritius as invariant regions to calibrate the DMSP OLS data.

C Population Considerations

As the population data (Center for International Earth Science Information Network - CIESIN - Columbia University; United Nations Food and Agriculture Programme - FAO; Centro Internacional de Agricultura Tropical - CIAT 2005, Center for International Earth Science Information Network - CIESIN - Columbia University 2015) are only available in five year intervals, I linearly interpolate the data for the intervening years. While a rather blunt method of imputation, there are two main reasons that this approach is appropriate. First, measuring population on a yearly time scale already involves significantly loss of information. Second, a parametric imputation approach that uses variables observed in all years would either only be able to use country level variables, or would require the collection of significant amount of data at the subnational level, which is prohibitively time consuming. As with nightlights, I correct for population in overlapping ethnic groups polygons for all analyses except the placebo tests in Section 3.1.

D Missing Data

Table D.1 presents the missingness of explanatory and control variables. Due to the fact that no variable has more than 10% of data missing I treat these observations as missing not at random and multiply impute them (Rubin 1987) using the `mice` package (van Buuren & Groothuis-Oudshoorn 2011), generating five imputed datasets. For all models with missing data, I estimate two chains on each imputed dataset and then pool all 10 chains together for inference.

	% Missing
Polyarchy	0.90
Lost Autonomy	2.67
GDP per capita	6.02

Table D.1: Missingness of control variables.

E Estimation and MCMC Diagnostics

I estimate the models using the Stan probabilistic programming language (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li & Riddell 2017) in R (R Core Team 2020) via the brms R package (Bürkner 2017, Bürkner 2018). Due to missingness in the variables, I multiply impute the missing values using the `mice` package (van Buuren & Groothuis-Oudshoorn 2011). I generate 5 imputed datasets, run two chains on each, and then perform inference on all 10 chains pooled together, averaging over the uncertainty in different imputed values (Little & Rubin 2002, 217-218).[1] I run four chains for 2,000 warmup iterations followed by 2,000 sampling iterations. All inference is based on the sampling iterations. Standard diagnostics indicate good convergence of the chains.

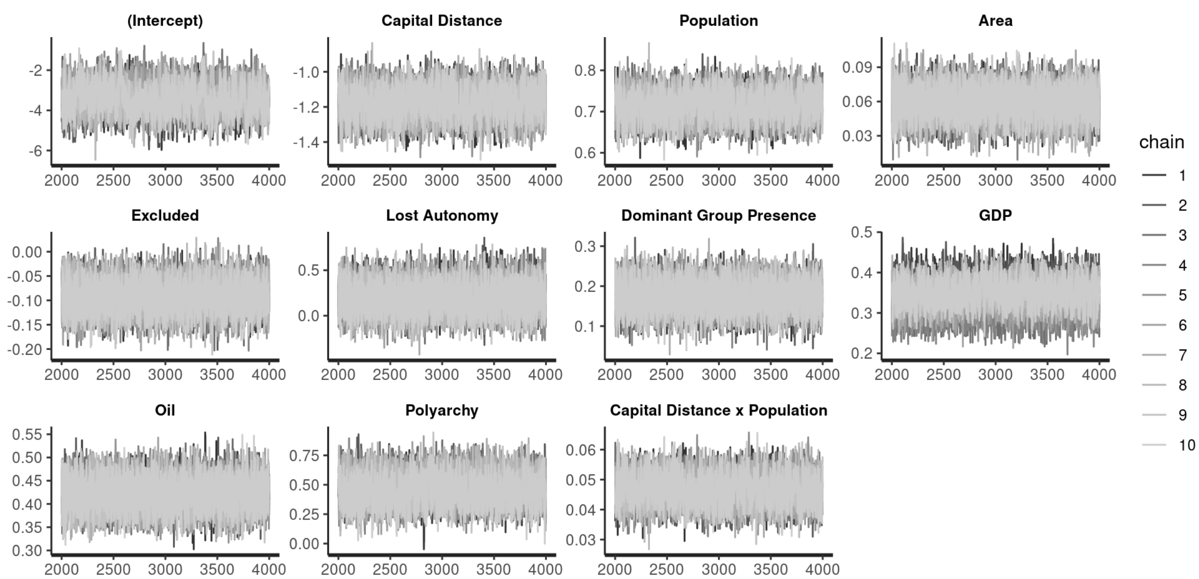


Figure E.1: Traceplot of samples for β in Model 3. Each shade of grey represents samples from one chain initialized at different starting values.

This section presents diagnostics of MCMC samples for Models 3 and 6 with the full slate of group and country level political control variables. Figures E.1 and E.2 display the traceplots for the regression coefficients β in Models 3 and 6, respectively. Each shade of grey represents a different chain, and the overlap between them provides evidence that the chains have converged to the stationary distribution. Figures E.3 and E.4 present plots of the Geweke diagnostic statistics for β in Models 3 and 6, respectively. The diagnostic tests whether the chain has converged to the stationary distribution by comparing the means of the first 10% and final 50% of the samples in each chain. Almost all estimates are

[1]Although it is possible to employ a model that jointly specifies the probability of an observation's absence alongside the parameters of interest, doing so is unnecessary in this case. When the proportion of missing information in a dataset is low, this "uncongeniality" between separate imputation and analysis models does not affect inference of imputed data (Meng 1994). The percentage of missing data in the data is .24%, so this should not affect the validity of my inferences.

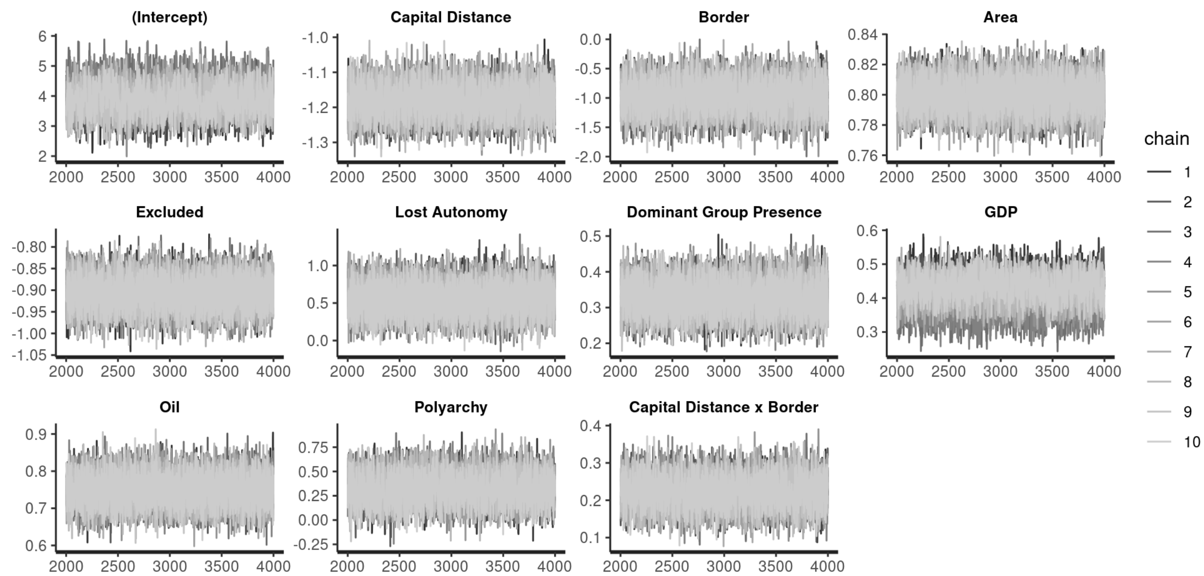


Figure E.2: Traceplot of samples for β in Model 6. Each shade of grey represents samples from one chain initialized at different starting values.

within $\pm$ 1.96 standard deviations of the mean, offering further evidence that the chains have converged to the stationary distribution.

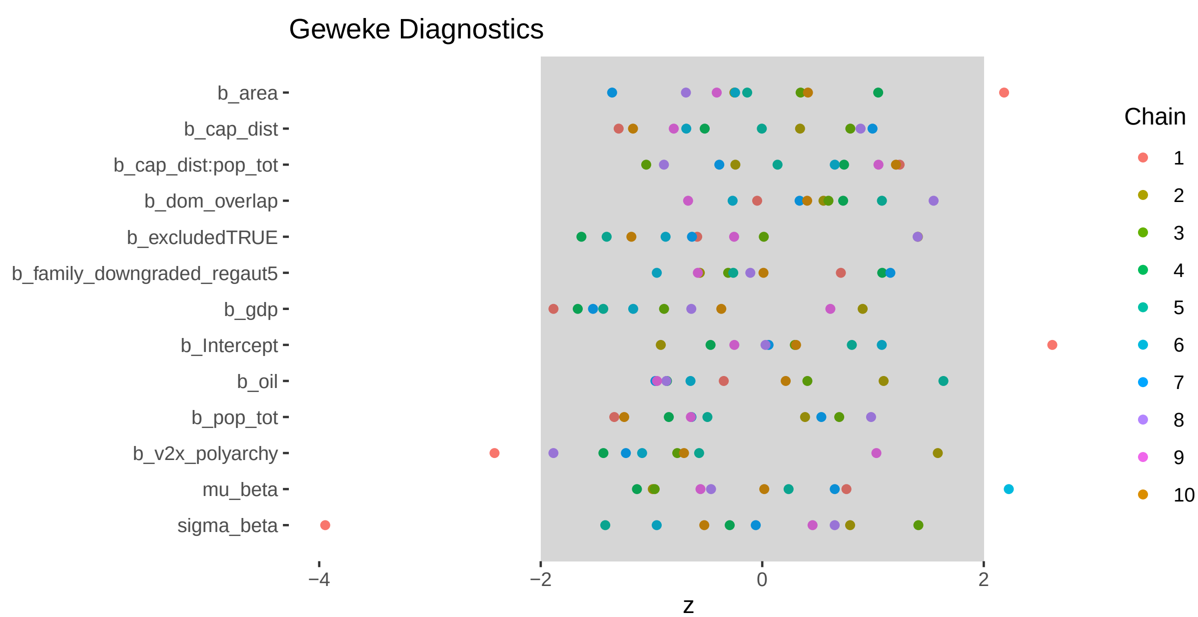


Figure E.3: Geweke diagnostic plot for β in Model 3. Dots are z-scores of the difference in means of the first 10% and final 50% of the samples in each chain.

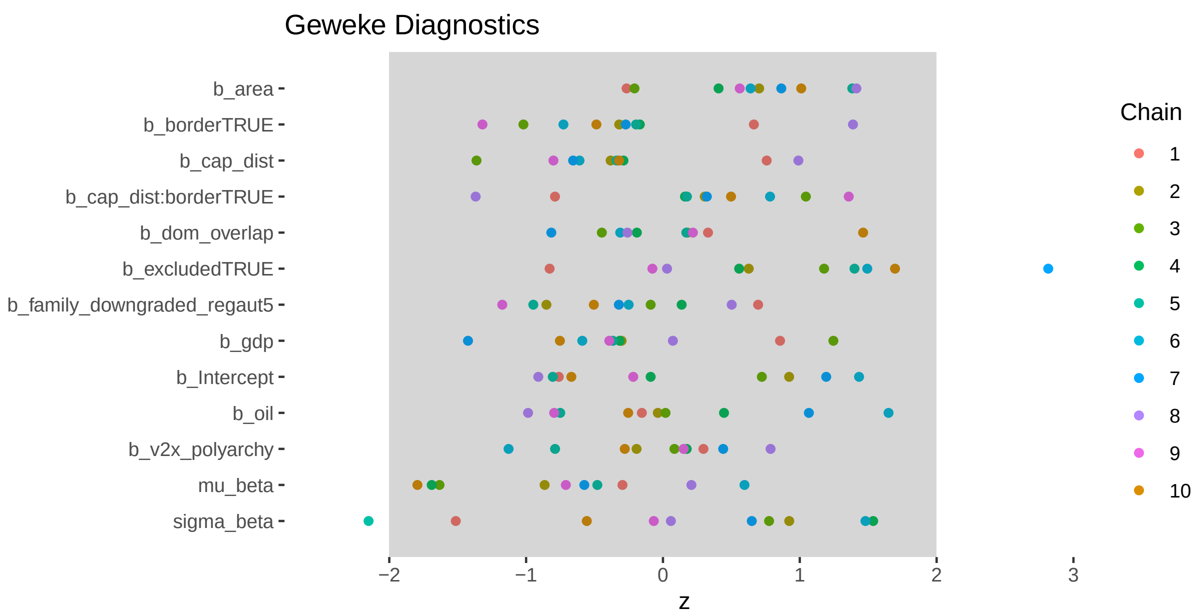


Figure E.4: Geweke diagnostic plot for β in Model 6. Dots are z-scores of the difference in means of the first 10% and final 50% of the samples in each chain.

F Prior Sensitivity

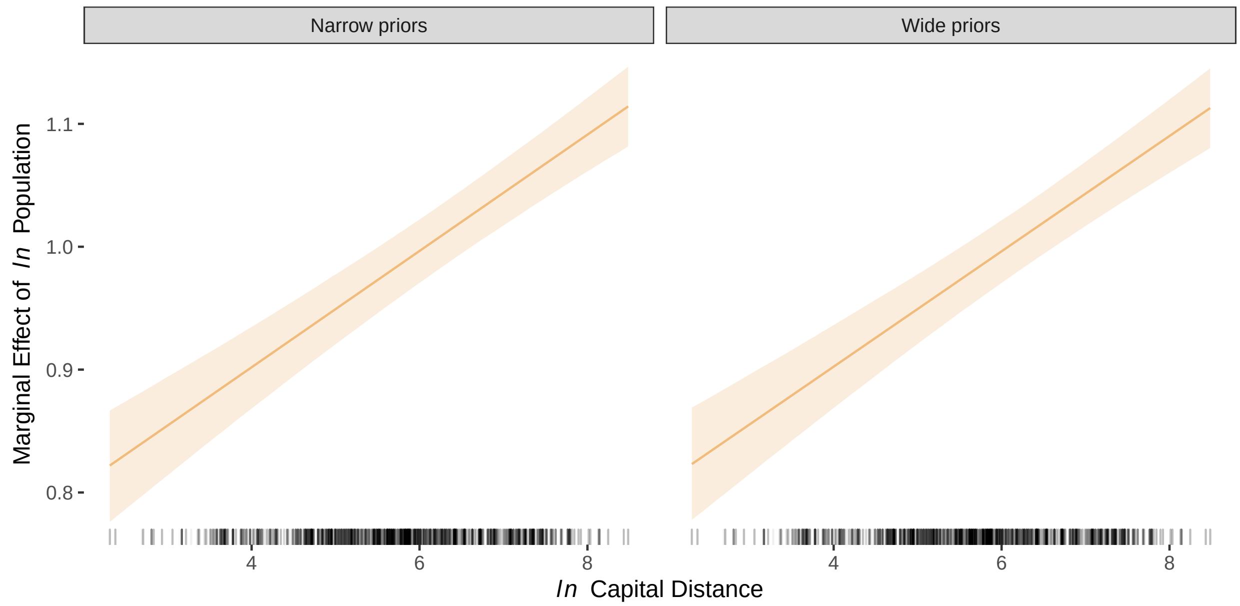


Figure F.1: Marginal effects of politically excluded ethnic group population on nighttime light levels, conditional on distance to the capital.

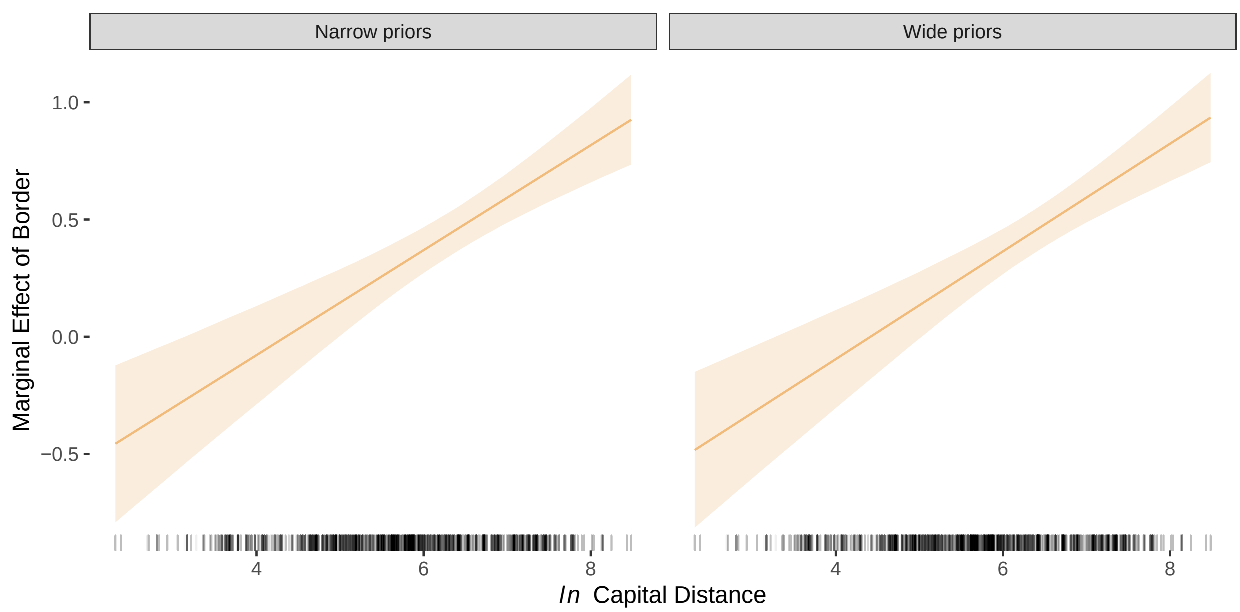


Figure F.2: Marginal effects of politically excluded ethnic group border on nighttime light levels, conditional on distance to the capital.

G Political Status

I also estimate models explaining the level of nightlights in a group's territory using only the subsample of politically excluded groups. I conduct two separate sub-analyses including all excluded groups and only truly marginalized groups. Results are similar in magnitude to those in the main paper across both sets of analyses.

G.1 Political exclusion

This section restricts the sample to politically excluded ethnic groups by eliminating dominant groups. This leaves those coded as senior partner, junior partner, self-exclusion, powerless, discriminated, and state collapse.

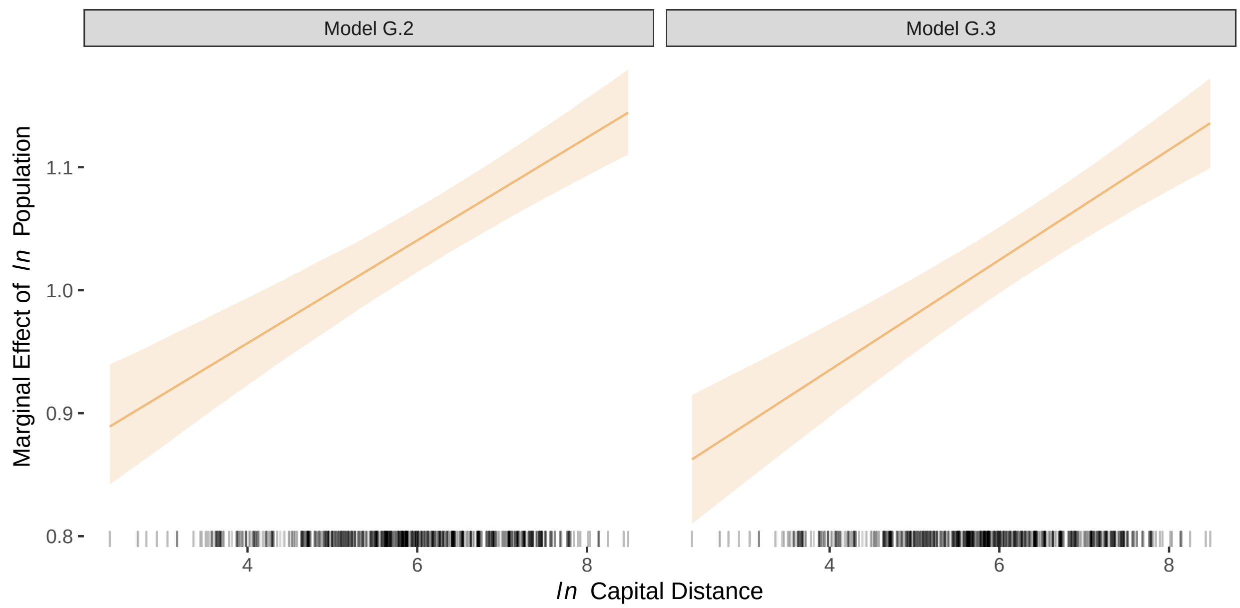


Figure G.1: Marginal effects of politically excluded ethnic group population on nighttime light levels, conditional on distance to the capital.

	Model G.1	Model G.2	Model G.3
ln Population	1.15*	0.79*	0.76*
	[1.13; 1.17]	[0.72; 0.86]	[0.68; 0.83]
ln Capital Distance		−1.10*	−1.13*
		[−1.28; −0.92]	[−1.32; −0.93]
ln Population × *ln* Capital Distance		0.04*	0.04*
		[0.03; 0.05]	[0.03; 0.06]
ln Area		0.08*	0.06*
		[0.05; 0.11]	[0.03; 0.09]
Dominant Group Presence			0.11*
			[0.03; 0.18]
Lost Autonomy			−0.10
			[−0.41; 0.21]
Oil			0.43*
			[0.36; 0.50]
ln GDP_{PC}			0.48*
			[0.39; 0.57]
Polyarchy			0.59*
			[0.31; 0.87]
(Constant)	−9.01*	−1.64*	−5.39*
	[−9.81; −8.21]	[−3.02; −0.39]	[−6.97; −3.76]
σ_α	2.07*	2.09*	1.46*
	[1.82; 2.33]	[1.84; 2.36]	[1.29; 1.68]
σ_γ	1.54*	1.45*	1.33*
	[1.12; 2.12]	[1.07; 2.03]	[0.98; 1.82]
WAIC	34538.40	34149.58	33892.14
5-fold RMSE	1.17	1.15	1.14
Observations	10929	10929	10929

* 0 outside 95% credible interval

Table G.1: Linear models explaining nightlights as a function of excluded ethnic group population and capital distance. The standard deviation of the country and year random intercepts are represented by σ_α and σ_γ, respectively. Continuous variables logged and standarized.

	Model G.4	Model G.5	Model G.6
Border	1.07*	−0.41	−0.54
	[0.94; 1.20]	[−0.96; 0.15]	[−1.09; 0.02]
ln Capital Distance		−1.28*	−1.20*
		[−1.37; −1.18]	[−1.29; −1.11]
Border × *ln* Capital Distance		0.15*	0.17*
		[0.06; 0.23]	[0.08; 0.25]
ln Area		0.86*	0.79*
		[0.84; 0.88]	[0.76; 0.81]
Dominant Group Presence			0.50*
			[0.41; 0.59]
Lost Autonomy			−0.04
			[−0.42; 0.35]
Oil			0.68*
			[0.60; 0.77]
ln GDP_{PC}			0.42*
			[0.31; 0.53]
Polyarchy			0.73*
			[0.41; 1.06]
(Constant)	8.35*	7.19*	3.16*
	[7.87; 8.86]	[6.41; 7.99]	[1.96; 4.33]
σ_α	2.65*	2.47*	1.95*
	[2.34; 3.01]	[2.15; 2.79]	[1.71; 2.22]
σ_γ	0.39*	0.40*	0.31*
	[0.29; 0.55]	[0.29; 0.55]	[0.22; 0.43]
WAIC	45128.99	38930.98	38568.70
5-fold RMSE	1.91	1.44	1.41
Observations	10929	10929	10929

* 0 outside 95% credible interval

Table G.2: Linear models explaining nightlights as a function of excluded ethnic group border and capital distance. The standard deviation of the country and year random intercepts are represented by σ_α and σ_γ, respectively. Continuous variables logged and standarized.

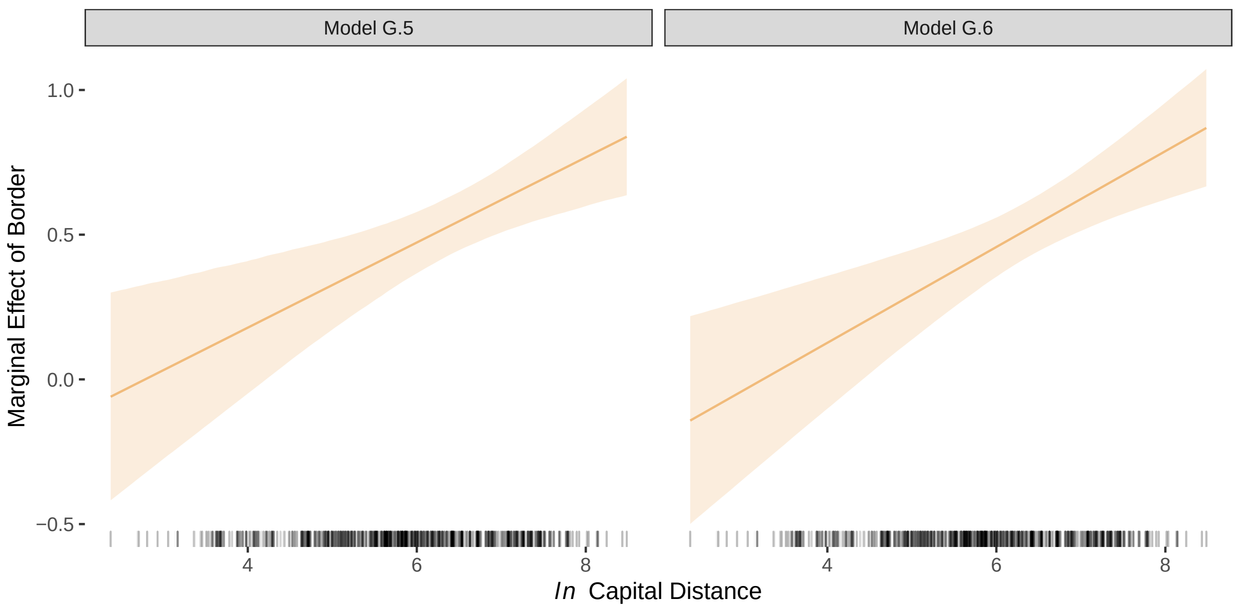


Figure G.2: Marginal effects of politically excluded ethnic group border on nighttime light levels, conditional on distance to the capital.

G.2 Political marginalization

This section restricts the sample to politically excluded ethnic groups by eliminating dominant, senior partner, and junior partner groups. This leaves those coded as self-exclusion, powerless, discriminated, and state collapse.

	Model G.7	Model G.8	Model G.9
ln Population	1.08*	0.71*	0.61*
	[1.06; 1.11]	[0.62; 0.79]	[0.50; 0.73]
ln Capital Distance		−1.12*	−1.28*
		[−1.32; −0.90]	[−1.55; −1.00]
ln Population × *ln* Capital Distance		0.05*	0.06*
		[0.03; 0.06]	[0.04; 0.08]
ln Area		0.06*	0.04*
		[0.02; 0.09]	[0.01; 0.08]
Dominant Group Presence			0.09
			[−0.04; 0.23]
Lost Autonomy			0.25
			[−0.09; 0.59]
Oil			0.45*
			[0.36; 0.54]
ln GDP_{PC}			0.61*
			[0.49; 0.72]
Polyarchy			0.23
			[−0.13; 0.60]
(Constant)	−8.04*	−0.69	−4.18*
	[−8.97; −7.12]	[−2.18; 0.62]	[−6.36; −1.97]
σ_α	2.24*	2.27*	1.62*
	[1.95; 2.57]	[1.99; 2.60]	[1.40; 1.88]
σ_γ	1.51*	1.43*	1.27*
	[1.11; 2.11]	[1.05; 2.00]	[0.94; 1.76]
WAIC	20864.17	20716.12	20512.62
5-fold RMSE	1.21	1.20	1.17
Observations	6494	6494	6494

* 0 outside 95% credible interval

Table G.3: Linear models explaining nightlights as a function of marginalized ethnic group population and capital distance. The standard deviation of the country and year random intercepts are represented by σ_α and σ_γ, respectively. Continuous variables logged and standarized.

	Model G.10	Model G.11	Model G.12
Border	1.06*	−1.00*	−0.88*
	[0.92; 1.20]	[−1.74; −0.25]	[−1.65; −0.12]
ln Capital Distance		−1.28*	−1.19*
		[−1.40; −1.16]	[−1.31; −1.07]
Border × *ln* Capital Distance		0.23*	0.22*
		[0.12; 0.34]	[0.11; 0.33]
ln Area		0.72*	0.65*
		[0.69; 0.75]	[0.62; 0.68]
Dominant Group Presence			0.56*
			[0.40; 0.72]
Lost Autonomy			0.46*
			[0.07; 0.86]
Oil			0.73*
			[0.62; 0.83]
ln GDP_{PC}			0.67*
			[0.55; 0.80]
Polyarchy			−0.13
			[−0.56; 0.30]
(Constant)	7.73*	8.53*	2.65*
	[7.17; 8.28]	[7.55; 9.50]	[1.23; 4.05]
σ_α	2.79*	2.69*	2.12*
	[2.44; 3.21]	[2.36; 3.09]	[1.86; 2.43]
σ_γ	0.34*	0.38*	0.23*
	[0.24; 0.47]	[0.27; 0.52]	[0.15; 0.33]
WAIC	25670.87	23127.11	22846.66
5-fold RMSE	1.75	1.44	1.41
Observations	6494	6494	6494

* 0 outside 95% credible interval

Table G.4: Linear models explaining nightlights as a function of marginalized ethnic group border and capital distance. The standard deviation of the country and year random intercepts are represented by σ_α and σ_γ, respectively. Continuous variables logged and standarized.

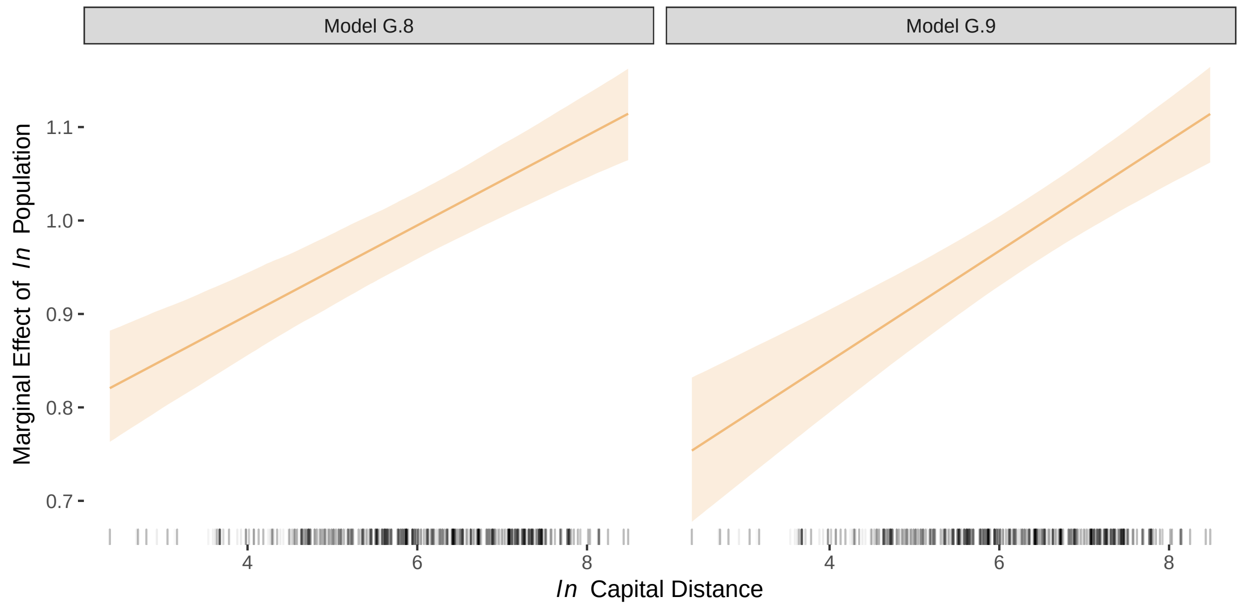


Figure G.3: Marginal effects of politically excluded ethnic group population on nighttime light levels, conditional on distance to the capital.

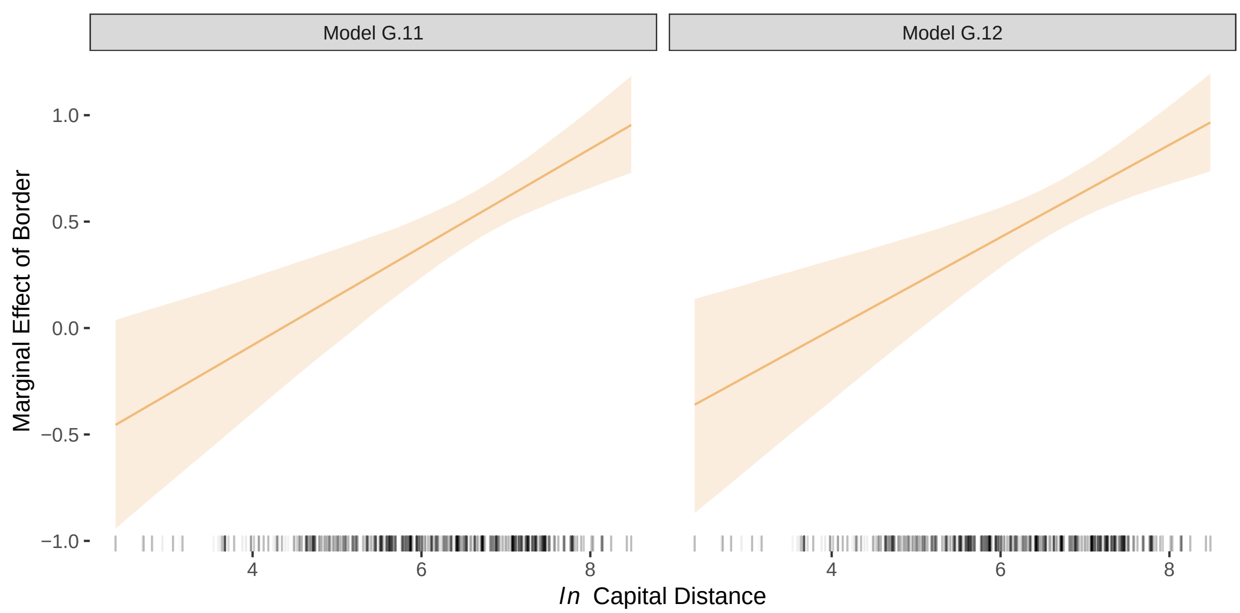


Figure G.4: Marginal effects of politically excluded ethnic group border on nighttime light levels, conditional on distance to the capital.

H Alternative Measures

The population Gini measure is calculated by treating each grid cell in the population data as an individual in the standard Gini index formula in Equation H.1:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}} \tag{H.1}$$

This excellently captures the theoretical concept of population concentration. While Weidmann (2009) uses the Herfindahl-Hirschman index to measure population concentration, his unit of analysis is ethnic group territory polygons, not grid cells within a polygon. Thus, his data will have no instances of a unit with zero population. As the Herfindahl-Hirschman index is a diversity measure, it ignores observations with a zero value. This property is inappropriate when many observations have zero population and these unpopulated grid cells indicate a more concentrated population. Each grid cell with no population contributes to a higher Gini coefficient because between two territories with equal population, the one with more unoccupied areas will have a more concentrated population overall.

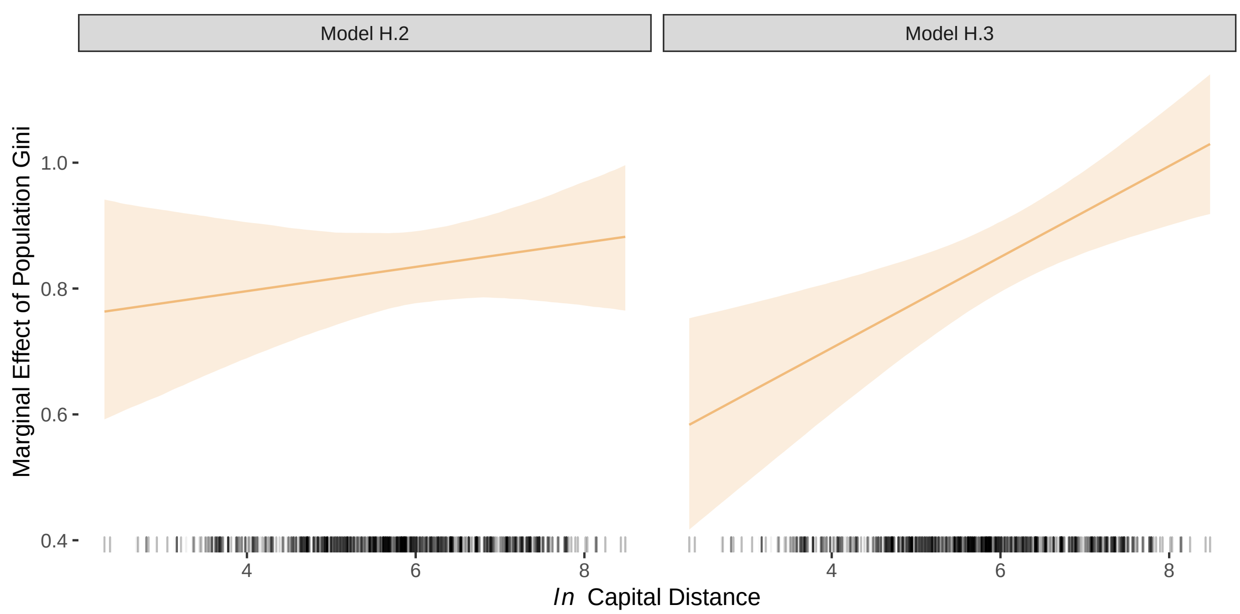


Figure H.1: Marginal effects of ethnic group population concentration on nighttime light levels, conditional on distance to the capital.

Figure H.1 presents results for the reestimated Models 2 & 3. The relationship between population Gini and nightlights is similar to that of total population. Effect sizes are smaller, and model fit is worse when comparing WAIC and RMSE. However, the relationship remains positive.

	Model H.1	Model H.2	Model H.3
Population Gini	2.22*	0.72*	0.42*
	[2.15; 2.29]	[0.45; 1.00]	[0.16; 0.68]
ln Capital Distance		−1.05*	−0.81*
		[−1.10; −1.00]	[−0.86; −0.76]
Population Gini × *ln* Capital Distance		0.02	0.07*
		[−0.02; 0.06]	[0.03; 0.11]
ln Area		0.88*	0.69*
		[0.86; 0.89]	[0.66; 0.71]
Excluded			−0.93*
			[−0.99; −0.86]
Dominant Group Presence			0.28*
			[0.20; 0.37]
Lost Autonomy			0.66*
			[0.29; 1.03]
Oil			0.76*
			[0.69; 0.84]
ln GDP_{PC}			0.40*
			[0.33; 0.47]
Polyarchy			0.29
			[−0.01; 0.59]
(Constant)	11.72*	7.06*	4.20*
	[11.31; 12.12]	[6.53; 7.59]	[3.47; 4.94]
σ_α	2.25*	2.25*	1.76*
	[1.99; 2.59]	[2.00; 2.55]	[1.55; 1.99]
σ_γ	0.29*	0.34*	0.24*
	[0.21; 0.40]	[0.24; 0.47]	[0.17; 0.34]
WAIC	52512.20	44949.95	43716.56
5-fold RMSE	1.91	1.42	1.35
Observations	12714	12714	12714

* 0 outside 95% credible interval

Table H.1: Linear models explaining nightlights as a function of ethnic group population Gini and capital distance. The standard deviation of the country and year random intercepts are represented by σ_α and σ_γ, respectively. Continuous variables logged and standarized.

I Alternate Secession Risk Measure

This section presents results from an alternate specification where nightlights per capita are treated as the outcome variable and separatism risk is measured via distance to the capital.

	Model I.1	Model I.2
ln Capital Distance	-0.05^{*}	-0.04^{*}
	$[-0.05;\ -0.05]$	$[-0.04;\ -0.04]$
ln Area		0.02^{*}
		$[0.02;\ 0.02]$
Excluded		-0.02^{*}
		$[-0.02;\ -0.01]$
Dominant Group Presence		0.02^{*}
		$[0.02;\ 0.03]$
Lost Autonomy		0.03^{*}
		$[0.00;\ 0.05]$
Oil		0.03^{*}
		$[0.03;\ 0.04]$
ln GDP_{PC}		0.03^{*}
		$[0.02;\ 0.04]$
Polyarchy		0.01
		$[-0.01;\ 0.03]$
(Constant)	0.87^{*}	0.28^{*}
	$[0.83;\ 0.92]$	$[0.22;\ 0.35]$
σ_{α}	0.15^{*}	0.10^{*}
	$[0.13;\ 0.16]$	$[0.08;\ 0.11]$
σ_{γ}	0.07^{*}	0.06^{*}
	$[0.05;\ 0.10]$	$[0.05;\ 0.09]$
WAIC	-23000.60	-25945.61
5-fold RMSE	0.10	0.09
Observations	12714	12714

* 0 outside 95% credible interval

Table I.1: Linear models explaining nightlights per capita as a function of capital distance. The standard deviation of the country and year random intercepts are represented by σ_{α} and σ_{γ}, respectively. Continuous variables logged and standarized.

J Fixed Effects Specifications

As a further robustness check I estimate versions of Models 2 and 5 using fixed instead of random effects. By eliminating the control variables in Models 3 and 6, I am able to remove time-invariant covariates from the models, allowing a fixed effects estimation strategy.

	Model J.1	Model J.2
ln Population	0.87^{*}	
	(0.03)	
Border		-0.21
		(0.37)
ln Capital Distance	-1.01^{*}	-1.69^{*}
	(0.08)	(0.06)
ln Population $\times$ *ln* Capital Distance	0.04^{*}	
	(0.00)	
Border $\times$ *ln* Capital Distance		0.28^{*}
		(0.06)
Country Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Adjsted R^2	0.88	0.63
RMSE	1.15	1.99
Observations	12714	12714

$^{*}p < 0.05$

Table J.1: Linear models explaining nightlights as a function of ethnic group population, borders, and capital distance with all predictors lagged one year.

The results of these models in Table J.1 are largely in line with those in Tables 1 and 2, indicating a positive marginal effect increasing in distance from the capital. These models account for all cross-sectional sources of variation between states, as well as yearly variations in investment patterns by central governments.

K Separatist Conflict Onset

This section present results form an analysis of separatist conflict onset in light of increasing subnational investment by governments as a preemption strategy. As in the main analyses, nightlights and capital distance and logged and lagged one year to deal with endogeneity concerns. A cubic polynomial of time since the last conflict spell is included (Carter & Signorino 2010), and active ongoing conflict years are dropped (McGrath 2015). Separatist conflict onset is taken from FORGE (Braithwaite & Cunningham 2020).

	Model K.1
ln Nightlights	-0.24^{*}
	(0.06)
ln Capital Distance	-0.13
	(0.29)
Country Fixed Effects	✓
Polynomial Time	✓
AIC	773.86
BIC	1808.31
Log Likelihood	-246.93
Deviance	493.86
Num. obs.	11955

$^{*}p < 0.05$

Table K.1: Logit analysis of separatist conflict onset

Nightlights are negatively and statistically significantly related to conflict onset in this analysis, while capital distance is not statistically significantly related.

References

Braithwaite, Jessica Maves & Kathleen Gallagher Cunningham. 2020. "When Organizations Rebel: Introducing the Foundations of Rebel Group Emergence (FORGE) Dataset." *International Studies Quarterly* 64(1):183–193.

Bürkner, Paul-Christian. 2017. "Brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80(1).

Bürkner, Paul-Christian. 2018. "Advanced Bayesian Multilevel Modeling with the R Package Brms." *The R Journal* 10(1):395.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. "Stan A Probabilistic Programming Language." *Journal of Statistical Software* 76(1).

Carter, David B. & Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271–292.

Cederman, Lars-Erik, Halvard Buhaug & Jan Ketil Rød. 2009. "Ethno-Nationalist Dyads and Civil War A GIS-Based Analysis." *Journal of Conflict Resolution* 53(4):496–525.

Cederman, Lars-Erik, Nils B. Weidmann & Kristian Skrede Gleditsch. 2011. "Horizontal Inequalities and Ethnonationalist Civil War: A Global Comparison." *The American Political Science Review* 105(3):478–495.

Cederman, Lars-Erik, Nils B. Weidmann & Nils-Christian Bormann. 2015. "Triangulating Horizontal Inequality: Toward Improved Conflict Analysis." *Journal of Peace Research* 52(6):806–821.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2015. Gridded Population of the World, Version 4 (GPWv4): Population Density. Technical report NASA Socioeconomic Data and Applications Center (SEDAC) Palisades, NY: .

Center for International Earth Science Information Network - CIESIN - Columbia University; United Nations Food and Agriculture Programme - FAO; Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Technical report NASA Socioeconomic Data and Applications Center (SEDAC) Palisades, NY: .

Hsu, Feng-Chi, Kimberly E. Baugh, Tilottama Ghosh, Mikhail Zhizhin & Christopher D. Elvidge. 2015. "DMSP-OLS Radiance Calibrated Nighttime Lights Time Series with Intercalibration." *Remote Sensing* 7(2):1855–1876.

Little, Roderick & Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd ed. ed. Hoboken, N.J.: Wiley.

McGrath, Liam F. 2015. "Estimating Onsets of Binary Events in Panel Data." *Political Analysis* 23(4):534–549.

Meng, Xiao-Li. 1994. "Multiple-Imputation Inferences with Uncongenial Sources of Input." *Statistical Science* 9(4):538–558.

R Core Team. 2020. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing.

Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley.

van Buuren, Stef & Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45(3).

Weidmann, Nils B. 2009. "Geography as Motivation and Opportunity Group Concentration and Ethnic Conflict." *Journal of Conflict Resolution* 53(4):526–543.

Wu, Jiansheng, Shengbin He, Jian Peng, Weifeng Li & Xiaohong Zhong. 2013. "Intercalibration of DMSP-OLS Night-Time Light Data by the Invariant Region Method." *International Journal of Remote Sensing* 34(20):7356–7368.