Supplemental Appendix For

The Urban Origins of Rebellion

Contents

This appendix proceeds in two parts. We begin with reporting summary statistics on all our relevant variables as well as a figure plotting ratio of civil war onset by location type. We then proceed to discuss a large number of sensitivity models used to evaluate the robustness of our results to important potential confounders, modeling choices, and endogeneity.

Summary Statistics

Table A1: Summary Statistics of All Variables

	Minimum	Median	Mean	Max	SD
Dependent variables					
Capital	0	0	0.174	1	0.379
Large City	0	0	0.355	1	0.479
Rural	0	0	0.445	1	0.498
Large City 10K	0	0	0.4	1	0.491
Large City 100K	0	0	0.337	1	0.473
Large City 1M	0	0	0.166	1	0.372
Independent variables					
Military Faction	0	0	0.141	1	0.349
Social Interest	0	0	0.061	1	0.240
Political Party	0	0	0.195	1	0.397
Ethnic Group	0	0	0.487	1	0.500
Relative Strength	0	0	0.844	1	0.841
Population (log)	5.721	9.598	9.661	13.990	1.507
GDP PC (log)	5.402	7.199	7.402	10.227	0.949
% Mountain	0	12.900	23.406	82.200	22.816
Ethnic Frac.	0.005	0.652	0.561	0.902	0.271
Religious Frac.	0	0.366	0.399	0.783	0.211
Youth Movement	0	0	0.471	1	0.212
Labor Union	0	0	0.017	1	0.127
Separatist	0	0	0.377	1	0.485
Representation	0	0	0447	1	0.207
Ideology	0	1	0.730	1	0.445
Military Faction (No Coups)	0	0	0.04	1	0.196
Instrumental variable					
Political Ban	0	0	0.94	1	0.292

Robustness Models and Sensitivity Analyses

Tables A2–A9 report the estimates from a large number of sensitivity analyses designed to illustrate our model's robustness to a large number of modeling and specification concerns to the extent possible with the data available. First, we illustrate our findings are not driven by the inclusion of controls for ethnic-based rebel groups and relative rebel group capacities by estimating models that include only our main independent variables of interest (*Military Faction*, *Social Interest*, and *Political Party*) in Table A2. Table A3 then ensures our results are robust to alternative operationalization thresholds of the *Large City* variable is not due to the threshold of 50,000 residents or above we used to define a city by dichotomizing this variable using a threshold of 10,000, 100,000, and 1 million residents, respectively.

In Table A5 we then illustrate that our results regarding *Social Interest* are robust to our decision to aggregate groups formed by labor unions and youth movements into one category by disaggregating this variable into these two constitutive categories. Here, our findings hold to the $p < .1$ level—at least with regards to Hypothesis H2—suggesting our decision to combine these two categories into one broad variable is defensible. Table A6 then illustrates that our findings are specific to the linkage between group type and formation location rather than by the specific goals of the group by accounting for secessionist (i.e., whether a group's initial goal was the pursuit of an independent state or autonomy within their region), increasing political participation, and idealogical goals.

Another potential concern is endogeneity, namely whether our DVs can substantially impact our explanatory variables rather than (only) the other way around. First, we do not believe that endogeneity is a real concern here, because the location of formation is unlikely to influence the origin type of the group forming. In other words, it is far less likely that being formed at a given city at year t impacted the composition of the group's densest constituent population at $t - 1$ rather than the other way around (Granger 1988). Nevertheless, in Table A7 we illustrate our results' robustness to these concerns, at least with respect to groups that originated in political parties, by using a two-step probit model

with an effective instrumental variable for political origin (see Rivers and Vuong 1988).

Accounting for endogeneity within time-invariant cross-sectional data is usually done by using an instrumental variable (IV) (Angrist and Pischke 2008, 182-184). Such instrumental variables must satisfy the exclusion restriction, namely to only affect the DV through the instrumented variable rather than via other channels. This makes identifying effective instruments for all our explanatory variables of interest challenging. However, we are able to identify an effective IV for one of our explanatory variables: Political Party. Here, we look at whether parties were practically fully banned before the group is recorded as formed, obtained as *de facto* illegalization from the "Democracy Dictatorship Revisited" dataset created by Cheibub et al. (2010).[1]

How is this variable, *Political Ban*, an effective IV? First, by definition, it cannot be affected by the occurrence of conflict or rebel group formation, because it must occur before either one of these outcomes can realize. Second, this means that if a political ban on parties impacts the probability of rebel group formation in any location type, the only way it can do so is by pushing illegalized political parties to challenge this ban by organizing as a rebel group. Due to these two reasons, Political ban does not violate the exclusion restriction, and can serve as an effective instrument, although—theoretically—only for rebel groups that origin in political parties. While accounting for only one of our four explanatory variables is not perfect, by doing so we hope to lend at least some empirical evidence to our above claim that endogeneity unlikely drives our result, as much as is feasible with our data.

Considering that our dependent variables are all binary (i.e., can take only values of 0 or 1), we employ a two-step probit method (Rivers and Vuong 1988). This approach illustrates the probability of the dependent variable being one given the values of the regressors, in the absence of endogeneity, which means that it makes it possible to trace the effects of changes in the (potentially endogenous) *Political Party* variable on the probability of formation in any of the four location types. Similarly to other instrumental variable based approaches (e.g.,

[1]Note that results remain unchanged when we use a measure of *de jure* illegalization instead.

two-stage least squares), these models rely on two equations to identify the hypothesized relationship(s). The first equation estimates the effects of the instrument *Political Ban* on our endogenous independent variable of interest—*Political Party*—adding all the same independent variables used in the second stage as controls. This equation is accordingly specified as follows:

$$Political\ Party_i = \gamma_i Political\ Ban_i + \gamma_X X_i + \epsilon_1 \tag{1}$$

Where *Political Party* is the endogenous regressors, *Political Ban* is the instrument and γ_i their coefficient, X_i are the rest of the independent variables with γ_X being their coefficients, and ϵ_1 the error term. Accordingly, the second equation is specified as:

$$\Phi\left(Y_i = \widehat{Political\ Party}_i + \gamma_X X_i + \epsilon_2\right) \tag{2}$$

Where Y_i corresponds to each dependent variable, $\widehat{Political\ Party}_i$ is the estimates from the first stage equation, and ϵ_2 the error term of the equation. Because each dependent variable is binary, $\Phi()$ is the probit function. To ensure that all values fall within zero and one, the estimates for $\widehat{Political\ Party}_i$ are standardized (i.e., divided by the standard error). The two-step probit models in Table A7 hence reasonably isolate the direct effect flowing from political party based groups to location formation rather than the other way around, and show that the findings are unlikely the result of simultaneous relationships between the two,[2] thus providing additional confirmation to the linkages between group type and formation location developed in the main article.

Next, recall that while we chose to employ a simple set of binary models to account for each dependent variable separately, models exist that account for such effects jointly. One such option is a multinomial model, where the probability of choosing one category is evaluated with respect to a reference option. For our purposes, such a model was not

[2]Although the instrument *Political Ban*$_i$'s effect is not statistically significant, it has the expected (positive) sign, suggesting it is viable (Angrist and Pischke 2008, 182-184).

applicable because a dependent variable that includes all three outcomes is likely violate the independence of irrelevant alternatives due to some overlap between our categories, e.g., large cities and capitals. Nevertheless, we illustrate that our findings are robust to such joint DV specifications in two ways.

First, we create a multinomial dependent variable that includes our three outcomes with some modifications to help alleviate the potential for IIA violations. Note that in the process, we omit all observations where formation was recorded as "foreign," which leaves us with a total of 304 groups/observations. In this new DV, the first category corresponds to our Capital DV; the second is similar to Large City, only here any large city that was also the country's capital was not and the remaining category hence corresponds to Rural=1. To this end, Table A8 reports the estimates of a multinomial logit model where the categories are Capital and Large City (as discussed here). Again, the reference category is Rural, meaning that each independent variable coefficient in each category lists the effect of said IV on each DV category compared with its impact on Rural. As Table A8, our results hold in sign and significance, even though the multinomial DV is still suspect to violations of the independence of irrelevant alternatives, which can bias coefficient estimates toward zero in some cases.

Next, to further ensure that using a joint DV does not impact our results, we estimate a multivariate probit model in Table A9 (Ashford and Sowden 1970). Compared with a multinomial logit, the multivariate probit allows one to model the impact of each independent variable on our different binary dependent variables, while still accounting for the mutual dependence between the latter, which means the results are hence not impacted by the independence of irrelevant alternatives (and prevent the need to omit observations of the foreign-found category). Briefly, the model makes it possible to assess the relationship between the different responses/DVs to each independent variable in terms of the correlation structure of said categories, described in the ρ covariances reported in Table A10. Indeed, as Table A10 shows, the estimated correlations between the DVs are statistically significant, suggesting potential impact on our results. However, as the estimates reported in Table A9

are almost identical to those reported in the main paper, any such correlations do not have a substantive impact our findings, which can be treated—for all intents and purposes—as independent and as pertaining to our explanatory variables of interest, specifically. The total set of sensitivity analyses reported in this section hence strongly suggest our results are robust to a large number of concerns, and hence that our hypotheses—especially Hypothesis H2—cannot be immediately rejected.

Finally Table A11 includes an indicator of military group formation where we removed all groups that initiated coups based on the information in the CAM dataset by Albrecht et al. (2021). Importantly, as we discuss the rational behind this decision in the main paper, the results are not robust when coups are removed – unsurprisingly considering that 43 our of the 60 cases of military group formation in our sample (or 72%) involve a military coups. This suggests that coups-focused groups operate based on incentives similar to other types of groups who supply politics in response for demand, in urban or other areas.

Table A2: Determinants of rebel group formation location – baseline models

	Capital	**Large City**	**Rural**
Military Faction	2.534***	1.897***	−1.407***
	(0.384)	(0.352)	(0.393)
Social Interest	1.359**	0.946**	−0.676
	(0.531)	(0.441)	(0.455)
Political Party	1.448***	0.994***	−0.445
	(0.359)	(0.280)	(0.278)
Constant	−1.255***	−0.323	−0.883***
	(0.422)	(0.340)	(0.341)
Akaike Inf. Crit.	272.115	442.667	496.620
N		380	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A3: Determinants of rebel group formation location – different city size thresholds

	10K	**100K**	**1M**
Military Faction	1.839***	1.708***	1.005***
	(0.397)	(0.368)	(0.388)
Social Interest	0.625	0.708	0.698
	(0.449)	(0.464)	(0.515)
Political Party	0.650**	1.064***	0.535
	(0.281)	(0.286)	(0.336)
Ethnic Group	−0.554**	−0.616**	−0.659**
	(0.240)	(0.254)	(0.322)
Relative Strength	0.153	0.035	0.037
	(0.150)	(0.154)	(0.177)
Constant	0.036	−0.240	−1.583***
	(0.365)	(0.372)	(0.441)
Akaike Inf. Crit.	455.277	424.543	326.476
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A4: Determinants of rebel group formation location – country fixed effects

	Capital	**Large City**	**Rural**
Military Faction	3.221***	2.529***	−1.413**
	(0.716)	(0.616)	(0.585)
Social Interest	1.769**	1.243**	−1.054
	(0.824)	(0.630)	(0.651)
Political Party	2.180***	0.601	−0.397
	(0.585)	(0.409)	(0.399)
Ethnic Group	−1.431**	−0.871**	0.521
	(0.660)	(0.392)	(0.350)
Relative Strength	0.222	0.222	0.174
	(0.353)	(0.274)	(0.248)
Constant	−27.816	18.344	−18.740
	(591,194.800)	(6,522.639)	(6,522.638)
Akaike Inf. Crit.	349.358	484.202	524.757
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A5: Determinants of rebel group formation location – disaggregated social interest groups

	Capital	**Large City**	**Rural**
Military Faction	2.220***	1.727***	−1.198***
	(0.404)	(0.370)	(0.409)
Youth Movement	1.283**	0.610	−0.532
	(0.611)	(0.504)	(0.502)
Labor Union	1.894*	1.749	−0.900
	(1.031)	(1.159)	(1.149)
Political Party	1.330***	0.916***	−0.352
	(0.365)	(0.285)	(0.283)
Ethnic Group	−0.696*	−0.516**	0.608***
	(0.365)	(0.249)	(0.228)
Relative Strength	0.216	0.025	−0.051
	(0.191)	(0.152)	(0.147)
Constant	−1.260***	−0.223	−1.000***
	(0.443)	(0.370)	(0.380)
Akaike Inf. Crit.	269.767	440.437	493.550
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A6: Determinants of rebel group formation location – separatism, representation, and ideology

	Capital	**Large City**	**Rural**
Military Faction	2.106***	1.813***	−1.097***
	(0.419)	(0.382)	(0.416)
Social Interest	1.793***	0.766*	−0.686
	(0.599)	(0.459)	(0.474)
Political Party	1.521***	0.914***	−0.393
	(0.395)	(0.293)	(0.291)
Ethnic Group	−0.164	−0.618**	0.394
	(0.388)	(0.271)	(0.245)
Relative Strength	0.157	0.041	0.016
	(0.193)	(0.154)	(0.150)
Separatist	−1.883***	0.230	0.622**
	(0.610)	(0.298)	(0.265)
Representation	−0.224	0.300	0.125
	(0.716)	(0.554)	(0.529)
Ideology	−0.312	0.184	0.185
	(0.386)	(0.298)	(0.274)
Constant	−1.138**	−0.362	−1.132***
	(0.508)	(0.403)	(0.403)
Akaike Inf. Crit.	260.796	444.547	491.124
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A7: Determinants of rebel group formation location – instrumental variable probit

	Probit stage			*First stage*
	Capital	**Large City**	**Rural**	
$\widehat{Political\ Party}$	0.311***	0.229***	−0.093	–
	(0.079)	(0.068)	(0.068)	
Military Faction	1.171***	0.983***	−0.687***	−0.138**
	(0.227)	(0.220)	(0.235)	(0.065)
Social Interest	0.843***	0.539*	−0.366	0.065
	(0.311)	(0.277)	(0.282)	(0.085)
Ethnic Group	−0.491**	−0.374**	0.395***	−0.091**
	(0.194)	(0.148)	(0.140)	(0.044)
Relative Strength	0.148	0.028	−0.039	0.027
	(0.107)	(0.091)	(0.089)	(0.027)
Political Ban	–	–	–	0.016
				(0.070)
Constant	−0.438*	0.073	−0.700***	0.376***
	(0.236)	(0.215)	(0.221)	(0.066)
Akaike Inf. Crit.	267.667	439.572	491.799	–
R^2	–	–	–	– 0.049
Adjusted R^2	–	–	–	0.033
N		378		

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A8: Determinants of rebel group formation location – multinomial logit

	Capital (only)	**Large City (no capital)**
Military Faction	2.553***	0.989*
	(0.498)	(0.510)
Social Interest	1.483**	0.285
	(0.620)	(0.640)
Political Party	1.481***	0.383
	(0.403)	(0.369)
Ethnic Group	−0.931**	−0.513*
	(0.385)	(0.305)
Relative Strength	0.204	0.002
	(0.211)	(0.193)
Constant	−0.507	−0.419
	(0.496)	(0.476)
Akaike Inf. Crit.	892.359	
N	304	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *Rural* is the reference category. *p<0.1; **p<0.05; ***p<0.01.

Table A9: Determinants of rebel group formation location – multivariate probit

	Capital	**Large City**	**Rural**
Military Faction	1.292***	1.061***	−0.743**
	(0.261)	(0.249)	(0.229)
Social Interest	0.847***	0.640*	−0.434
	(0.283)	(0.328)	(0.326)
Political Party	0.783***	0.603***	−0.316
	(0.204)	(0.209)	(0.210)
Ethnic Group	−0.529	−0.412**	0.455**
	(0.217)	(0.171)	(0.197)
Relative Strength	0.119	0.006	0.010
	(0.105)	(0.107)	(0.101)
Constant	−1.372***	−0.509	−0.207
	(0.198)	(0.188)	(0.201)
LR χ^2		230.067***	
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A10: Correlation matrix of MVP dependent variable categories

	ρ_1	ρ_2	ρ_3
ρ_1	–		
ρ_2	0.667***	–	
	(0.077)		
ρ_3	−0.509***	−0.904***	–
	(0.101)	(0.027)	

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

Table A11: Determinants of rebel group formation location – no military coups

	Capital	**Large City**	**Rural**
Military Faction (No Coups)	−0.131	0.916	−0.256
	(0.720)	(0.566)	(0.581)
Social Interest	1.012*	0.640	−0.471
	(0.524)	(0.453)	(0.465)
Political Party	0.936***	0.760***	−0.278
	(0.327)	(0.275)	(0.279)
Ethnic Group	−0.971***	−0.671***	0.704***
	(0.340)	(0.239)	(0.224)
Relative Strength	0.450***	0.176	−0.151
	(0.170)	(0.142)	(0.143)
Constant	−0.685*	0.080	−1.176***
	(0.401)	(0.353)	(0.373)
Akaike Inf. Crit.	300.528	461.141	501.647
N		378	

Note: Coefficients are reported with standard errors clustered by country in parentheses. *p<0.1; **p<0.05; ***p<0.01.

References

Albrecht, Holger, Kevin Koehler, and Austin Schutz. 2021. "Coup agency and prospects for democracy." International Studies Quarterly 65(4): 1052-1063.

Angrist, Joshua D., and J$\acute{}$orn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Ashford, J. R. and R. R. Sowden. 1970. "Multi-variate probit analysis." *Biometrics* 26(3): 535-546.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, insurgency, and civil war." *American political science review* 97(1): 75-90.

Granger, Clive WJ. 1988. "Causality, cointegration, and control." *Journal of Economic Dynamics and Control* 12(2-3): 551-559.

Rivers, Douglas, and Quang H. Vuong. 1988. "Limited information estimators and exogeneity tests for simultaneous probit models." *Journal of Econometrics* 39(3): 347-366.