

Administrative Unit Proliferation

Guy Grossman*

Department of Political Science
University of Pennsylvania

Janet I. Lewis†

Department of Government
Harvard University

September 10, 2013

1 Online Appendix

This appendix reports several key aspects of our data collection, coding, summary statistics and spatial analysis. First, it summarizes how we collected information about district (LC5) elections and the composition of local-level resource allocation committees– District Executive Committees (DECs)– in order to measure key variables in our empirical analyses. Second, it describes the coding of the dependent variable. Third, it provides summary statistics for all variables used in our analyses. Fourth, it provides further information on the copula-based binary logit choice model we use for accommodating spatial correlation across observational units. Specifically it provides a formalization of the homoskedastic version of the Bhat and Sener likelihood function. Fifth, it provides additional information on the robustness checks.

1.1 Local Politics Data Collection Procedure

Most of the data used for this project were assembled from public sources, especially from the Ugandan Bureau of Statistics (census) and the Ugandan Electoral Commission (presidential elections returns). We collected original data, however, to generate variables capturing political dynamics at the local level. These include, inter alia, information on the share of DEC seats held by individuals from each county during each

*Corresponding author. 208 S. 37th Street (225 Stiteler Hall), Philadelphia, PA 19104, USA. ggros@sas.upenn.edu; 215-898-4209

†589 McNair Road Annapolis, MD 21402, USA. janetilewis@gmail.com; 650-906-9443

inter-election period and data on district chairperson elections.¹

We hired a team of Ugandan research assistants to collect these data. The research assistants contacted knowledgeable individuals in each district—typically local journalists, current and former politicians, and current and former civil servants—and asked a series of questions about each inter-election period using a standardized instrument. Most of the information was collected by phone, although on occasion our research assistants travelled or were based outside of Kampala (e.g. in Gulu, Lira, Mityana, Masindi, Tororo, and Yumbe) to meet key informants in person. These data were collected in two rounds: between July 2009 and December 2009 and between April and June 2012. The second round of data collection was necessary to reduce missingness and to conduct reliability and validity checks for the information assembled in 2009. For each district, research assistants obtained information from at least two local experts, and if those two peoples' responses contradicted one another, the research assistant contacted two additional people. If three out of the four respondents provided the same information, that information was deemed reliable; if not, the research assistant contacted additional people.

While such reported accounts are not ideal, official records of such information are generally unavailable, which may explain why this is the first decentralization study in Uganda using sub-district data. Further, the information we collected is generally common knowledge within districts—particularly among individuals who served as key informants and those who are politically and socially active in a district's civic affairs. As our paper indicates and our fieldwork in Uganda suggests, inter-county politics is a highly salient local issue, and thus politically-active individuals are aware of details such as the home county of prominent local officials. This may explain why in places where we asked two research assistants to collect the same information independently, we found very high rates of inter-coder reliability.

In order to code a county's share of seats in the DEC, our research assistants asked key informants to name all those who served on the DEC in a given inter-election period, including each member's ethnicity and "home county." We use the term home county rather than county of residence since some officials keep a residence in the district headquarters, while originating from a different, more remote county. The home county of district chairperson candidates is well-known locally, and generally is the area where the candidates' extended family resides. Note that the DEC is typically comprised of between five and seven members.

¹For district chairperson elections, research assistants collected the following information: the name, home county, and ethnicity of the winner and runner-up, whether the majority of each county's voters supported the winner or the runner up, and whether the winner of the district chairperson election ran unopposed.

1.2 Dependent Variable Coding

Our key dependent variable (*Splinter county*) denotes whether county j split to form a new district in electoral wave t ; that is, in one of three periods between elections in Uganda (local and presidential elections occur at the same time). Those periods are from 1996 to 2001, from 2001 to 2006, and from 2006 to 2011. To determine a county's split status, we used Ugandan news sources (particularly articles from *New Vision* and *Monitor*), Statoids,² and where discrepancies existed, interviews with local officials from the areas in question and information provided by the Ministry of Local Government.

In most cases, districts were requested by the District Council, approved and announced by the government, and created—meaning a new district headquarters was built and staffed—all within a single inter-election period. However, in some cases the requests and central government announcement came before an election, whereas the creation of the district occurred only after the election. This was the case for seven districts whose creation was approved and announced prior to the 2006 elections, but came into being in the months following those elections.³ In these cases we decided to consider the split as occurring when the new districts were approved and publicly announced by the central government. This is the case because, from a theoretical standpoint, when examining the determinants of district splits, we seek to isolate whether factors (such as the composition of the DEC) that existed during a given inter-election period influence decisions at the local and central government levels regarding district splits in that period— or for electoral variables, whether events in a particular election influence such decisions in the following period. Since it is the decision at the local level to request and approve a new district that interests us, we isolate this decision in coding our dependent variable. As far as we know, there are no cases in which the Ugandan government announced the creation of a district and the district was not created.

²Available at www.statoids.com

³These seven districts are Budaka, Bukedea, Bulamogi, Buliisa, Kabula, Kongasis, and Manjiya.

1.3 Descriptive Statistics

Table 1 provides descriptive statistics of the variables used in the regression analysis.

Table 1: Descriptive Statistics Table

	Mean	sd	Min	Max	Observations
Dependent variables					
Splinter county	0.185	0.389	0	1	460
Museveni county vote share	0.661	0.232	0.0249	0.988	460
Key independent variables					
Share of DEC seats	0.354	0.295	0	1	460
Support LC5 loser	0.220	0.414	0	1	460
Ethnic marginalization	0.0848	0.279	0	1	460
Development index	-0.0105	0.849	-1.119	3.001	460
Control variables					
LC5 Elections Opposed	0.902	0.297	0	1	460
Ethnic Fractionalization	0.365	0.253	0.0115	0.910	460
Breakup lag	0.330	0.471	0	1	460
Log county population size	11.73	0.616	8.928	13.12	460
Log county area size (km ²)	6.785	1.259	1.564	9.163	460
N. counties in district	3.754	1.830	1	9	460

1.4 Spatial Analysis and Distribution of District Splits

The homoskedastic version of the likelihood in Bhat and Sener, which we used for our spatial robustness check, can be written as

$$L(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^N \prod_{t=1}^T \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{it}) y_{it}}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_{it})} \times \left[1 + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{t=1}^T (-1)^{y_{it}+y_{jt}} \theta_{ijt} \left\{ 1 - \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{it}) y_{it}}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_{it})} \right\} \left\{ 1 - \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{jt}) y_{jt}}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_{jt})} \right\} \right]$$

where $\theta_{ijt} = \tanh\left(\frac{\gamma_1 e_{ijt} + \gamma_2 p_{ijt} + \gamma_3 f_{ijt} + \gamma_4 s_{ij}}{D_{ij}}\right)$ is a parameter that reflects spatial-temporal dependence between counties i and j . The hyperbolic tangent function, \tanh , maps its argument from the $(-\infty, \infty)$ interval to the $(-1, 1)$ interval, which is a necessary constraint for this likelihood. D_{ij} is the distance between counties i and j while e_{ijt} , p_{ijt} , f_{ijt} , and s_{ij} are dummy variables indicating whether both i and j are similar “as of” electoral wave t where e_{ijt} is equal to one if both county i and county j were below the median level of economic development during or prior to wave t . Similarly, p_{ijt} equals one if both counties i and j were below the median level of political representation during or prior to wave t , while f_{ijt} equals one if county i and j were both ethnically marginalized during or prior to wave t . Finally, s_{ij} indicates whether counties i and j were in the same district at the start of the first inter-election period (1996-2000). This specification for θ_{ijt} reflects a theory that similarly marginalized counties that are geographically close are more likely to splinter from their districts. The likelihood for a typical logit model is a special case of the above likelihood where $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$. The logarithm of this likelihood is maximized subject to the constraint that the term in square brackets is positive. As mentioned above, we find a relatively weak spatial correlation: a likelihood ratio test fails to reject the null hypothesis that all four gamma coefficients are zero (p-value=0.13).

2 Additional Tables

Table 2: District Splits Robustness check I (alternative vars)

	M6	M7	M8
DEC share /population share	-0.81*** (0.20)	-0.78*** (0.20)	
County Share of seats in DEC			-3.55*** (0.84)
Social services marginalization	-1.70*** (0.66)		
Economic marginalization index		-1.00*** (0.29)	
Development index			-1.56*** (0.50)
Ethnic marginalization	1.60*** (0.46)	1.62*** (0.47)	1.48*** (0.46)
N. counties in district	0.53*** (0.10)	0.54*** (0.11)	0.42*** (0.11)
Breakup lag	-0.80** (0.37)	-0.85** (0.37)	-0.88** (0.38)
Support LC5 chair loser	0.42 (0.34)	0.39 (0.34)	0.31 (0.34)
LC5 Elections opposed	0.51 (0.56)	0.37 (0.56)	0.21 (0.57)
Museveni vote share past election	-4.02 (2.90)	-4.99* (2.79)	-5.01* (2.82)
Museveni vote share past election ²	4.75* (2.65)	5.53** (2.55)	5.49** (2.57)
Log county population (census, 2002)	0.02 (0.29)	-0.21 (0.30)	0.18 (0.31)
Ethnic Fractionalization (census, 2002)	-0.93 (0.81)	-0.47 (0.79)	-0.76 (0.81)
Eastern Region	0.01 (0.52)	-0.31 (0.53)	-0.17 (0.55)
Northern Region	-0.98 (0.66)	-1.23* (0.65)	-1.24* (0.69)
Western Region	-1.74*** (0.54)	-1.92*** (0.55)	-1.71*** (0.55)
Electoral waves indicators	X	X	X
Intercept	-4.47 (3.65)	-1.18 (3.82)	-5.00 (3.78)
Observations	428	428	428
AIC	351.96	352.62	348.25
Log Likelihood	-157.98	-158.31	-156.12

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: District Splits Robustness check II (demeaned vars)

	M9	M10
DEC share /population share (between)	-0.43* (0.25)	-0.47* (0.26)
DEC share /population share (within)	-1.13*** (0.36)	-1.37*** (0.38)
Ethnic marginalization (between)	3.72** (1.57)	3.69** (1.56)
Ethnic marginalization (within)	0.33 (0.64)	0.33 (0.69)
Development index	-1.58*** (0.51)	-1.79*** (0.53)
Support LC5 chair loser (between)	0.44 (0.49)	0.52 (0.49)
Support LC5 chair loser (within)	-0.27 (0.63)	-0.50 (0.65)
LC5 Elections opposed (between)	1.14* (0.67)	1.03 (0.69)
LC5 Elections opposed (within)	-0.08 (1.09)	-0.05 (1.15)
Museveni vote share past election (between)	-9.79* (5.50)	-10.75* (5.59)
Museveni vote share past election (within)	-5.37 (3.72)	-4.10 (3.84)
Museveni vote share past election ² (between)	11.76** (5.31)	12.82** (5.41)
Museveni vote share past election ² (within)	5.27 (3.38)	3.94 (3.49)
N. counties in district (between)	1.41*** (0.24)	1.49*** (0.25)
N. counties in district (within)	0.09 (0.15)	0.05 (0.15)
Breakup lag (between)	-0.73 (0.48)	-0.72 (0.49)
Breakup lag (within)	1.77 (1.20)	2.21* (1.23)
Regional indicators	X	X
Electoral wave indicators	X	X
County demographic controls	X	X
Intercept	-3.48** (1.64)	5.04 (4.17)
Observations	428	428
AIC	325.45	323.51
Log Likelihood	-138.72	-135.75

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: District Splits: Separate Regression for Each Wave

	1996-2000	2001-2005	2006-2010
DEC share /population share	-0.56* (0.29)	-3.20*** (0.87)	-0.69* (0.41)
Ethnic marginalization	1.94* (1.02)	1.32 (0.98)	1.44 (0.95)
Development index	-2.79* (1.53)	-1.52* (0.84)	-1.61* (0.83)
Support LC5 chair loser	1.60** (0.70)	-0.96 (0.84)	0.10 (0.61)
LC5 Elections opposed	-1.97 (2.06)	-0.43 (1.28)	2.74* (1.56)
Museveni vote share past election	-5.62 (7.62)	-0.27 (7.64)	-2.69 (5.35)
Museveni vote share past election ²	7.00 (6.74)	0.61 (6.22)	4.24 (5.34)
Log county population (census, 2002)	0.79 (0.74)	-0.74 (0.63)	-0.87 (0.65)
Ethnic Fractionalization (census, 2002)	1.91 (1.84)	-4.52** (2.11)	-1.65 (1.47)
N. counties in district	0.63*** (0.23)	0.74*** (0.28)	0.56* (0.30)
Breakup lag	0.00 (.)	-0.83 (0.75)	-1.11* (0.65)
Eastern Region	0.33 (1.74)	-1.34 (1.34)	0.11 (1.08)
Northern Region	0.31 (2.05)	-2.88* (1.56)	-1.06 (1.32)
Western Region	-1.50 (1.58)	-3.22* (1.66)	-1.81* (0.98)
Intercept	-14.78 (9.69)	9.39 (8.09)	6.76 (7.51)
random Intercept	0.86 (0.94)	0.00 (0.41)	0.67 (0.56)
Observations	148	147	133
AIC	113.53	107.61	148.19
Log Likelihood	-41.76	-37.81	-58.09

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$