[Appendix B. *For online publication only*] Supplement to "Liberation Technology: Mobile Phones and Political Mobilization in Africa"

Marco Manacorda* Andrea Tesei†

This version: October 2019

B.1 Descriptive evidence on trends in coverage and protests

In this section we present additional evidence on the trends in coverage and protests across countries. Figure B.1 shows trends in mobile phone coverage by country. Continent-wide coverage starts from a value of 8.8 percent in 1998, reaching 64 percent in 2012. This very fast continental growth masks large differences across countries. The figure shows that among early adopters, such as Morocco and South Africa, coverage was virtually ubiquitous by the end of the period. This is in contrast with countries like Ethiopia and Mali where, as of 2012, still less than 10 percent of the population was covered.

Turning to protests, Figure B.2 reports the evolution in log protests per capita (plus 1 to account for zeros) measured in GDELT separately by country. As the range of variation of this variable is very different across countries, we standardize these series to their value in 1998. One can observe an increase in protests around 2008 in countries like Madagascar and Guinea that experienced food riots. The variation in the data is - in all cases - dwarfed by the very rapid surge in protests at the beginning of the current decade, with clear spikes in countries like Egypt, Libya and Tunisia, where the Arab Spring took place.

Figure B.3 reports the cross-sectional correlation between mobile phone coverage and protests per capita in GDELT (again expressed in logs of protests per capita plus 1) across all countries. Data are averages across the period for each country weighted by population weights. The data illustrate a clear positive correlation between these two series, with countries with full coverage, such as South Africa, showing rates of protests per capita around 50 log points higher than countries with virtually no coverage, such as Ethiopia. Results, not reported, are similar when using ACLED and SCAD.

* Queen Mary University of London, CEPR, CEP (LSE), IZA & CESifo; contact: m.manacorda@qmul.ac.uk.

† Queen Mary University of London, CEPR, CEP (LSE) & CESifo; contact: a.tesei@qmul.ac.uk.

B.2 Validation of data on protests from different sources

In this section, we investigate the correlation between different data sources on protests occurrence. Figure B.4 reports the evolution in protests per capita measured in GDELT and ACLED, separately by country. As the scale of the different series varies across countries, we report the residuals from regressions of log protests per capita (plus 1, to account for zeros) on country dummies and year dummies. By taking residuals, we also account for differential reporting probabilities across sources, countries and time. Despite the difference in scale (note that the ranges of variation on the left- and right-hand axes are different), one can appreciate a very strong positive correlation between the two series in most countries. In countries such as Burkina Faso, Madagascar and Tunisia, to name a few, one can see that the series line up remarkably well. This is less true in other countries such as Algeria, Benin or Ghana. Note that the series refer to average protests per capita in each country/year.

As our analysis ultimately focuses on cells within countries, we also explore the correlation between protests per capita from the different sources across these cells. Figure B.6 reports on the vertical axis the intensity of protests per capita measured in GDELT and on the horizontal axis the intensity of protests from ACLED. Both series are obtained as residuals of logs of the relevant variables (plus 1 to account for zeros) on cell and country fixed effects, separately for each country. Regressions are weighted by population size. We superimpose on the data an estimated regression line, separately for each country. The pooled regression coefficient across all countries alongside the associated standard error clustered at the level of cell is reported at the bottom of the figure. One can clearly see that, even within countries, there is a very clear positive correlation between the two series. This is true in almost all countries, and the pooled regression coefficient of protests from GDELT on protests from ACLED is 1.703 (s.e. 0.142). We obtain qualitatively very similar results when we compare GDELT and SCAD in Figures B.5 and B.7. Taken together these figures suggest that, despite some unavoidable measurement error, the three data sets convey very similar information.

B.3 Afrobarometer data

B.3.1 Assignment of Afrobarometer observations to grid cells

In the analysis we use data from Afrobarometer, rounds 3 to 5, spanning from 2005 to 2012 (information on available data by country and round is reported in Table B.1 below). We have assigned individuals in Afrobarometer to the 0.5° X 0.5° cells following the procedure described below.

The Afrobarometer provides information on individuals' country, district and town/village of residence. First, we match observations in the Afrobarometer to data from GeoNames. We restrict to populated places in GeoNames (*i.e.*, we exclude, for example, mountains or

lakes), defined as towns, villages or other places where people live and work. GeoNames also provides alternate names for each place, which are typically other names by which the place is known or the name in the local language.

We match observations in the Afrobarometer based on their town/village of residence to place names in GeoNames. If an observation does not match, we match sequentially on the first, second and third alternate place name in GeoNames. For unmatched observations, we proceed sequentially replicating the same procedure but matching on districts of residence in the Afrobarometer.

As a second step, we assign place names in GeoNames to the 0.5° X 0.5° cells. Importantly, even within a country, the same place name in GeoNames can be shared by more than one populated place, meaning that we cannot always uniquely assign a place name to a cell. When two places share the same name and hence potentially belong to more than one cell, we expand the dataset and we assign that place to each of these multiple cells. We construct an adjustment factor for each observation in this dataset so that a place name X cell has an associated weight equal to the relative population of a cell expressed as a fraction of the total population among all cells to which that given place can potentially belong to. Clearly, for cells that are univocally assigned to a cell this population, this weight is equal to 1.

Combining steps one and two, we are able to assign observations in the Afrobarometer to grid cells. The resulting dataset has a number of observations larger than the original Afrobarometer, as individuals whose place of residence can potentially belong to different cells will have as many observations in the data as the potential cells of residence. Afrobarometer data include sampling weights. We rescale sampling weights by the population weights described above. This is equivalent to assuming that these individuals have been sampled at random among all those living in all the potential cells of residence and guarantees that the sum of sampling weights in this new dataset is the same as in the original Afrobarometer dataset.

In total we are able to assign 78,167 individuals in Afrobarometer to at least one cell, or 81 percent of total respondents. In total 49 percent of matched individuals have a unique cell identifier, while the rest are assigned to at least two cells.

B.3.2 Procedure to predict mobile phone use in Afrobarometer

The Afrobarometer only provides information on mobile phone use for observations in rounds 4 and 5. In addition, the phrasing of the question differs across rounds. We first characterize the determinants of mobile phone use in round 5 of Afrobarometer by running an ordered probit model of frequency of mobile phone use (5 categories, ranging from "never" to "several times a day") on a number of socio-economic characteristics, country plus year fixed effects, and the fraction of the population in reach of signal in the cell from GSMA. The estimated coefficients are reported in column (2) of Table B.2 below. Mobile

phone use increases and then decreases with age, increases with education and is higher for males compared to females and in cities compared to the countryside. Importantly, it is also strongly positively correlated with mobile phone coverage from GSMA. We use estimates from this model to predict mobile phones usage for all individuals in the Afrobarometer, including in rounds 3 and 4. We assign to each individual a dummy equal to 1 if the estimated probability of using a mobile phone at least once a day exceeds 50 percent. For all observations in round 4 that report using a mobile phone every day we assign a probability of one irrespective of the model prediction. Based on this procedure, 71 percent of the individuals are predicted to use a mobile phone.

B.3.3 Descriptive statistics Afrobarometer

Table B.3 presents descriptive statistics on the Afrobarometer data. In total, the sample is composed of 78,167 individuals, living in 4,332 cells. The upper panel reports individual-level characteristics, while the bottom panel the corresponding cell characteristics. Cells in the Afrobarometer are more populated compared to the average cell in Africa. They also display higher than average mobile phone coverage, at 68 percent of the cell area. Consistently, the fraction of population reporting using a mobile phone at least once a day is 71 percent. During the period of observation, on average 12 percent of the Afrobarometer respondents report having participated in at least one protest in the previous year, and between 65 and 70 percent report distrusting or disapproving of the president. Those more likely to participate in a protest are young (peak at age 33 years), educated men - see column (1) of Table B.2.

B.3.4 OLS results in Afrobarometer cells

Table B.4 below reports OLS estimates of model (5.1) in the text on the sample of cells/years available in the Afrobarometer. Regression results based on GDELT, ACLED and SCAD in the first three columns are qualitatively similar to the ones obtained on the entire sample (see Appendix Table A.3). Importantly, Afrobarometer data on the fraction of people participating in protests in column (4) are qualitatively in line with those from GDELT, ACLED and SCAD. The estimates suggest that a 1 s.d. fall in GDP growth is associated with an increase of around 3.5 p.p. in the protest differential between areas with full and zero coverage (-0.874 x 0.04). At a baseline protest participation of around 12 p.p. this is equivalent to an increase of around 29 percent.

B.4 Misreporting of protests

2SLS estimates of the coefficient on the interaction term in Table 1 are systematically larger for GDELT compared to ACLED and SCAD. An explanation for this is systematic misreporting of protests. We claim in the paper that all datasets are likely to suffer from both type-1 and type-2 measurement error. Here we investigate the consequences of these errors for the estimates of our regression model.

Let us focus on the occurrence of at least one protest per day in a cell (*one per day*). This somewhat simplifies the derivation. Let $y_i^* = 0/1$ be a latent variable that denotes whether one protest occurs in a cell in a given day i and let $y_i = 0/1$ be a variable denoting whether a protest is recorded in the data. The probability that an event is recorded depends on both type-1 and type-2 errors, respectively denoted by $\alpha_1 = P(y_i = 0|y_i^* = 1)$ and $\alpha_2 = P(y_i = 1|y_i^* = 0)$. It follows that the number of reported protests per cell per year is:

$$\begin{aligned}\Sigma_i Pr(y_i = 1) = \Sigma_i[Pr(y_i = 1|y_i^* = 1)Pr(y_i^* = 1) + Pr(y_i = 1|y_i^* = 0)Pr(y_i^* = 0)] = \\ = (1 - \alpha_1 - \alpha_2)\Sigma_i Pr(y_i^* = 1) + 365\alpha_2\end{aligned} \tag{B.4.1}$$

Let $p^* = \frac{\Sigma_i Pr(y_i^*=1)}{365}$ denote the true average incidence of protests over a year. Hence the log number of reported protests, the dependent variable in the regressions, is:

$$ln\Sigma_i Pr(y_i = 1) \approx cons + Aln\Sigma_i Pr(y_i^* = 1) \tag{B.4.2}$$

where, $A = \frac{(1-\alpha_1-\alpha_2)p^*}{(1-\alpha_1-\alpha_2)p^*+\alpha_2}$. As it seems plausible that $1 - \alpha_1 \geq \alpha_2$, *i.e.* that the probability that an event is reported as true if it is true is greater than the probability that it is reported as true when it is instead false, it follows that $0 \leq A \leq 1$. It follows that estimates based on the error ridden measure of the log number of protests will be attenuated by a factor A, *i.e.* that all data sets will provide conservative estimates of the parameters of interest.

Estimates using as a dependent variable the log *one per day* number of protests are reported in column (5) of Table 2 in the paper. Estimates are similar to those obtained when the dependent variables is the log count of total protests (in Table 1 in the paper) and patterns are similar across datasets, with the largest (in absolute value) point estimates obtained when using GDELT. This implies that - if the above model is the right model - A is the largest for GDELT.

One can also prove that A is larger the larger $\frac{1-\alpha_1}{\alpha_2}$. As A is larger for GDELT compared to ACLED and SCAD, this means that the relative probability that an event is reported as true when true relative to when is false is larger in GDELT than in the two other datasets. As we have good reasons to believe that α_2 is small in both ACLED and SCAD compared to GDELT, this suggests in turn that α_1 must be much larger in ACLED and SCAD than in GDELT. In sum, although GDELT is likely to suffer from considerable

type-2 measurement error, this is dwarfed by the extent of type-1 measurement error in the other two data sets.

Additional references for section 2 in the paper

BBC (2010), 'Mozambique 'Blocked Texts' During Food Riots', *BBC News Africa* .

Bush, R. (2010), 'Food Riots: Poverty, Power and Protest', *Journal of Agrarian Change* **10**(1), 119–129.

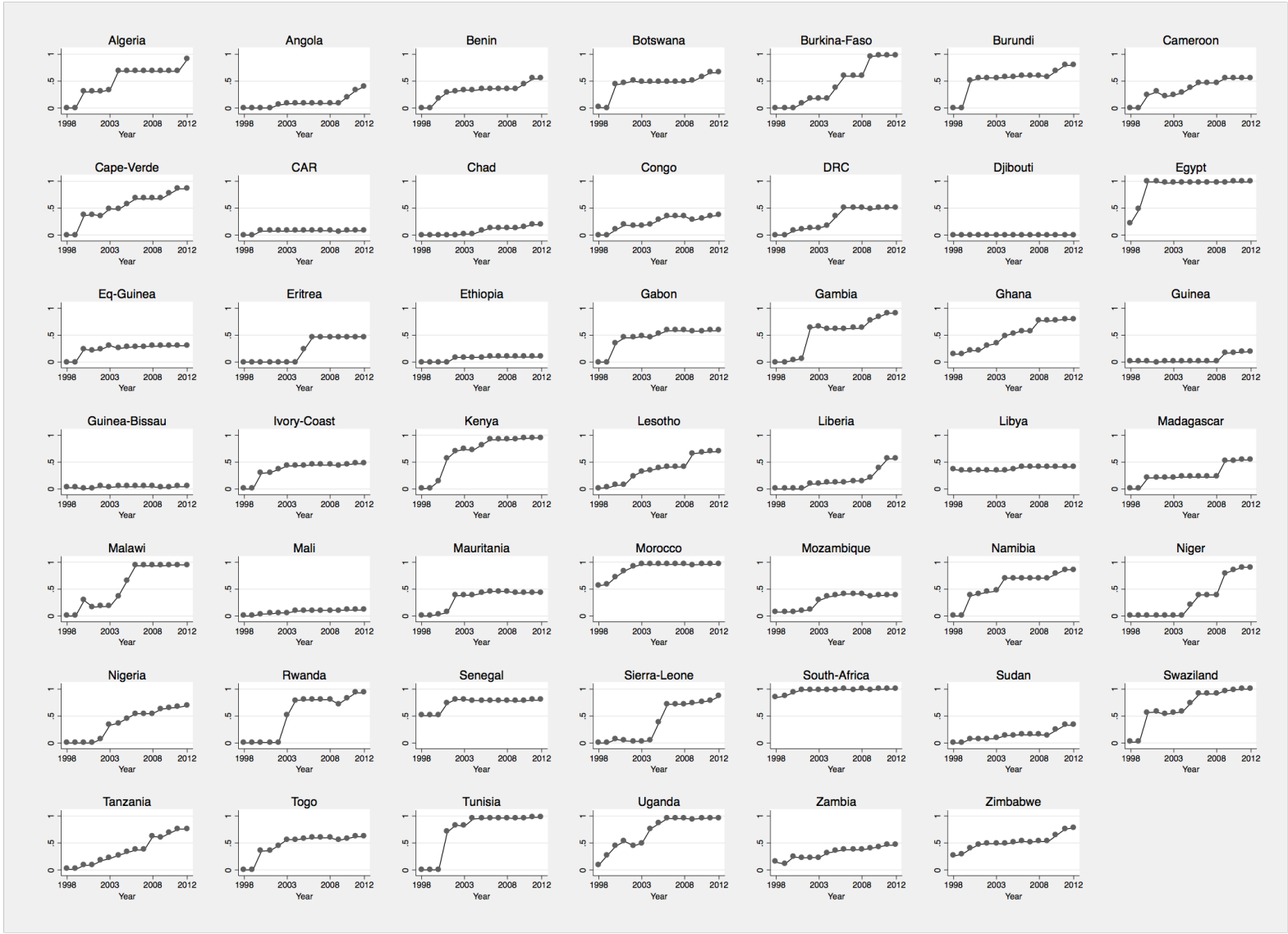
Engels, B. (2015), 'Different Means of Protest, Same Causes: Popular Struggles in Burkina Faso', *Review of African Political Economy* **42**(143), 92–106.

Fisher, M. (2011), 'In Tunisia, Act of One Fruit Vendor unleashes Wave of Revolution through Arab World', *Washington Post* **26**.

Harsch, E. (2008), 'Price Protests expose State Faults', *Africa Renewal* **22**(2), 15–21.

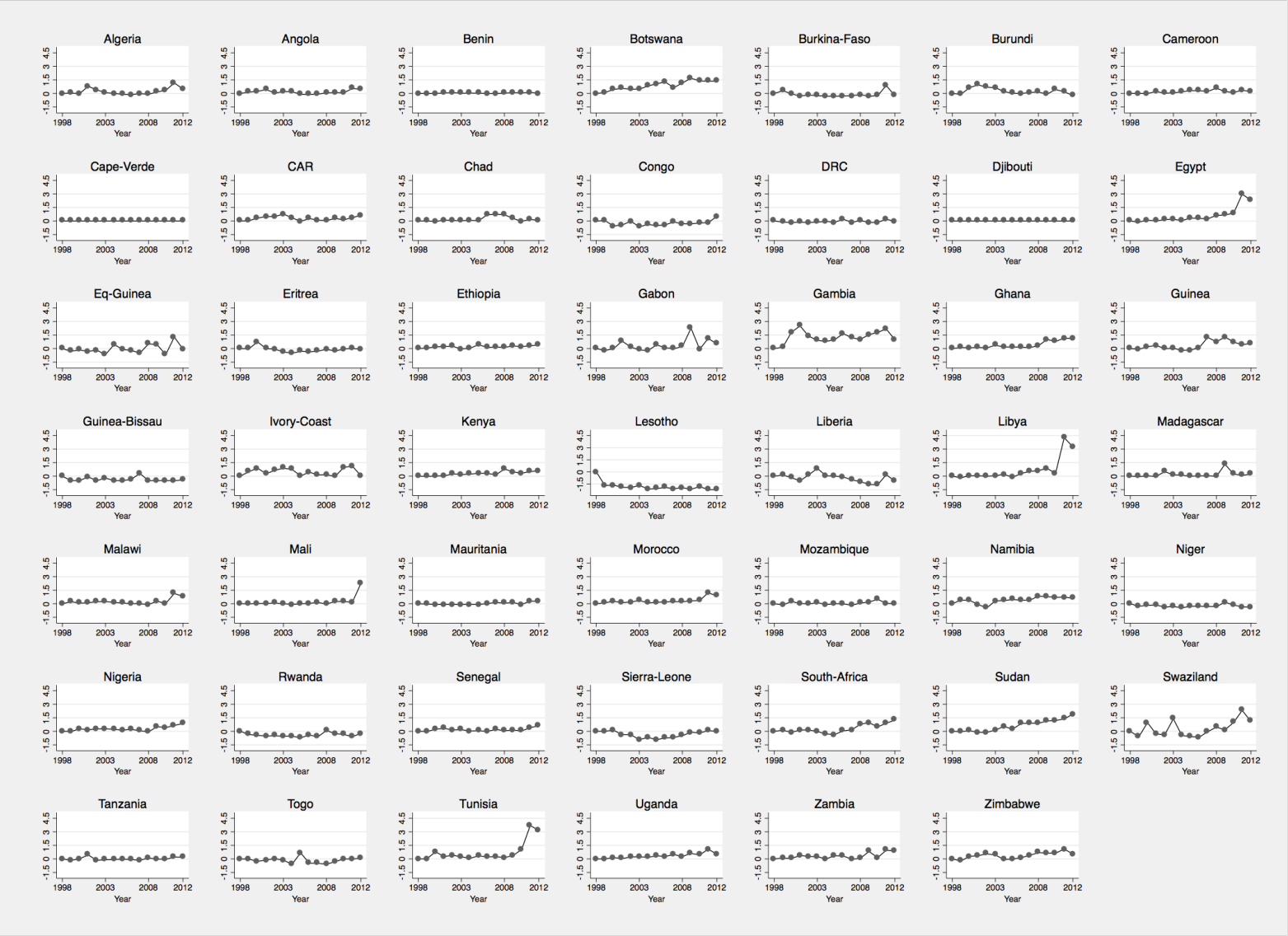
Howard, P. N. & Hussain, M. M. (2011), 'The Role of Digital Media', *Journal of democracy* **22**(3), 35–48.

Figure B.1 Trends in the fraction of population covered by mobile phone coverage by country



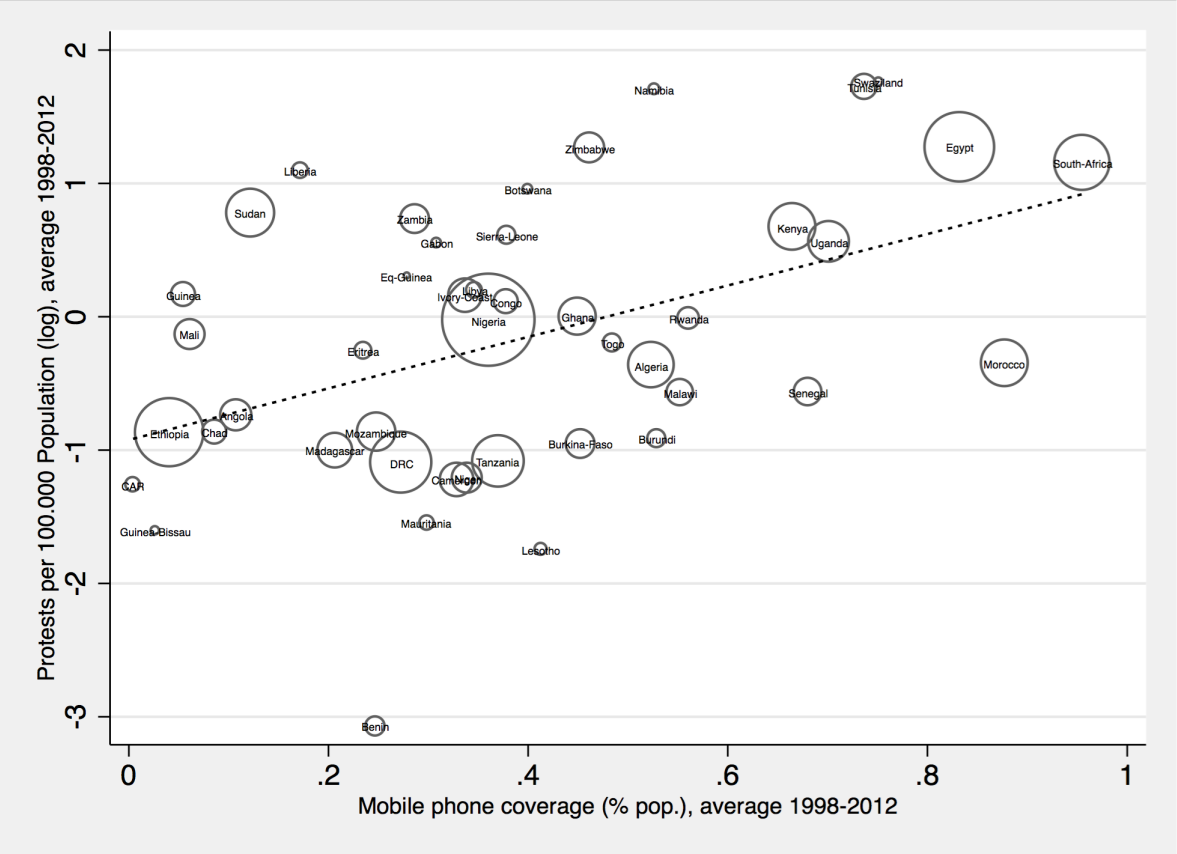
Notes. The figure reports the fraction of the population covered by mobile phone signal by country and time. Series are obtained as population-weighted averages of the fraction of each country's 0.5° X 0.5° degree cell that is covered by the signal in each year.

Figure B.2 Trends in protests by country - GDELT



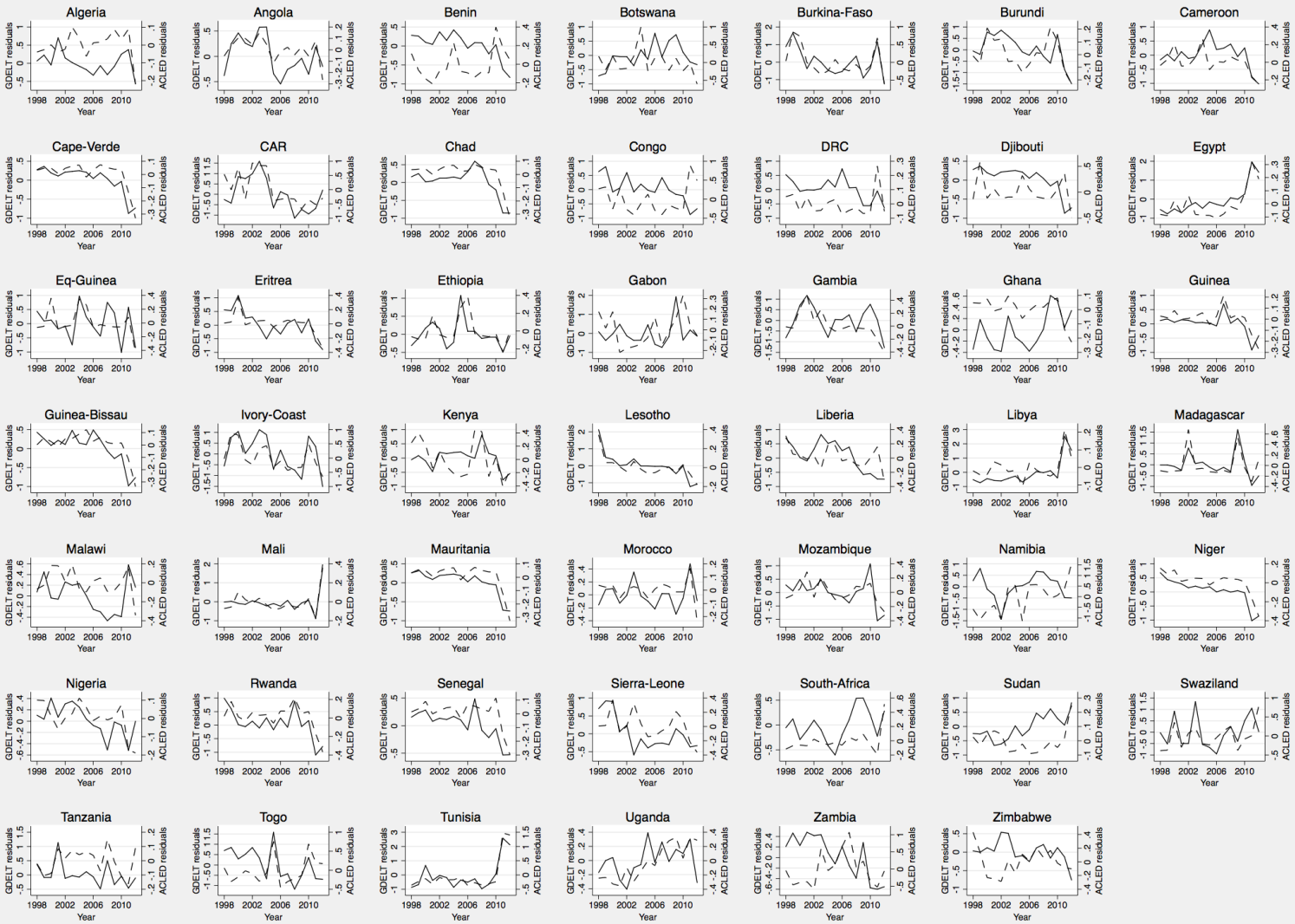
Notes. The figure reports the evolution of log protests per 100,000 individuals (plus 1) by country based on GDELT. All series are standardized to their value in the first year.

Figure B.3 Cross-sectional relationship between mobile phone coverage and protests



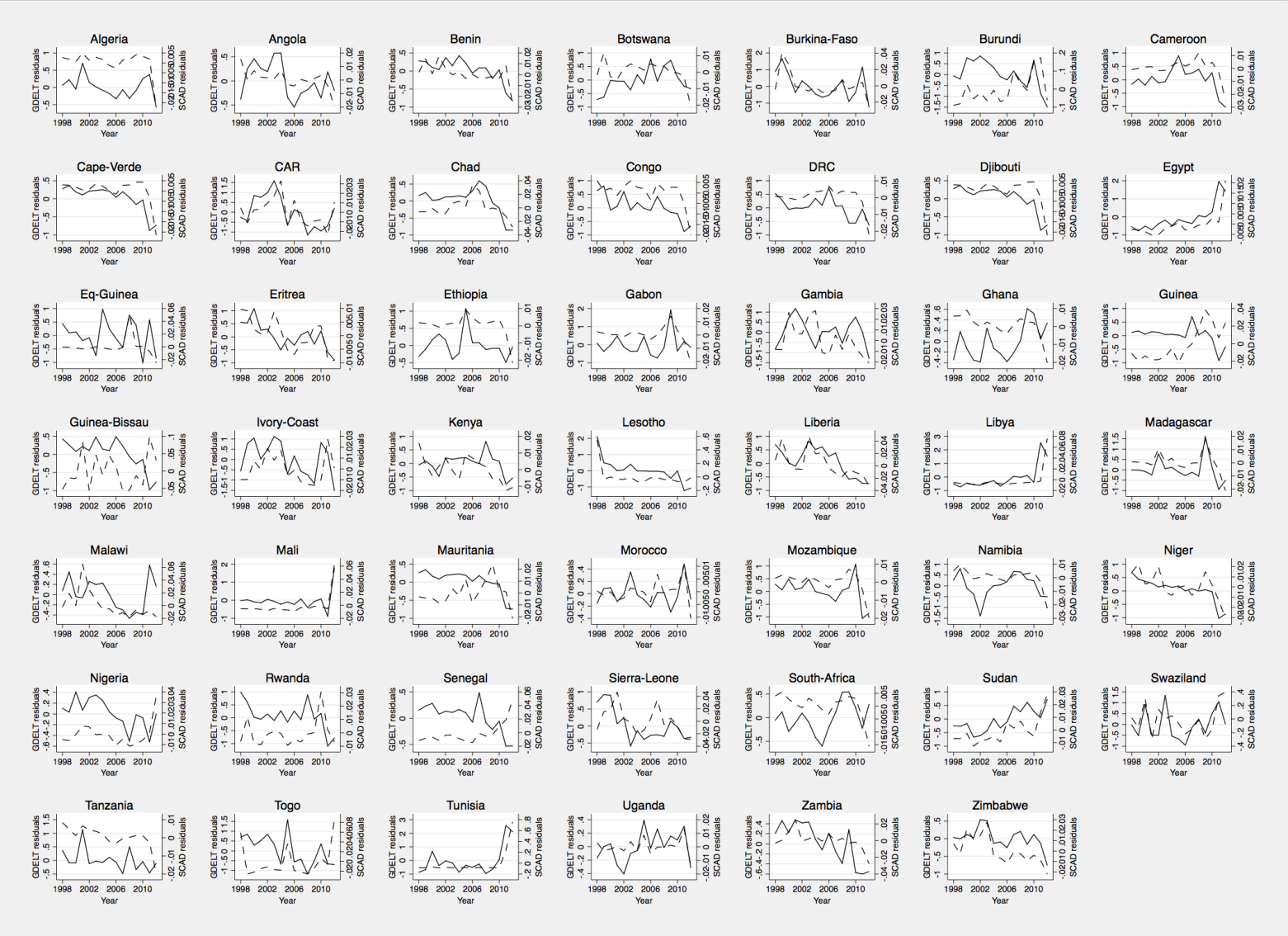
Notes. The figure reports log protests per 100,000 individuals based on GDELT on the vertical axis and the fraction of the population covered by mobile phone signal on the horizontal axis. Averages between 1998 and 2012 by country reported. The size of each circle is proportional to the country population.

Figure B.4 Correlation between reported protests in GDELT and ACLED across countries



Notes. The figure reports log protests per 100,000 individuals in GDELT (solid line) and ACLED (short-dashed line) by country and year. Residuals from regressions on country and year fixed effects reported.

Figure B.5 Correlation between reported protests in GDELT and SCAD across countries

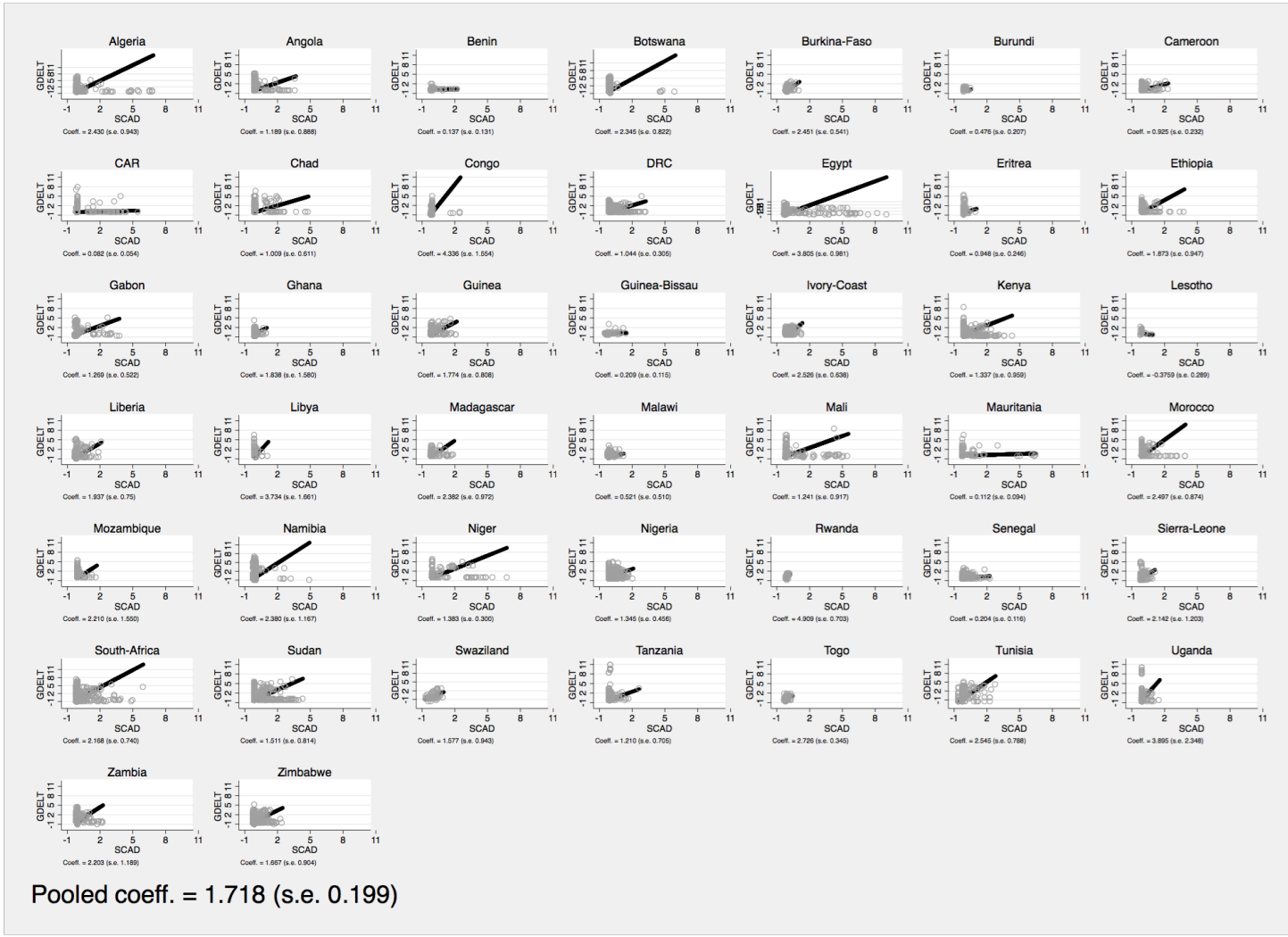


Notes. The figure reports log protests per 100,000 individuals in GDELT (solid line) and SCAD (short-dashed line) by country and year. Residuals from regressions on country and year fixed effects reported.

Figure B.6 Within-country correlation between reported protests in GDELT and ACLED

Notes. The figure reports the relationship between protests in GDELT (on the vertical axis) and ACLED (on the horizontal axis) within each country. Each point refers to a cell X year observation. All series are expressed in logs (plus 1 to account for zeros). Residuals from regressions on cell fixed effects and year X country fixed effects reported. A GLS best-fit regression line (and the associated slope coefficients and standard errors) of protests in GDELT on protests in ACLED with weights equal to the population in each cell in each year is also reported.

Figure B.7 Within-country correlation between reported protests in GDELT and SCAD



Notes. The figure reports the relationship between protests in GDELT (on the vertical axis) and SCAD (on the horizontal axis) within each country. Each point refers to a cell X year observation. All series are expressed in logs (plus 1 to account for zeros). Residuals from regressions on cell fixed effects and year X country fixed effects reported. A GLS best-fit regression line (and the associated slope coefficients and standard errors) of protests in GDELT on protests in SCAD with weights equal to the population in each cell in each year is also reported.

Table B.1 Afrobarometer country-rounds availability

	Round 3 (2005/06)	Round 4 (2008/09)	Round 5 (2011/13)
Benin	1,190 [37]	1,184 [33]	592 [33]
Botswana	1,182 [53]	920 [42]	880 [41]
Burkina-Faso	-	968 [53]	576 [40]
Burundi	-	-	1,200 [15]
Cameroon	-	-	656 [55]
Cape-Verde	765 [5]	656 [10]	719 [7]
Ghana	1,165 [71]	960 [60]	1,376 [66]
Guinea	-	-	1,136 [42]
Ivory-Coast	-	-	1,136 [57]
Kenya	1,246 [45]	960 [34]	2,135 [31]
Lesotho	1,161 [10]	1,192 [9]	1,197 [9]
Liberia	-	797 [28]	873 [25]
Madagascar	1,333 [191]	1,152 [183]	1,012 [216]
Malawi	1,199 [34]	1,152 [23]	1,523 [40]
Mali	1,187 [101]	960 [115]	986 [94]
Mozambique	1,198 [111]	1,088 [85]	1,936 [99]
Namibia	1,139 [82]	1,024 [49]	1,097 [52]
Nigeria	2,200 [193]	1,781 [197]	1,936 [182]
Senegal	1,200 [47]	1,030 [25]	1,176 [34]
Sierra-Leone	-	-	550 [28]
South-Africa	2,171 [212]	2,220 [188]	1,400 [130]
Swaziland	-	-	456 [7]
Tanzania	1,203 [102]	1,024 [68]	2,144 [94]
Togo	-	-	368 [14]
Uganda	2,400 [60]	2,431 [46]	1,444 [57]
Zambia	1,200 [103]	1,200 [68]	1,176 [71]
Zimbabwe	914 [44]	1,000 [42]	1,888 [48]

Notes. The table reports the number of individuals by country in rounds 3 to 5 of Afrobarometer. The number of cells identified for each country in each round is reported in parenthesis.

Table B.2 Individual correlates of protest participation and mobile phone use

	(1)	(2)
	Protest Participation	*Mobile phone use*
Age/100	0.148	2.392
	[0.06]*	[0.00]***
Age/100 *sq.*	-0.224	-3.317
	[0.01]***	[0.00]***
Female	-0.035	-0.234
	[0.00]***	[0.00]***
City	0.008	0.444
	[0.14]	[0.00]***
Adults in household	0.004	0.022
	[0.01]***	[0.00]***
Years of education	0.005	0.124
	[0.00]***	[0.00]***
Coverage		0.359
		[0.00]***
Observations	76,068	30,760

Notes. The table reports individual-level regressions based on Afrobarometer data. The dependent variable in column (1) is a dummy variable equal to 1 if the respondent attended a demonstration or protest during the previous year; in column (2) is an ordered variable for frequency of mobile phone use (from 0 or "never" to 4 or "several times a day"). This latter variable is only available for Round 5. Method of estimation in column (1): OLS; in column (2): Ordered Probit (marginal coefficients reported). Regression in column (1) includes the same controls as in columns (3) to (8) of Table 1 in the paper. Regression in column (2) includes country and year fixed effects. All regressions weighted by individual sampling weights. p-values for wild cluster bootstrap standard errors at the level of country are reported below each coefficient. See also notes to Table 1 in the paper.

Table B.3 Descriptive statistics Afrobarometer

	Avg.	Std. Dev.	Min.	Max.
		Individuals (78,167)		
Protest participation	0.12	0.32	0	1
Mobile phone	0.71	0.45	0	1
Age	36.85	14.73	18	130
Years of education	4.97	3.63	0	16
Gender	0.5	0.5	0	1
Christian	0.63	0.48	0	1
Muslim	0.2	0.4	0	1
Other religion	0.17	0.38	0	1
Adults in household	3.07	2.33	0	40
Worse economic condition (personal)	0.35	0.48	0	1
Worse economic condition (country)	0.38	0.49	0	1
Distrust president	0.65	0.48	0	1
Disapprove president	0.70	0.46	0	1
		Cells (4,332)		
Population (1000s)	246.02	431.49	0	7,497
Mobile phone coverage (percent)	0.68	0.36	0	1
Protests per $100,000$ *pop.* $-$ *GDELT*	1.99	6.98	0	528.17
Protests per $100,000$ *pop.* $-$ *ACLED*	0.18	0.74	0	206.83
Protests per $100,000$ *pop.* $-$ *SCAD*	0.07	0.30	0	9.12
Country GDP growth (percent)	0.06	0.02	-0.18	0.15

Notes. The table reports descriptive statistics for individuals in Afrobarometer (upper panel) as well as the corresponding cell characteristics (lower panel). Data in the upper panel are weighted by individual sampling weights. Data in the lower panel are weighted by cell population.

Table B.4 Protests, GDP growth and mobile phone coverage in Afrobarometer cells: OLS

	(1)	(2)	(3)	(4)
	$GDELT$	$ACLED$	$SCAD$	$Afrobarometer$
$Coverage$	0.594	0.166	0.176	-0.013
	[0.02]**	[0.40]	[0.01]***	[0.69]
$\Delta GDP\ X\ Coverage$	-8.981	-4.209	-2.459	-0.708
	[0.04]**	[0.23]	[0.03]**	[0.03]**
Observations	4,286	4,286	4,286	4,220

Notes. The table reports the same specifications as in Table A.3 of the typeset Appendix estimated for the sample of cells/years available in Afrobarometer, where the dependent variables are: log protests per 100,000 people from GDELT (column 1), ACLED (column 2) and SCAD (column 3); fraction participating in a protest from Afrobarometer (column 4). p-values for wild cluster bootstrap standard errors at the level of country are reported in parenthesis. See also notes to Table A.3.