Online Appendix

"Cell Phone Access and Election Fraud: Evidence from a Spatial Regression Discontinuity Design in Afghanistan"

Robert M. Gonzalez*

May 18, 2020

*Department of Economics, Moore School of Business, University of South Carolina, Columbia, SC 29208 (e-mail: robert.gonzalez@moore.sc.edu)

Appendix A The Audit and Recount Process

In its final report of the 2009 election (Electoral Complaints Commission, 2010), the ECC reported that after receiving an increasing number of complaints on ballot stuffing and other irregularities, they decided on September 8, 2009 to conduct an audit of polling stations nationwide. To this end, they ordered the IEC to conduct an audit and recount of stations satisfying the following criteria:

- A1: Stations in which 600 or more valid votes were cast
- B1: Stations with more than 100 votes in which one candidate received 95 percent or more of the total votes cast
- C1: Stations satisfying both A1 and B1

After the initial samples were drawn, however, three additional categories were created due to a misunderstanding of the ECC orders by the IEC[1]. The three new categories expanded the scope of the audit. The categories were:

- A2: Stations with 600 or more votes cast (Excluding those in A1)
- B2: Stations in which a candidate received 95 percent or more of the total valid votes cast (Excluding those in B1)
- C2: Stations satisfying both A2 and B2

After the audit process, the IEC reported that 3,376 stations classified in at least one of these categories. Out of this sample, the ECC and IEC investigated 10 percent of the ballots within each category. Some of the physical indicators used to determine fraud were whether the ballot box was tampered, all required materials were included, visual inspection of the ballots, reviews of the tally results and the actual ballot counts, among others. For the purpose of this study I aggregate the six categories described above into three broader categories:

- Category A: Stations with 600 or more votes cast. Defined as A1+A2+C1+C2 from the categories above.
- Category B: Stations in which one candidate received 95 percent or more of the total votes cast. Defined as B1+B2+C1+C2

[1] The misunderstanding was mainly due to the definition used to classify votes as "valid".

- Category C: Stations satisfying Categories A and B above. Defined as C1+C2.

The number of polling stations within each category is 1,706 in category A (545 from A1 + 299 from A2 + 741 from C1 + 121 from C2), 2,532 in category B (1269 from B1 + 401 from B2 + 741 from C1 + 121 from C2), and 862 in category C (741 from C1 + 121 from C2).

Appendix B Additional Figures and Tables

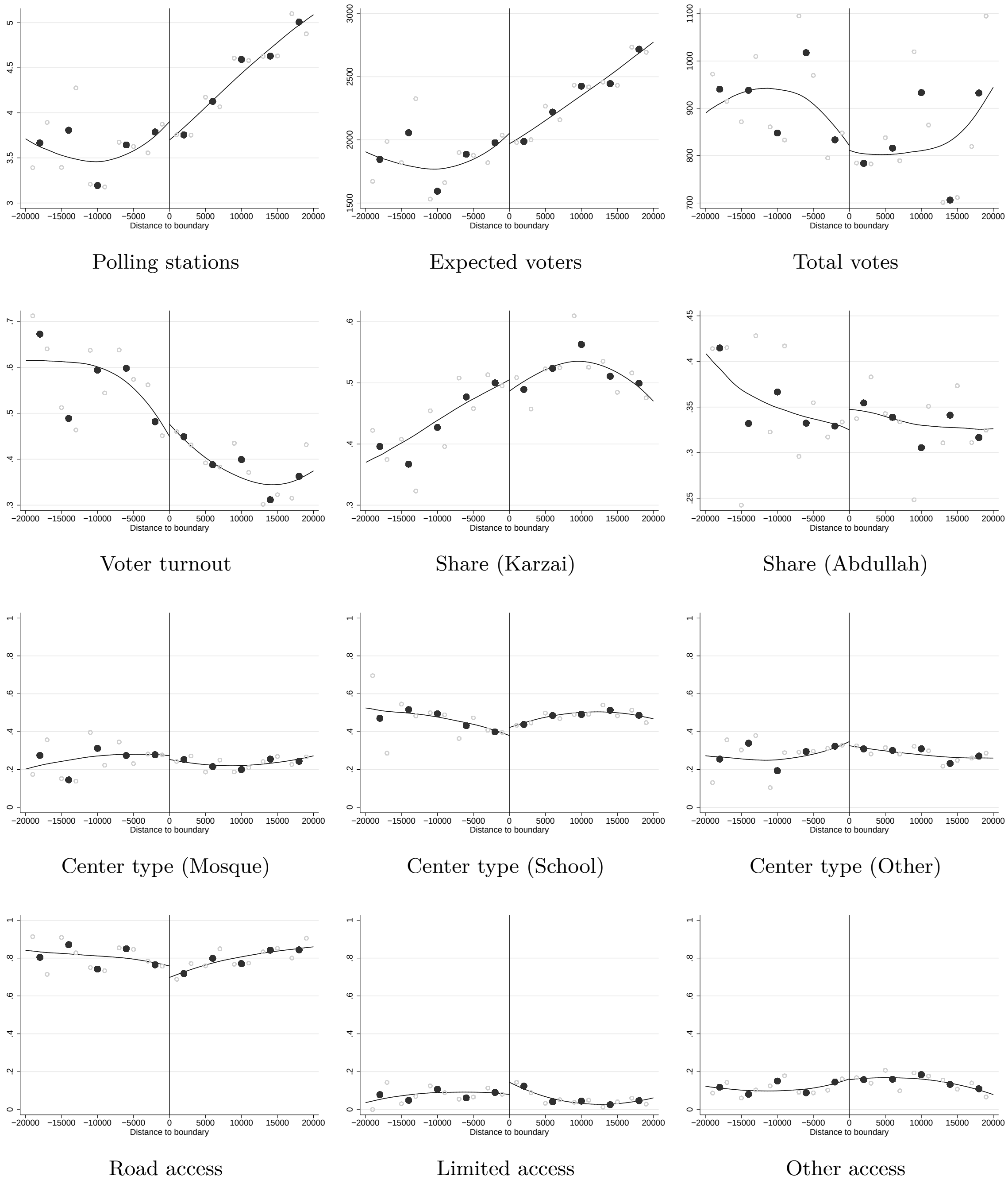
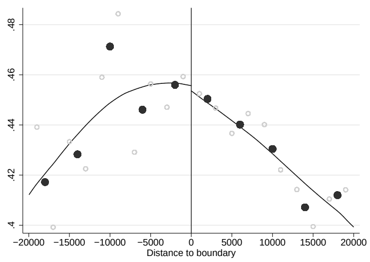
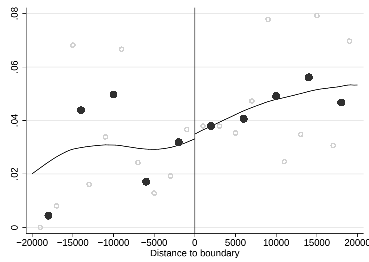


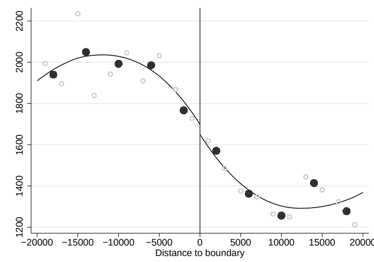
Figure B1: Binned Averages for various Covariates (Covariate RD Plots) - *Continues*



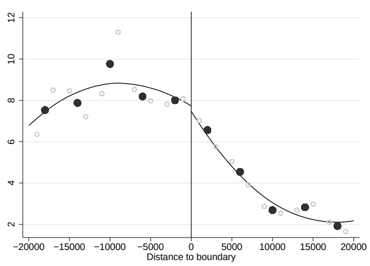
Female stations



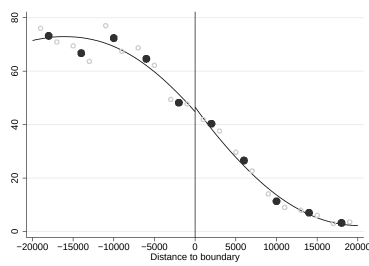
Kuchis stations



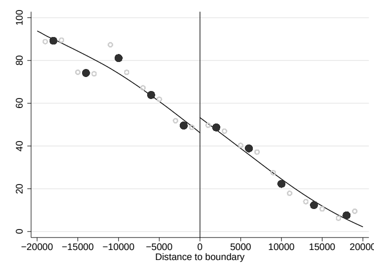
Elevation (m)



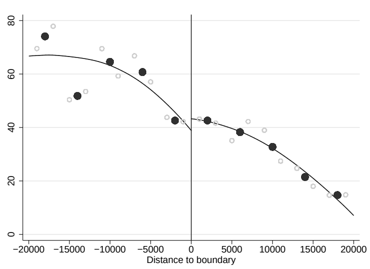
Slope (percent)



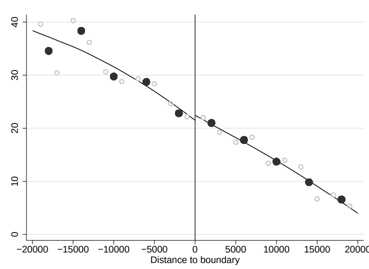
Distance (primary road)



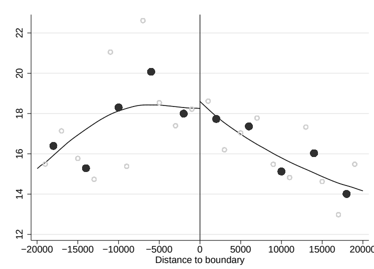
Distance (secondary road)



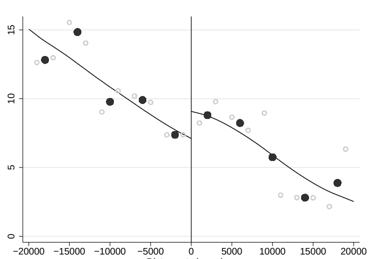
Distance (district hospital)



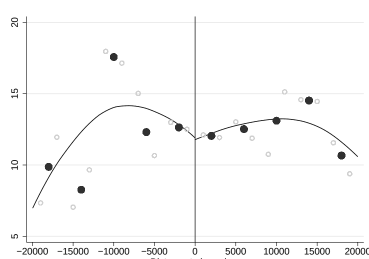
Distance (basic clinic)



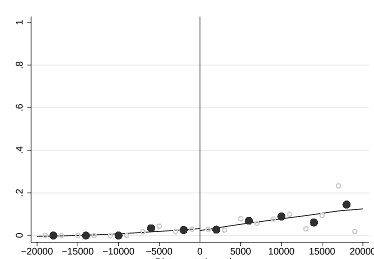
Distance (primary river)



Distance (secondary river)

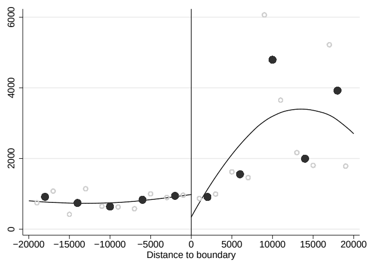


Distance (seasonal river)

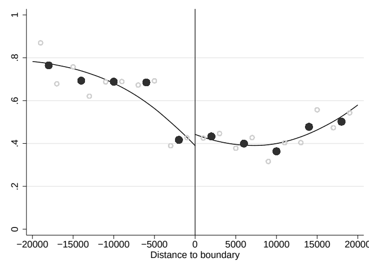


District/Provincial capital

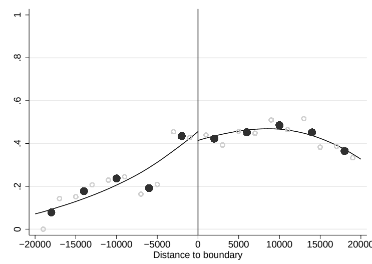
Figure B1: Binned Averages for various Covariates (Covariate RD Plots) - *Continues*



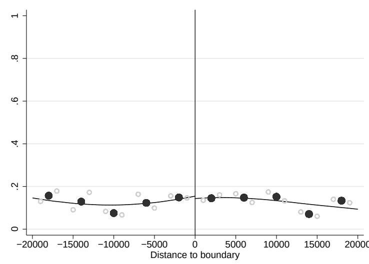
Population



Dari language



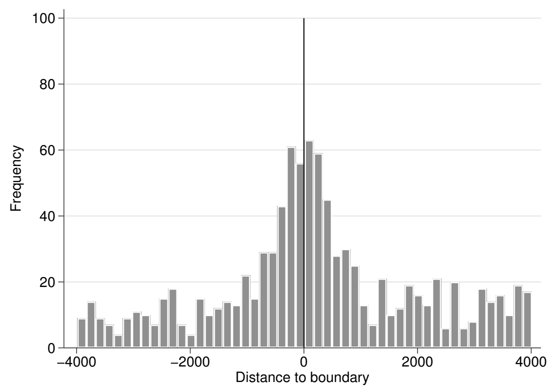
Pashto language



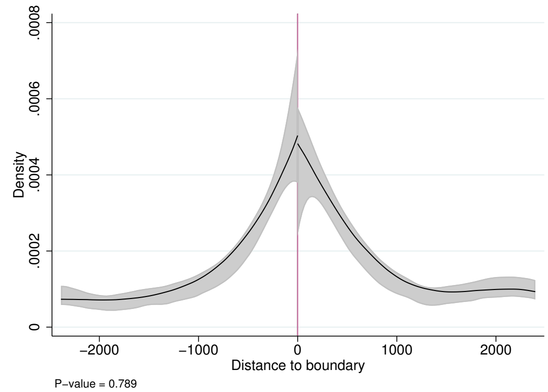
Other language

Figure B1: Binned Averages for various Covariates (Covariate RD Plots) - *Continued*

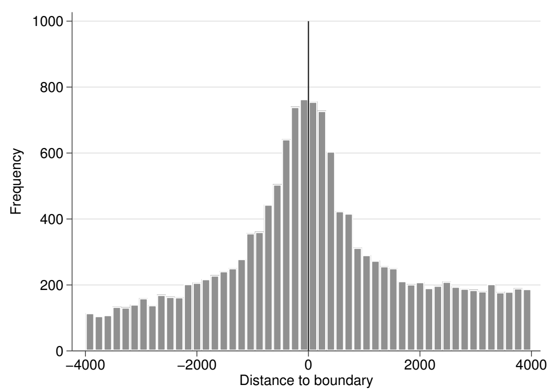
Notes: Solid dots give the average value of the specified covariate for polling centers falling within 4000-meter distance bins. Hollow dots give the average value of the specified covariate for polling centers falling within 2000-meter distance bins. “Distance to boundary” refers to the distance between a polling center and the closest point in the cell phone coverage boundary. “Negative” values of distance give the distance of polling centers/villages in non-coverage areas. The solid line trends give the predicted values from a regression of the outcome variable on a second degree polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 20,000 meters.



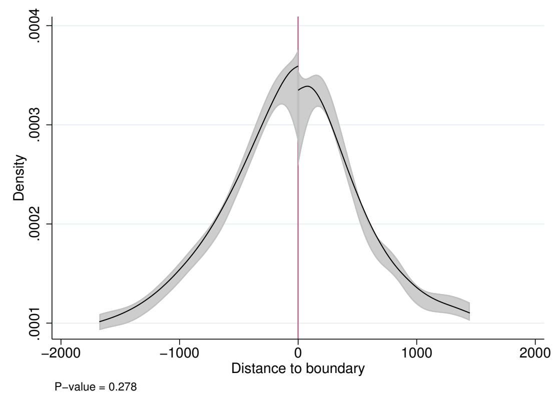
(a) Distance to coverage boundary (polling centers)



(b) Density breaks test (polling centers)



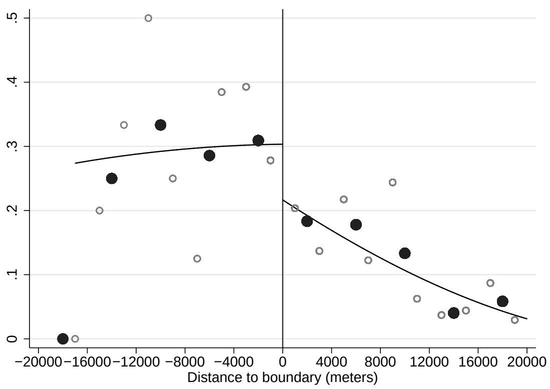
(c) Distance to coverage boundary (villages)



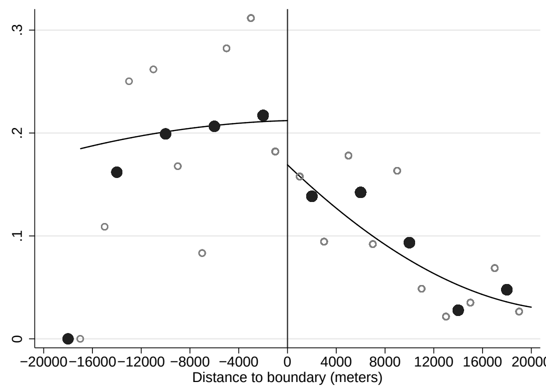
(d) Density breaks test (villages)

Figure B2: Histograms and densities of the forcing variable

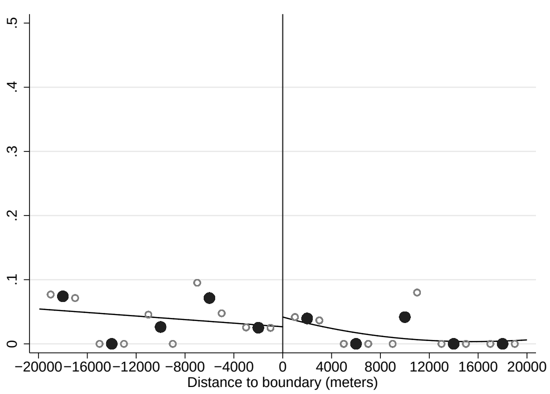
Notes: "Distance to boundary" refers to the distance between a polling center (Panels a and b) and villages (Panels c and d) to the closest point in the coverage boundary. Distance is measured in meters. Bin width of 160 meters. The distance to boundary is normalized so that "negative" values of distance give the distance of polling centers/villages in non-coverage areas. Panels (b) and (d) provide results from Panel (b) uses the test for breaks in the density of the forcing variable proposed in Cattaneo et al. (2019) and uses the code discussed in Cattaneo et al. (2018). P-value for test presented in figure caption.



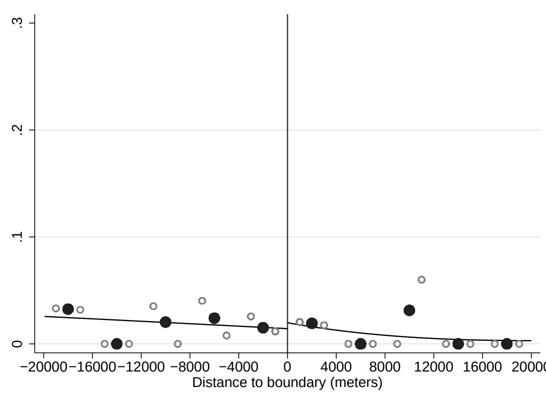
(a) At least one station with Category C fraud



(b) Share of votes under Category C fraud



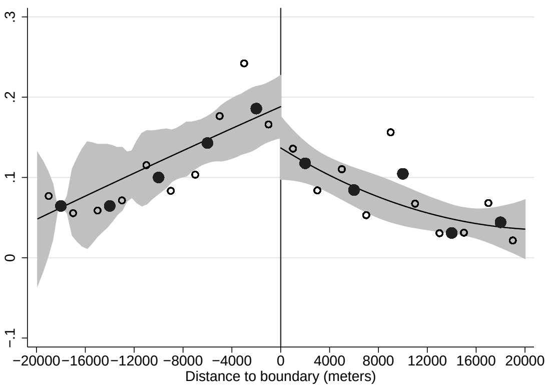
(c) At least one station with Category C fraud



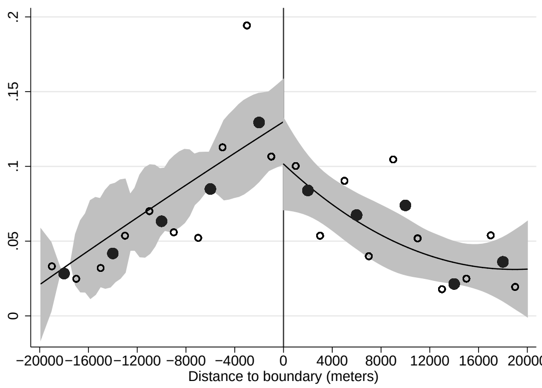
(d) Share of votes under Category C fraud

Figure B3: Binned Averages for Category C fraud (RD plot)

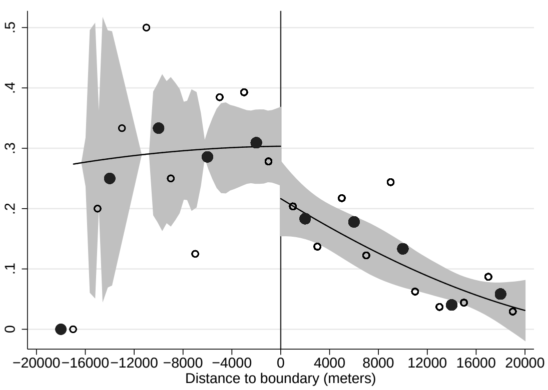
Notes: South/East region (Panels A and B). North/West region (Panels C and D). Solid dots give the average share of votes classifying in Category C fraud for polling centers falling within 4000-meter distance bins. Hollow dots give the average share of votes classifying in Category C fraud for polling centers falling within 2000-meter distance bins. Refer to section II in the text for a detailed description of Category C fraud. "Distance to boundary" refers to the distance between a polling center and the closest point in the cell phone coverage boundary. "Negative" values of distance give the distance of polling centers/villages in non-coverage areas. The solid line trends give the predicted values from a regression of the outcome variable on a second degree polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 20,000 meters.



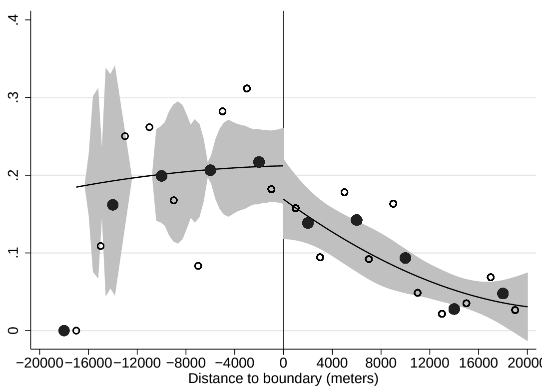
(a) At least one station with Category C fraud



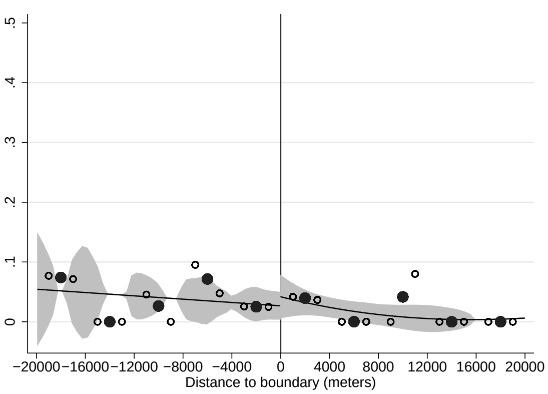
(b) Share of votes under Category C fraud



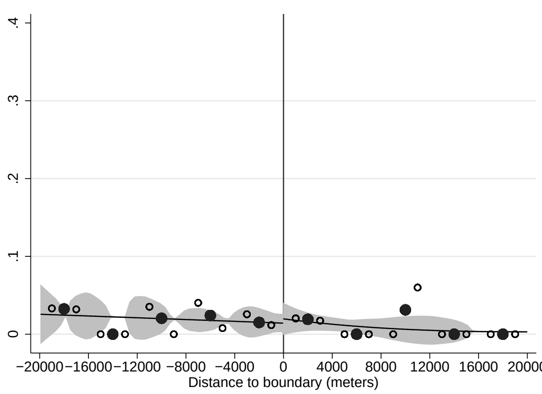
(c) At least one station with Category C fraud



(d) Share of votes under Category C fraud



(e) At least one station with Category C fraud



(f) Share of votes under Category C fraud

Figure B4: Binned Averages for Category C fraud (RD plot)

Notes: South/East region (Panels A and B). North/West region (Panels C and D). Solid dots give the average share of votes classifying in Category C fraud for polling centers falling within 4000-meter distance bins. Hollow dots give the average share of votes classifying in Category C fraud for polling centers falling within 2000-meter distance bins. Refer to section II in the text for a detailed description of Category C fraud. “Distance to boundary” refers to the distance between a polling center and the closest point in the cell phone coverage boundary. “Negative” values of distance give the distance of polling centers/villages in non-coverage areas. The solid line trends give the predicted values from a regression of the outcome variable on a second degree polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 20,000 meters.

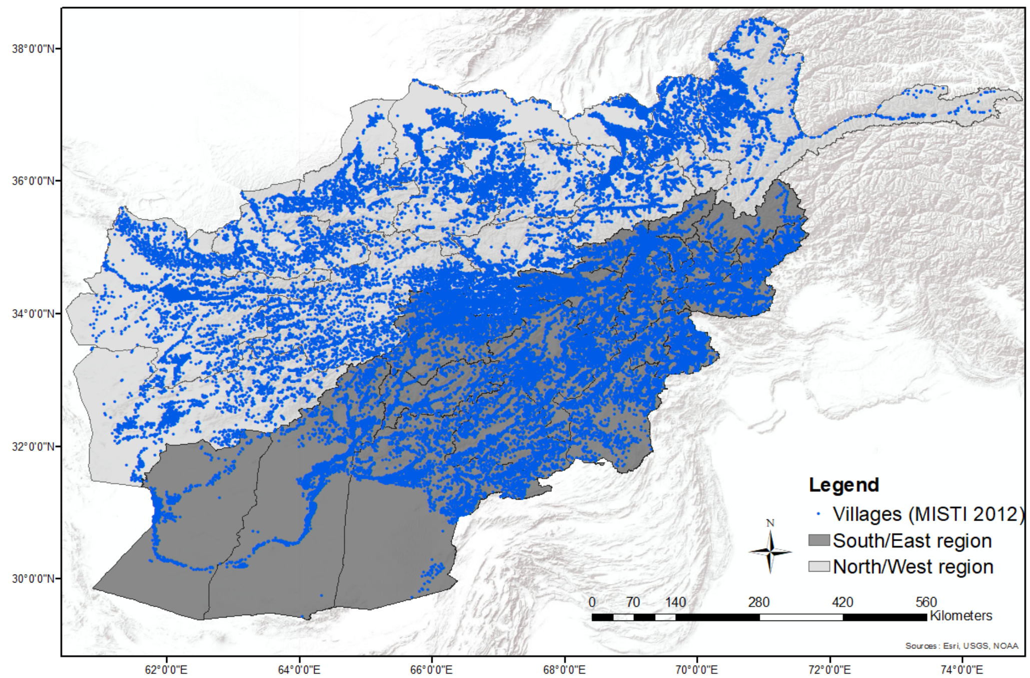
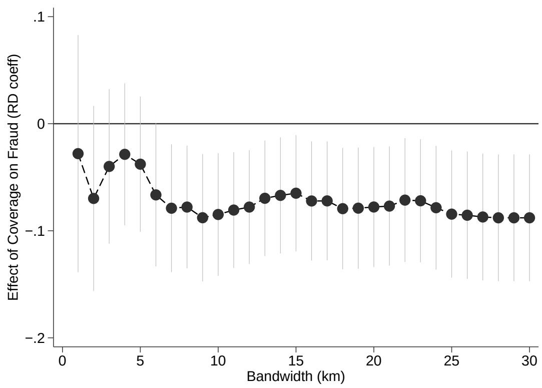
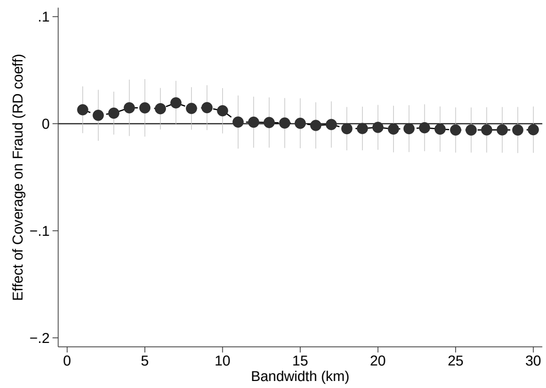


Figure B5: Afghan provinces and regions

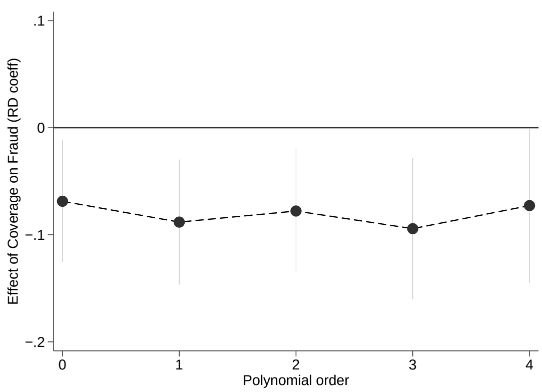
Notes: Regions of Afghanistan. Darker shade indicates the Southeastern provinces. Regions defined using International Security Assistance Forces (ISAF) regional command center definitions. Lines demarcate the provinces of Afghanistan. Blue dots indicate the location of Afghan villages (MISTI, 2013). Map overlaid on USGS topographic map.



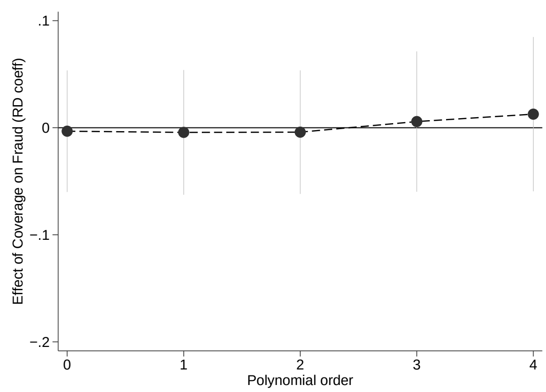
(a) Bandwidth choice (SE region)



(b) Bandwidth choice (NW region)



(c) Distance polynomial (SE region)



(d) Distance polynomial (NW region)

Figure B6: Sensitivity of Results to Bandwidth Choice and Polynomial order (Category C fraud)

Notes: Panels (a) and (b): Each dot indicates the RD estimate using the specified bandwidth. Range spikes indicate 95% confidence intervals of the estimates. Panels (c) and (d): Each dot indicates the RD estimate using the specified order in the RD polynomial. Range spikes indicate 95% confidence intervals of the estimates.

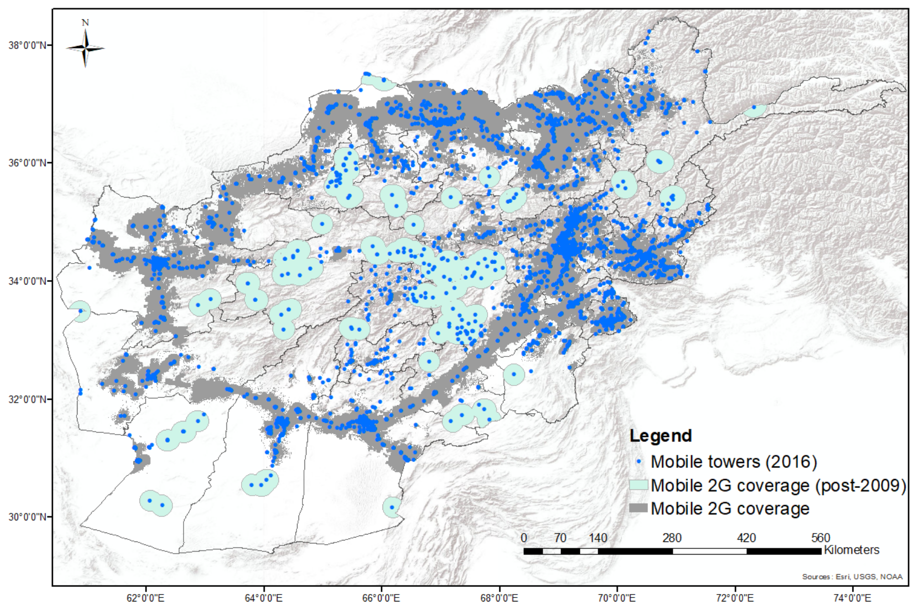
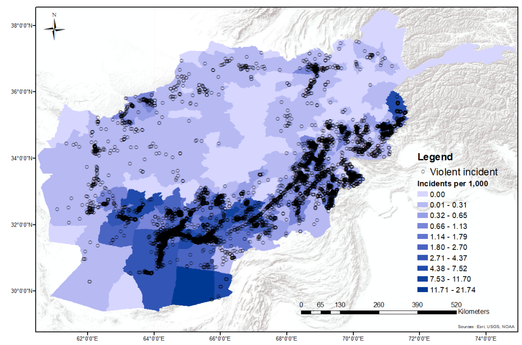
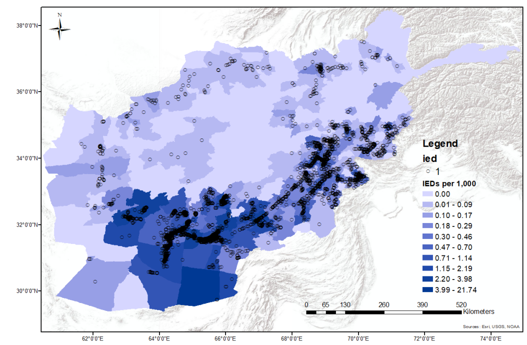


Figure B7: Mobile towers in 2016 but not in 2009

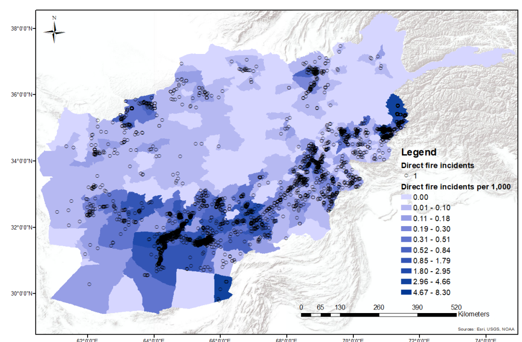
Notes: Blue dots indicate the location of towers in Q1 of 2016. Data obtained from the Afghan Telecommunications Regulatory Authority (ATRA). Gray shaded areas indicate coverage in 2009. Light blue shaded areas indicate coverage areas post 2009 and used in the falsification exercise in section 3.4.1.



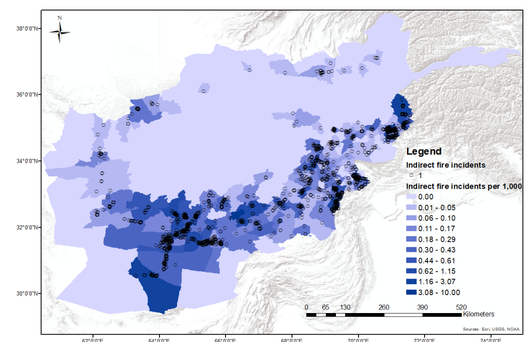
(a) All violent incidents



(b) IEDs



(c) Direct fire



(d) Indirect fire

Figure B8: Insurgent Violence up to Election day (Jan. 1, 2009 - Aug. 20, 2009)

Notes: Dots indicate the location of the violent incident. Color scale of shaded districts gives the rate of violence as the number of incidents per 1,000 inhabitants in the district where the incident took place. Data are significant actions collected by Afghan forces and ISAF. Refer to Condra et al. (2018) for detailed description of the data.

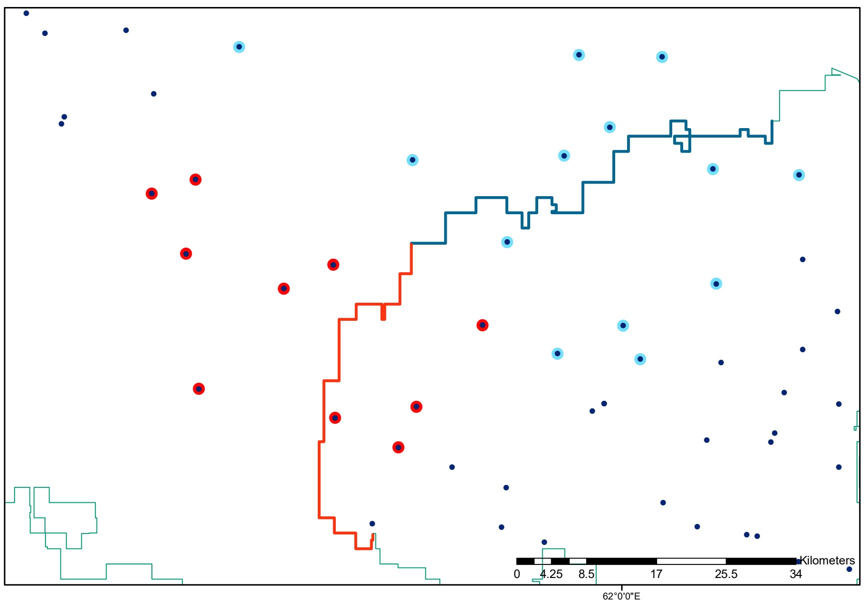


Figure B9: Example of Boundary Segments with Corresponding polling centers

Notes: Example of two boundary segments highlighted in red and blue with polling center (highlighted in corresponding color) belonging to the neighborhood of each segment. Neighborhoods determined by all polling centers closest to a specific segment of the boundary.

Table B1: Sample and Imputations

	Full sample		Within 10 km of boundary		Within 6 km of boundary		Within 4 km of boundary	
	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
Panel A. Unrestricted sample								
Not imputed	5,904	95.84	3,329	95.77	2,388	94.99	1,845	94.81
Imputed based on:								
Settlement	169	2.74	103	2.96	89	3.54	72	3.70
Nearest center	81	1.31	42	1.21	35	1.39	27	1.39
District capital	6	0.10	2	0.06	2	0.08	2	0.10
Total	6,160	100	3,476	100	2,514	100	1,946	100
Panel B. Restricted sample								
Not imputed	2,331	96.04	1,377	95.49	1,106	95.18	912	95.10
Imputed based on:								
Settlement	67	2.76	49	3.40	41	3.53	36	3.75
Nearest center	26	1.07	15	1.04	14	1.20	10	1.04
District capital	3	0.12	1	0.07	1	0.09	1	0.10
Total	2,427	100	1,442	100	1,162	100	959	100

Notes: "Not imputed" refers to centers for which data were available after the merging of 2009 fraud data and 2010 geographic coordinate data. Imputations based on settlement give the polling center the coordinates of the village or settlement center where the polling center is located. Imputations based on nearest center give the polling center the coordinates of the polling center that, within the district, has the closest ID code to it. This is done because the assignment of ID codes followed a spatial order for the most part. Imputations based on district center simply give the polling center the coordinates of the district's capital where the center is located. *Restricted sample* refers to sample where at least one polling center is located on each side of a defined neighborhood. Refer to section 3.3 for the definition of neighborhood. The restricted sample constitutes the main estimation sample.

Table B2: Mean comparison for various polling center characteristics: Southeast Region

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage (1)	No Coverage (2)	S.E. (3)	Coverage (4)	No Coverage (5)	S.E. (6)	RD coeff. (7)	S.E. (8)
Electoral outcomes								
No. of stations	4.26	4.043	(0.26)	4.12	4.048	(0.28)	0.13	(0.276)
No. of expected voters	2325	2152	(145.6)	2255	2146	(157.3)	131.93	(158.9)
Total votes	850.6	878.4	(98.36)	868.8	860.1	(109.2)	24.67	(104.5)
Voter turnout	0.40	0.461	(0.04)*	0.43	0.455	(0.04)	-0.01	(0.039)
Vote share:								
Karzai	0.60	0.608	(0.05)	0.60	0.598	(0.05)	0.02	(0.047)
Abdullah	0.23	0.15	(0.05)*	0.23	0.162	(0.05)	0.05	(0.044)
Polling center characteristics								
Polling center type:								
Mosque	0.24	0.278	(0.04)	0.24	0.285	(0.04)	-0.04	(0.039)
School	0.40	0.281	(0.04)***	0.38	0.27	(0.04)***	0.10	(0.042)**
Other type	0.36	0.44	(0.05)*	0.38	0.444	(0.05)	-0.06	(0.047)
Polling center access (2010):								
Road access	0.62	0.652	(0.06)	0.59	0.644	(0.06)	-0.07	(0.059)
Limited access	0.11	0.096	(0.03)	0.13	0.104	(0.03)	0.04	(0.036)
Other access	0.27	0.252	(0.05)	0.29	0.252	(0.05)	0.03	(0.052)
Share female stations	0.43	0.447	(0.02)	0.43	0.45	(0.02)	-0.02	(0.020)
Share Kuchis stations	0.05	0.0363	(0.01)	0.03	0.0406	(0.01)	0.00	(0.013)
Geographic characteristics								
Elevation (meters)	1789.00	1961	(73.54)**	1812.00	1937	(70.94)*	-142.20	(66.975)**
Slope (percent)	5.67	7.277	(0.76)**	6.30	7.226	(0.77)	-0.90	(0.796)

(Continues)

Table B2: Mean comparison for various polling center characteristics: Southeast Region - *Continued*

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage (1)	No Coverage (2)	S.E. (3)	Coverage (4)	No Coverage (5)	S.E. (6)	RD coeff. (7)	S.E. (8)
Economic development characteristics								
Distance (km) to:								
Primary road (2005)	32.30	45.99	(3.33)***	40.11	44.75	(3.21)	-4.12	(3.022)
Secondary road (2005)	36.03	42.83	(4.15)	43.73	41.37	(4.22)	3.00	(3.880)
District hospital (2005)	25.22	31.5	(2.60)**	28.17	30.99	(2.32)	-2.14	(2.279)
Basic health center (2005)	20.01	22.75	(3.02)	21.90	21.83	(2.81)	1.20	(2.942)
Primary river	16.89	18.15	(1.86)	18.30	18.1	(1.80)	1.11	(1.905)
Seasonal river	6.27	5.988	(1.24)	7.24	5.315	(1.41)	1.48	(1.136)
Distance to seasonal river	15.71	12.63	(2.50)	15.98	12.25	(2.95)	3.09	(2.696)
District/Province capital	0.05	0.0397	(0.02)	0.04	0.0444	(0.02)	-0.01	(0.017)
Demographic characteristics (of closest settlement)								
Population (2012-2013)	1207	1185	(240.0)	1009	1249	(214.3)	-246.5	(207.9)
Language spoken (2012-2013):								
Dari	0.25	0.205	(0.05)	0.23	0.211	(0.05)	0.02	(0.044)
Pashto	0.67	0.685	(0.06)	0.66	0.693	(0.05)	-0.02	(0.052)
Other	0.08	0.109	(0.03)	0.11	0.0963	(0.04)	0.01	(0.034)
Observations								

Notes: Columns (1), (2), (4), and (5) give the means of the corresponding variable. Columns (3) and (6) give the clustered standard errors for the difference in means in parenthesis. Sample restricted to neighborhoods with at least one observation on each side of the boundary. *, **, and *** indicate 10, 5, and 1 percent significance respectively. *No. of expected voters* is the number of voters predicted by the IEC prior to election day. *Total votes* cast is the actual number of votes tallied at the center. *Voter turnout* is defined as the number of votes cast at the center divided by the expected number of voters. *Vote share* is the share of votes received by the each of the two main candidates divided by total votes. I report *Total votes*, *Voter turnout*, and *Vote share* for centers without evidence of fraud. The remaining variables are defined in section II.A of the text. Year values in parenthesis indicate the year the data was collected. Variables without year indication were collected in 2009 or are time-invariant variables. Refer to section II.B for a description of the RD model used in columns (7) and (8).

Table B3: Mean comparison for various polling center characteristics (Conley S.E.)

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage	No Coverage	S.E.	Coverage	No Coverage	S.E.	RD coeff.	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraud outcomes (Category C fraud)								
All regions	0.08	0.11	(0.02)**	0.08	0.12	(0.02)**	-0.04	(0.019)**
			[0.02]**			[0.02]**		[0.020]**
East and South	0.14	0.20	(0.03)**	0.13	0.20	(0.03)**	-0.08	(0.032)**
			[0.03]**			[0.03]**		[0.033]**
North and West	0.01	0.01	(0.01)	0.01	0.01	(0.01)	0.00	(0.010)
			[0.01]			[0.01]		[0.010]
Electoral outcomes								
No. of stations	4.09	3.73	(0.16)**	3.86	3.78	(0.18)	0.09	(0.175)
			[0.16]**			[0.17]		[0.173]
No. of expected voters	2194.00	1944.00	(94.04)***	2069.00	1979.00	(100.00)	94.33	(101.198)
			[92.99]***			[99.35]		[100.161]
Total votes	871.80	866.60	(56.34)	835.00	863.70	(62.41)	-28.22	(59.992)
			[55.72]			[61.72]		[59.076]
Voter turnout	0.43	0.50	(0.02)***	0.45	0.49	(0.02)	-0.03	(0.023)
			[0.02]***			[0.02]		[0.023]
Vote share:								
Karzai	0.50	0.49	(0.03)	0.50	0.49	(0.03)	0.01	(0.030)
			[0.03]			[0.03]		[0.030]
Abdullah	0.34	0.33	(0.03)	0.35	0.33	(0.03)	0.01	(0.031)
			[0.03]			[0.03]		[0.030]

(Continues)

Table B3: Mean comparison for various polling center characteristics (Conley S.E.) - *Continued*

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage (1)	No Coverage (2)	S.E. (3)	Coverage (4)	No Coverage (5)	S.E. (6)	RD coeff. (7)	S.E. (8)
Polling center characteristics								
Polling center type:								
Mosque	0.24	0.26	(0.03)	0.25	0.26	(0.03)	-0.01	(0.029)
			[0.03]			[0.03]		[0.029]
School	0.46	0.37	(0.03)***	0.44	0.37	(0.03)**	0.06	(0.034)*
			[0.03]***			[0.03]**		[0.033]*
Other type	0.30	0.37	(0.03)**	0.30	0.37	(0.03)*	-0.05	(0.032)
			[0.03]**			[0.03]*		[0.032]
Polling center access (2010):								
Road access	0.76	0.78	(0.04)	0.73	0.78	(0.04)	-0.05	(0.039)
			[0.04]			[0.04]		[0.038]
Limited access	0.08	0.07	(0.02)	0.10	0.07	(0.02)	0.03	(0.023)
			[0.02]			[0.02]		[0.022]
Other access	0.16	0.15	(0.03)	0.17	0.15	(0.03)	0.02	(0.032)
			[0.03]			[0.03]		[0.032]
Share female stations	0.44	0.45	(0.01)	0.44	0.45	(0.01)	-0.01	(0.012)
			[0.01]			[0.01]		[0.012]
Share Kuchis stations	0.04	0.03	(0.01)	0.03	0.04	(0.01)	0.00	(0.010)
			[0.01]			[0.01]		[0.010]

(Continues)

Table B3: Mean comparison for various polling center characteristics (Conley S.E.) - *Continued*

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage	No Coverage	S.E.	Coverage	No Coverage	S.E.	RD coeff.	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Geographic characteristics								
Elevation (meters)	1570.00	1782.00	(58.06)***	1617.00	1756.00	(50.71)***	-128.72	(50.617)**
			[57.59]***			[51.89]***		[51.586]**
Slope (percent)	5.72	7.57	(0.53)***	6.46	7.66	(0.58)**	-1.01	(0.602)*
			[0.53]***			[0.58]**		[0.593]*
Economic development characteristics								
Distance (km) to:								
Primary road (2005)	35.30	48.62	(2.43)***	40.72	47.07	(2.31)***	-5.70	(2.290)**
			[2.51]***			[2.41]***		[2.370]**
Secondary road (2005)	44.65	52.60	(3.37)**	50.08	49.84	(3.12)	0.49	(3.056)
			[3.41]**			[3.19]		[3.110]
District hospital (2005)	37.56	45.42	(2.76)***	40.25	42.68	(2.73)	-2.34	(2.739)
			[2.86]***			[2.78]		[2.789]
Basic health center (2005)	20.12	24.31	(1.87)**	21.78	22.68	(1.74)	-0.38	(1.837)
			[1.83]**			[1.75]		[1.836]
Primary river	17.45	18.38	(1.40)	18.00	17.90	(1.29)	0.19	(1.319)
			[1.45]			[1.34]		[1.366]
Secondary river	8.11	8.71	(1.12)	8.96	8.03	(1.10)	0.61	(1.040)
			[1.13]			[1.13]		[1.053]
Seasonal river	12.66	11.05	(1.54)	13.16	10.30	(1.77)	2.70	(1.669)
			[1.62]			[1.83]		[1.715]
District/Province capital	0.05	0.03	(0.01)**	0.04	0.03	(0.01)	0.01	(0.013)
			[0.01]**			[0.01]		[0.013]

(Continues)

Table B3: Mean comparison for various polling center characteristics (Conley S.E.) - *Continued*

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage (1)	No Coverage (2)	S.E. (3)	Coverage (4)	No Coverage (5)	S.E. (6)	RD coeff. (7)	S.E. (8)
Demographic characteristics (of closest settlement)								
Population (2012-2013)	1287.00	957.00	(223.69)	1004.00	1018.00	(145.37)	-51.25	(141.521)
			[222.92]			[146.84]		[142.455]
Language spoken (2012-2013):								
Dari	0.43	0.46	(0.04)	0.43	0.43	(0.04)	0.00	(0.040)
			[0.04]			[0.04]		[0.040]
Pashto	0.44	0.43	(0.04)	0.43	0.45	(0.04)	-0.01	(0.041)
			[0.04]			[0.04]		[0.042]
Other	0.13	0.12	(0.03)	0.14	0.12	(0.03)	0.01	(0.026)
			[0.02]			[0.03]		[0.026]
Observations	891	551		601	456		601	456

Notes: Columns (1), (2), (4), and (5) give the means of the corresponding variable. Columns (3) and (6) give the clustered standard errors for the difference in means in parenthesis and Conley (1999) standard errors in brackets for the difference in means. Estimation of standard errors uses code in Hsiang (2010). Conley (1999) standard errors use a distance cutoff of 50 kilometers and a Bartlett spatial weighting kernel. Sample restricted to neighborhoods with at least one observation on each side of the boundary. *, **, and *** indicate 10, 5, and 1 percent significance respectively. "Coverage" refers to cell phone coverage. *No. of expected voters* is the number of voters predicted by the IEC prior to election day. *Total votes* cast is the actual number of votes tallied at the center. *Voter turnout* is defined as the number of votes cast at the center divided by the expected number of voters. *Vote share* is the share of votes received by the each of the two main candidates divided by total votes. I report *Total votes*, *Voter turnout*, and *Vote share* for centers without evidence of fraud. The remaining variables are defined in the "Data and Variables" section of the text. Year values in parenthesis indicate the year the data was collected. Variables without year indication were collected in 2009 or are time-invariant variables. Refer to section 3.2.2 for a description of the RD model used in columns (7) and (8).

Table B4: Sensitivity of Results to the Addition of Baseline Covariates

	All regions		South and East region		Northwest region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. At least one station with Category C fraud						
Inside coverage	-0.081**	-0.080***	-0.176***	-0.167***	0.025	0.010
	(0.032)	(0.028)	(0.056)	(0.051)	(0.025)	(0.023)
Observations	1074	2039	532	1087	527	952
Mean Outside coverage	0.183	0.141	0.311	0.285	0.0350	0.0402
Bandwidth (km)	7.278	-	6.100	-	7.675	-
Neighborhoods	230	237	95	101	133	137
Panel B. Share of votes under Category C fraud						
Inside coverage	-0.043**	-0.039**	-0.064	-0.080**	0.016	0.006
	(0.020)	(0.020)	(0.039)	(0.035)	(0.013)	(0.014)
Observations	1064	2039	528	1087	503	952
Mean Outside coverage	0.124	0.0933	0.221	0.198	0.0176	0.0193
Bandwidth (km)	7.152	-	5.963	-	7.030	-
Neighborhoods	228	237	95	101	132	137

Notes: Results use equation (3) and add a set of baseline covariates. Refer to section 3.2.1 for a description. Covariates used are: the number of stations, number of female stations, elevation, slope, population, distance to primary and secondary roads, distance to health facility, and distance to a primary river. Optimal bandwidth chosen as in Calonico et al. (2014). Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to section 3.3.2 for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B5: Effect of Mobile Coverage on Category C Fraud, Polling Center-level Clustering

	All regions		Southeast region		Northwest region	
	Optimal Bandwidth (1)	Polynomial in Distance (2)	Optimal Bandwidth (3)	Polynomial in Distance (4)	Optimal Bandwidth (5)	Polynomial in Distance (6)
Panel A. At least one station with Category C fraud						
Inside coverage	-0.077***	-0.082***	-0.160***	-0.171***	0.027	0.011
	(0.029)	(0.024)	(0.051)	(0.042)	(0.023)	(0.022)
Observations	1074	2039	532	1087	527	952
Mean Outside coverage	0.183	0.141	0.311	0.285	0.035	0.040
Bandwidth (km)	7.278	-	6.100	-	7.675	-
Panel B. Share of votes under Category C fraud						
Inside coverage	-0.039*	-0.041**	-0.067*	-0.094***	0.019	0.006
	(0.022)	(0.018)	(0.040)	(0.033)	(0.014)	(0.012)
Observations	1064	2039	528	1087	503	952
Mean Outside coverage	0.124	0.093	0.221	0.198	0.018	0.019
Bandwidth (km)	7.152	-	5.963	-	7.030	-

Notes: Results use equation (1). Refer to section II.B for a description. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2), (4), and (6) use a third degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the polling center level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B6: Effect of Mobile Coverage on Category C Fraud, District Fixed Effects and Clustering

	All regions		Southeast region		Northwest region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. At least one station with Category C fraud						
Inside coverage	-0.042	-0.052	-0.116**	-0.144**	0.029	0.020
	(0.034)	(0.032)	(0.057)	(0.061)	(0.023)	(0.022)
Observations	1083	2039	532	1087	538	952
Mean Outside coverage	0.182	0.141	0.311	0.285	0.035	0.040
Bandwidth (km)	7.373	-	6.139	-	8.101	-
Districts	188	239	87	119	95	120
Panel B. Share of votes under Category C fraud						
Inside coverage	-0.022	-0.027	-0.059*	-0.087**	0.020	0.010
	(0.021)	(0.019)	(0.035)	(0.040)	(0.013)	(0.013)
Observations	1073	2039	529	1087	510	952
Mean Outside coverage	0.125	0.093	0.221	0.198	0.017	0.019
Bandwidth (km)	7.235	-	5.993	-	7.243	-
Districts	187	239	87	119	94	120

Notes: Results use equation (1). Refer to section II.B for a description. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2), (4), and (6) use a third degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use district fixed effects and standard errors clustered at the district level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B7: Effect of Mobile Coverage on Alternative Measures of Fraud

	All regions		Southeast region		Northwest region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. At least one disqualified station within polling center						
Inside coverage	-0.023	-0.012	-0.060**	-0.049*	0.020	0.025
	(0.017)	(0.014)	(0.027)	(0.025)	(0.019)	(0.019)
Observations	1082	2039	545	1087	573	952
Mean Outside coverage	0.045	0.037	0.067	0.061	0.018	0.018
Bandwidth (km)	7.357	-	6.624	-	9.982	-
Panel B. At least one station with more than 600 votes cast						
Inside coverage	-0.062*	-0.050	-0.096	-0.106*	0.011	-0.025
	(0.036)	(0.036)	(0.063)	(0.063)	(0.037)	(0.031)
Observations	1081	2039	484	1087	542	952
Mean Outside coverage	0.255	0.234	0.399	0.363	0.093	0.143
Bandwidth (km)	7.340	-	4.803	-	8.404	-

Notes: Results use equation (1). Refer to section II.B for a description. Panel title indicated outcome used. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2), (4), and (6) use a third degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the polling center level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B8: Summary Statistics for Violence Data

	All regions combined		Southeast region		Northwest region	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Panel A: Likelihood of at least one incident within 1km of polling center						
Attack	0.213	0.41	0.316	0.46	0.068	0.25
IED	0.148	0.35	0.218	0.41	0.045	0.21
Direct Fire	0.103	0.30	0.128	0.33	0.030	0.17
Indirect Fire	0.065	0.25	0.100	0.30	0.011	0.10
Panel B: Number of incidents within 1km of polling center						
Attack	1.000	4.35	1.235	4.42	0.151	0.81
IED	0.418	1.63	0.493	1.47	0.081	0.44
Direct Fire	0.362	2.52	0.384	2.66	0.045	0.29
Indirect Fire	0.220	1.65	0.358	2.28	0.025	0.28
Observations	6,160		2,642		1,913	

Notes: Data are significant actions (SIGACTs) collected by Afghan forces and ISAF. Refer to Condra et al. (2018) for detailed description of the data. Insurgent attacks data or significant actions (SIGACTs) are obtained from time-stamped and georeferenced records collected by the International Security Assistance Forces (ISAF) and Afghan forces in Afghanistan. The three main outcomes: the total number of attacks (Attack), the number of IEDs (IED), and direct fire attacks (Direct Fire) within a 1-kilometer radius of polling centers and from January 1, 2009 to August 20, 2009. This encompasses all violence occurring on the election year and up to election day. Direct fire refers to attacks using small arms and rocket-propelled grenades.

Table B9: Summary Statistics for Complaints Data

			Coverage		No coverage	
	Mean (1)	S.D. (2)	Mean (3)	S.D. (4)	Mean (5)	S.D. (6)
Number of complaints	0.54	1.30	0.62	1.37	0.33	1.03
Share of complaints submitted by individuals	0.91	0.26	0.91	0.26	0.91	0.25
Share of complaints submitted by females	0.13	0.28	0.13	0.29	0.08	0.23
Observations	2785		2071		714	

Notes: Individuals refers to citizens, candidates, or candidates' representatives and excludes government officials (polling officials, IEC or ECC officials, and other government organizations). Refer to section 4.3 for more details.

Table B10: Tribes and Tribal Confederations of Southeast Afghanistan

	Number	Share of Total	Share of ethnic group	Share of confederacy
Panel A: Pashtun Confederations and Tribes				
Durrani confederation	4,343	0.24	0.37	1.00
Durrani Pashtun	1,608	0.09	0.14	0.37
Barakzai Pashtun	902	0.05	0.08	0.21
Noorzai Pashtun	502	0.03	0.04	0.12
Panjpai Durrani Pashtun	401	0.02	0.03	0.09
Alizai Pashtun	328	0.02	0.03	0.08
Mixed Durrani	208	0.01	0.02	0.05
Popalzai Pashtun	200	0.01	0.02	0.05
Ashakzai Pashtun	194	0.01	0.02	0.04
Ghilzai confederation	4,469	0.25	0.38	1.00
Ibrahim Ghilzai	3,325	0.18	0.28	0.74
Turan Pashtun	874	0.05	0.07	0.20
Baezai Mohamand Powindah	110	0.01	0.01	0.02
Kudi Lodi Powindah	82	0.00	0.01	0.02
Mian Khel Powindah	39	0.00	0.00	0.01
Other Ghilzai	16	0.00	0.00	0.00
Miani Powindah	14	0.00	0.00	0.00
Kukozai Mohamand Powindah	9	0.00	0.00	0.00
Other Pashtun	3,052	0.17	0.26	1.00
Safi Pashtun	1,261	0.07	0.11	0.41
Shinwari Pashtun	509	0.03	0.04	0.17
Pashtun	279	0.02	0.02	0.09
Wardak Pashtun	235	0.01	0.02	0.08
Salarzai	113	0.01	0.01	0.04
Kom	107	0.01	0.01	0.04
Jadran Pashtun	87	0.00	0.01	0.03
Other	461	0.02	0.04	0.14
All Pashtuns	11,864	0.66	1.00	.

(Continues)

Table B10: Tribes and Tribal Confederations of Southeast Afghanistan-*Continued*

	Number	Share of Total	Share of ethnic group	Share of confederacy
Panel B: Hazara Tribes				
Besud Hazara	1,226	0.07	0.29	.
Other Hazara	896	0.05	0.21	.
Dai Chopan Hazara	615	0.03	0.14	.
Dai Khitai Hazara	309	0.02	0.07	.
Muhammad Kwaja Hazara	222	0.01	0.05	.
Khatai Hazara	215	0.01	0.05	.
Jaghuri Hazara	187	0.01	0.04	.
Dai Kundi Hazara	154	0.01	0.04	.
Chahar Dasta Hazara	134	0.01	0.03	.
Polada Hazara	134	0.01	0.03	.
Jaghatus Hazara	94	0.01	0.02	.
Faoladi Hazara	55	0.00	0.01	.
Dai Zangi Hazara	21	0.00	0.00	.
Uruzgani Hazara	13	0.00	0.00	.
All Hazara	4,275	0.24	1.00	.
Panel C: Other Ethnic Groups				
Tajik	1,117	0.06	1.00	.
Uzbek	174	0.01	1.00	.
Baluch	157	0.01	1.00	.
Mixed	163	0.01	1.00	.
Other (non-Pashtun)	246	0.01	1.00	.

Notes: Tribal confederations created using definitions based on Tribal Hierarchy and Dictionary of Afghanistan (2007). Tribal maps obtained via the Culture and Conflict Studies.

Table B11: Effect of Mobile Coverage on Number of Insurgent Attacks near Polling Centers: Excluding Election Day

	All regions		Southeast region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)
Panel A. Number of violent incidents within 1km radius of poling center				
Inside coverage	0.575	0.724	0.517	1.064
	(0.599)	(0.478)	(1.029)	(0.883)
Observations	1278	2118	682	1163
Mean Outside coverage	0.81	0.66	1.49	1.50
Bandwidth (km)	10.2	-	9.53	-
Neighborhoods	236	239	101	102
Panel B. Number of IEDs within 1km radius of polling center				
Inside coverage	0.228	0.178	0.367	0.276
	(0.196)	(0.161)	(0.313)	(0.299)
Observations	866	2118	528	1163
Mean Outside coverage	0.14	0.076	0.23	0.17
Bandwidth (km)	4.20	-	4.86	-
Neighborhoods	222	239	96	102
Panel C. Number of direct fire incidents within 1km radius of polling center				
Inside coverage	0.345	0.412	0.368	0.590
	(0.313)	(0.268)	(0.549)	(0.494)
Observations	1473	2118	817	1163
Mean Outside coverage	0.43	0.36	0.83	0.82
Bandwidth (km)	13.6	-	13.4	-
Neighborhoods	238	239	102	102

Notes: Results use equation (1) and outcome variables denoted in panel names. Refer to section III.B for more details on the outcome variables used. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2) and (4) use a second degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B12: Effect of Mobile Coverage on Number of Insurgent Attacks near Polling Centers: Entire time period (January 1, 2006-August 20, 2009 (Election Day))

	All regions		Southeast region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)
Panel A. Number of violent incidents within 1km radius of poling center				
Inside coverage	3.565	3.482	5.024	5.607
	(3.171)	(2.784)	(5.492)	(5.240)
Observations	1339	2118	712	1163
Mean Outside coverage	2.46	1.93	4.59	4.35
Bandwidth (km)	11.2	-	10.6	-
Neighborhoods	236	239	101	102
Panel B. Number of IEDs within 1km radius of polling center				
Inside coverage	0.721	0.326	1.056	0.402
	(0.522)	(0.485)	(0.877)	(0.899)
Observations	900	2118	540	1163
Mean Outside coverage	0.62	0.36	1.05	0.81
Bandwidth (km)	4.55	-	5.14	-
Neighborhoods	222	239	97	102
Panel C. Number of direct fire incidents within 1km radius of polling center				
Inside coverage	2.161	2.127	3.008	3.400
	(1.669)	(1.534)	(2.998)	(2.878)
Observations	1533	2118	774	1163
Mean Outside coverage	1.11	0.92	2.18	2.06
Bandwidth (km)	14.5	-	12.5	-
Neighborhoods	238	239	102	102

Notes: Results use equation (1) and outcome variables denoted in panel names. Refer to section III.B for more details on the outcome variables used. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2) and (4) use a second degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B13: Effect of Mobile Coverage on Number of Insurgent Attacks near Polling Centers: 5KM Radius

	All regions		Southeast region	
	Optimal Bandwidth	Polynomial in Distance	Optimal Bandwidth	Polynomial in Distance
	(1)	(2)	(3)	(4)
Panel A. Number of violent incidents within 5km radius of poling center				
Inside coverage	1.914	2.426	3.193	5.287
	(2.696)	(2.228)	(4.635)	(4.150)
Observations	1284	2118	756	1163
Mean Outside coverage	7.99	6.09	14.6	13.6
Bandwidth (km)	10.3	-	11.8	-
Neighborhoods	236	239	101	102
Panel B. Number of IEDs within 5km radius of polling center				
Inside coverage	1.796	0.719	3.197*	1.996
	(1.108)	(1.099)	(1.860)	(2.004)
Observations	844	2118	519	1163
Mean Outside coverage	1.90	1.12	3.23	2.53
Bandwidth (km)	3.98	-	4.69	-
Neighborhoods	222	239	96	102
Panel C. Number of direct fire incidents within 5km radius of polling center				
Inside coverage	0.322	1.343	0.172	2.186
	(1.340)	(0.930)	(2.510)	(1.847)
Observations	1433	2118	860	1163
Mean Outside coverage	4.49	3.57	8.39	7.96
Bandwidth (km)	13.1	-	14.3	-
Neighborhoods	238	239	102	102

Notes: Results use equation (1) and outcome variables denoted in panel names. Refer to section III.B for more details on the outcome variables used. Optimal bandwidth chosen as in Calonico et al. (2014). Columns (2) and (4) use a second degree polynomial in distance to boundary. Polynomial order determined using Akaike's criterion as suggested in Black et al. (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to section II.C for a description of how boundary neighborhoods are created. ***, **, * indicate 10, 5, and 1 percent significance, respectively.

Table B14: Summary of News Articles Addressing Tower Shutdowns

	Full Sample (Feb. 2008-Dec. 2019)		Up to 2009 Election (Feb. 2008-Dec. 2009)	
	Number	Percent	Number	percent
Panel A: References to Timing of Shutdowns				
Night	29	93.5	10	90.9
Day	1	3.2	1	9.1
Complete shutdown	1	3.2	0	0.0
Total articles (timing reference)	31	100	11	100
Timing not specified	10	-	0	-
Panel B: References to Location of Shutdowns				
Helmand	9	26.5	1	14.3
Kandahar	3	8.8	1	14.3
Ghazni	5	14.7	1	14.3
Southern Afghanistan	8	23.5	4	57.1
Other	9	26.5	0	0.0
Total articles (location reference)	34	100	7	100
Location not specified	7	-	4	-
Total articles	41		11	

Notes: Night/day refers to whether the article specified if the shutdowns were during the night or during the day. References to locations refers to whether the article mentioned the location of the shutdowns. Total is the number of articles. Shares calculated out of the total articles making reference to timing or location of the shutdowns. Other provinces referenced in the location of shutdowns where: Balkh, Farah, Zabul, Paktika, Ghor, Wardak, Herat, Kunduz,Baghlan, Nangharar, Takhar, Khost, Paktya, Logar, and Kunar.

Appendix C Model Extensions

C.1 The Transmitter's Problem

Consider a polling center serving n voters. Furthermore assume that there exists a nationwide phone hotline to report electoral fraud. Given the widespread use of cell phones in the developing world and Afghanistan in particular, suppose individual i uses a cell phone if he decides to report fraud. Reporting fraud carries a physical cost $c(D)$ where D indicates the accessibility of the medium (cell phones) used to report fraud. In the context of this study D is an indicator for whether the polling center is located in an area with cell phone coverage.[2] Specifically, let the cost of reporting fraud equal $\underline{c}$ if the center is on an area with cell phone coverage (i.e., $D = 1$) and $\bar{c}$ otherwise with $\bar{c} > \underline{c}$.[3]

Furthermore, assume that reporting fraud gives i a utility gain λ_i that can be interpreted as a "warm glow" parameter or i's satisfaction from his pro-social behavior. The individual's net payoff from reporting fraud is therefore given by $\lambda_i - c(D)$. He will then decide to report fraud if:

$$\lambda_i \geq c(D) \tag{1}$$

Assuming λ_i is distributed among voters at the center with probability function $G(\lambda)$ then the probability of an individual making a report given coverage status D is given by $\rho(D) = 1 - G(D)$.[4]

C.2 Nonlinear Reporting Cost Function

Since reporting fraud via cell phones increases in cost as polling centers are farther away from coverage areas, I consider the case where this function is nonlinear. Specifically, assume reporting fraud carries a physical cost $c(d)$ where d defines how accessible the medium (cell phones) used to report fraud is. In the spatial context of this study d can be interpreted as the shortest distance from the polling center where fraud takes place to a geographic boundary that defines whether

[2] In reality D should indicate whether there is coverage in the area where the individual decides to report fraud (the polling center, his house, etc.). This information is unavailable hence I only consider coverage at the polling center. This implicitly assumes that the call to report fraud is made right at the center. However, since centers were located within settlements, it is likely that any calls are made within the "catchment area" of the center so that using the center as a reference in determining coverage should not greatly affect the analysis of the model.

[3] Without loss of generality, I assume that the reporting cost on the non-coverage side $\bar{c}$ is constant, however, this cost might increase as polling center are further away from the coverage boundary. Refer to Appendix C for a discussion of an alternative specification of the reporting cost function that uses a smooth, non-linear function on the non-coverage side.

[4] $G(.)$ is actually a function of $c(D)$, which in turn, is a function of D. Refer to Appendix C for an extension of Equation 1 that considers the possibility of free-riding when reporting fraud.

there is cell phone coverage. Using this interpretation and assuming that $d < 0$ (i.e., a negative distance) arbitrarily defines a polling center located in the non-coverage side of the boundary one can specify a function $D = \mathbb{1}\{d \geq 0\}$ that indicates coverage such that the cost of reporting fraud is given by:

$$c(d) = D\underline{c} + (1 - D)h(d) \tag{2}$$

where $\underline{c}$ is the marginal cost of making a call when the center is in the coverage side. I assume this cost is equal for everyone in the coverage side. $h(d)$ is a smooth cost function faced by individuals on the non-coverage side with $h(d) > \underline{c}$ for all $d < 0$ and $h(0) = \bar{c}$ with $\bar{c} > \underline{c}$.[5]

C.3 Free-riding and Fraud Reporting

I extend the *transmitter's problem* presented in section C.1 by allowing for the possibility of "free-riding" in the reporting process. Free-riding can be a concern in this context if individuals assess that the probability that a fraudulent official is punished conditional on his report is trivial. If that is the case, then the probability of making a report does not change regardless of accessibility to the reporting medium, in this context, coverage status. I show that individuals have an incentive to report fraud, despite the free-riding problem, as long as there is some utility gain from the reporting process itself (i.e., the warm glow parameter specified in section C.1)

More specifically, assume that since reporting fraud is costly, individual i assesses the likelihood that the center is actually audited (and hence the fraudulent candidate is penalized) as a result of his report.[6] More specifically, let ϕ_1 and ϕ_0, with $\phi_1 \geq \phi_0$ denote i's subjective assessment of the probability that the candidate will be punished given that he reports him and does not report him respectively. When the candidate is punished i gets a net utility value ξ_i that can be interpreted as a utility gain from the fraudulent candidate being punished net of any affinity or benefits that the individual might receive from non-punishment.[7] Additionally, assume as in section C.1, that the act of reporting fraud gives i a utility gain λ_i.

The individual's net payoffs from reporting fraud are therefore given by $\xi_i + \lambda_i - c(d)$ when the fraudulent candidate is punished and $\lambda_i - c(d)$ otherwise. Lastly, if i decides *not* to report, he simply obtains ξ_i when the candidate is punished and zero otherwise. Assuming linear utility, he

[5] A possible parameterization for $c(d)$ could be $c(d) = \bar{c} \cdot \exp(-\beta d)$

[6] I assume that once a fraudulent center is audited, the candidate and polling center manager are penalized. Therefore I do not consider any "concealment technology" as in Cremer and Gahvari (1994)

[7] The idea is that the individual might obtain some "justice has been served" satisfaction while at the same time punishment to a candidate of his liking might bring some disutility.

will then decide to report fraud if:

$$\phi_1[\xi_i + \lambda_i - c(d)] + (1-\phi_1)[\lambda_i - c(d)] \geq \phi_0 \xi_i$$
$$\lambda_i + (\phi_1 - \phi_0)\xi_i \geq c(d) \tag{3}$$

Notice that even when there is a "free-riding" problem (i.e., i believes that his report does not affect the probability that the candidate is punished ($\phi_1 = \phi_0$) and hence the decision rule above reduces to: Report if $\lambda_i \geq c(d)$) an individual i might still have an incentive to report fraud as he derives utility from doing this alone. Therefore "free-riding" will lower the willingness to report fraud but not eliminate it completely.

C.4 The Voter's Problem

An individual i considers a campaign promise and the possibility of a violent outcome on election day when deciding whether to vote. In the spirit of Dekel et al. (2008), assume each voter is characterized by parameter a_i representing the net utility the individual obtains from simply tendering his vote to the candidate. This can be interpreted as i's affinity towards the candidate. The candidate offers individual i a campaign promise $p_{l,i}$ regardless of the election result but conditional on i tendering the vote to him. Assuming the individual is an expected income maximizer then i's expected payoff from voting is given by $V_i^1 = a_i + p_{l,i}$ while i's expected payoff from not voting is simply $V_i^0 = 0$. Since elections in conflict zones are often characterized by violence, the individual takes into account an exogenous probability δ_j that a violent event takes place at polling center j and as a result receives a negative payoff P. This consideration is particularly important in the Afghan context as the Taliban issued several warnings targeting polling centers and voters on election day (Gall (2009), Filkins (2009)). The individual therefore decides to vote if $\delta(V_i^1 - P) + (1-\delta)V_i^1 \geq V_i^0$.[8]

Given i's payoffs, the minimum price per legal vote (i.e., the campaign promise) that guarantees i's vote is therefore given by:

$$p_{l,i} = \delta_j P - a_i \tag{4}$$

where $p_{l,i} = 0$ if the affinity parameter a_i is sufficiently large as to offset the negative payoff of violence (i.e., $a_i > \delta_j P$).[9]

[8] For simplicity, this specification of the model considers only the problem of whether to vote but not the problem of for whom to vote. Refer to Appendix C for an extension of the "Voter's Problem" that considers two candidates and hence the decision becomes whether to vote and for whom to vote.

[9] Alternatively, one can specify the reservation campaign promise $p_{l,i}$ as equal to the $max\,\{0, \delta_j P - a_i\}$

C.5 Two Candidates

This section introduces the possibility of two candidates in the *voter's problem.* A key distinction from the model presented in section C.4 is that the voter not only has to decide whether to vote but also for whom to vote taking into account each candidate's campaign promise. In terms of the *candidate's problem*, the level of fraud in equilibrium changes slightly when considering a second candidate. In essence, this introduces an additional channel of fraud, namely what I refer to as a *challenger effect.* Broadly speaking, in order for a candidate to entice voters to vote for him, he has to pay a legal price for their vote that matches the highest value between the expected net payoff from violence and the opposing candidates' campaign offer to voters.

The Voter's Problem Suppose voters have to decide between two candidates indexed by k. As in Dekel et al. (2008), assume that each voter i is characterized by parameters U_i^k and a_i^k that represent the utility the individual obtains from k's victory and from simply tendering his vote to k, respectively. Candidate k offers individual i a campaign promise $p_{l,i}^k$ regardless of the election result but conditional on i tendering his vote to k[10]. Letting $\psi_{k|l}$ denote the probability that candidate k wins given that i tendered his vote to candidate $l \in \{X, Y\}$ and letting the individual be an expected income maximizer then i's expected payoff from voting for k is given by:

$$V_i^k = \psi_{k|k}(U_i^k + a_i^k + p_{l,i}^k) + \psi_{\sim k|k}(U_i^{\sim k} + a_i^k + p_{l,i}^k) \tag{5}$$

where $\sim k$ denotes "not" k. Similarly, letting $\psi_{k|0}$ denote the probability that candidate k wins given that i did not vote, then i's expected payoff from not voting is simply:

$$V_i^0 = \psi_{k|0}U_i^k + \psi_{\sim k|0}U_i^{\sim k} \tag{6}$$

Since elections in conflict zones are often characterized by violence, the individual takes into account an exogenous probability δ_j that a violent event takes place at the polling center and as a result receives a very negative payoff P. The individual therefore decides to vote if:

$$\delta(V_i^k - P) + (1 - \delta)V_i^k \geq V_i^0 \tag{7}$$

[10]This campaign promise can be interpreted in either of two ways: Voting is not secret so that candidate k knows which individuals voted for him and hence he pays them the campaign promise $p_{l,i}^k$, or voting is secret but once a voter commits *a priori* to tender the vote to k he does not change his vote the day of the election.

Given expressions (5), (6) and (7), the individual will decide to vote and vote for k if the following two conditions hold.

$$p_{l,i}^{k} + a_i^k + \psi_{k|k} U_i^k + \psi_{\sim k|k} U^{\sim k} \geq \delta P + \psi_{k|0} U_i^k + \psi_{\sim k|0} U^{\sim k} \quad (8)$$

$$p_{l,i}^{k} + a_i^k + \psi_{k|k} U_i^k + \psi_{\sim k|k} U^{\sim k} \geq p_{l,i}^{\sim k} + \psi_{k|\sim k} U_i^k + \psi_{\sim k|\sim k} U^{\sim k} \quad (9)$$

Assuming, for simplicity, that the individual believes that his vote is "non-pivotal"[11], then the two expressions above simply reduce to:

$$p_{l,i}^{k} + a_i^k \geq \delta P \quad (10)$$

$$p_{l,i}^{k} + a_i^k \geq p_{l,i}^{\sim k} + a_i^{\sim k} \quad (11)$$

The minimum price per legal vote (i.e., the campaign promise) that candidate k must pay is therefore given by:

$$p_{l,i}^{k} = max \left\{ 0, \delta P - a_i^k, p_{l,i}^{\sim k} + a_i^{\sim k} - a_i^k \right\} \quad (12)$$

The Candidate's Problem Candidate k must decide how many votes (both legal and fraudulent) to buy from each center j. Assume that the auditing agency can differentiate between fraudulent and legal votes so that, once audited, any fraudulent votes are dropped and the candidate only receives legal votes v_l^k. In case where the center is not audited the candidate simply keeps all votes $v_l^k + v_f^k$. Given that the assessed probability of an audit is given by π and assuming that the candidate has quasilinear preferences over votes, then the maximization problem of the candidate is given by:

$$\max_{v_l^k, v_f^k} \quad \pi v_l^k + (1-\pi)[v_l^k + (v_f^k)^\alpha]$$

$$\text{subject to} \quad p_f v_f^k + \sum_{i=1}^{v_l^k} p_{l,i}^k \leq E^k$$

where fraudulent votes enter non-linearly (with $\alpha \leq 1$) to capture the possibility that fraudulent and legal votes are not perfect substitutes as specified in Callen and Weidmann (2013) and E^k is some campaign endowment of candidate k. To simplify the analysis, assume that the candidate does not observe the affinity parameters a_i^k but knows their distribution among the voters in center j so

[11] This assumption simply states that the individual believes that tendering the vote to k will not alter the probability that k wins. More specifically, $\psi_{k|k} = \psi_{k|\sim k} = \psi_{k|0}$ for $k \in \{X, Y\}$.

that $\sum_{i=1}^{v_l^k} p_{l,i}^k$ is simply given by $p_l^k v_l^k$ where p_l^k uses the expected value of these affinity parameters. The solution to the problem above provides an optimal relationship between fraudulent votes and their price p_f that is given by:

$$v_f^k = \left[\frac{\alpha(1-\pi)p_l^k}{p_f}\right]^{\frac{1}{1-\alpha}} \tag{13}$$

Substituting the expressions for prices p_f and p_l^k in order to obtain the equilibrium level of fraud gives:

$$v_f^k = \left[\frac{\alpha(1-\pi)\cdot max\left\{0, \delta P - a^k, p_l^{\sim k} + a^{\sim k} - a^k\right\}}{\pi F}\right]^{\frac{1}{1-\alpha}} \tag{14}$$

From this expression it is clear that the probability π that the center is audited (i.e., the level of social monitoring) decreases the equilibrium fraud level in the center, however, notice also that the "social monitoring effect" is one among others that explain fraud. To see this more clearly, I rewrite expression (14) by separating the different components of fraud:

$$v_f^k = \left[\alpha \cdot \underbrace{\frac{1-\pi}{\pi}}_{\substack{\text{Social}\\\text{monitoring}\\\text{effect}}} \cdot \underbrace{\frac{1}{F}}_{\substack{\text{Fine}\\\text{effect}}} \cdot max\left\{0, \underbrace{\delta P - a^k}_{\substack{\text{Violence}\\\text{effect}}}, \underbrace{p_l^{\sim k} + a^{\sim k} - a^k}_{\substack{\text{Challenger}\\\text{effect}}}\right\}\right]^{\frac{1}{1-\alpha}} \tag{15}$$

References

Black, D. A., J. Galdo, and J. A. Smith (2007). Evaluating the regression discontinuity design using experimental data. *Working Paper, University of Michigan*.

Callen, M. and N. B. Weidmann (2013). Violence and election fraud: Evidence from afghanistan. *British Journal of Political Science*, 53–75.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal 18*(1), 234–261.

Cattaneo, M. D., M. Jansson, and X. Ma (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1–7.

Condra, L. N., J. D. Long, A. C. Shaver, and A. L. Wright (2018). The logic of insurgent electoral violence. *American Economic Review 108*(11), 3199–3231.

Conley, T. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*, 1–45.

Cremer, H. and F. Gahvari (1994). Tax evasion, concealment and the optimal linear income tax. *Scandinavian Journal of Economics*.

Dekel, E., M. O. Jackson, and A. Wolinsky (2008). Vote buying: General elections. *Journal of Political Economy 116*(2).

Electoral Complaints Commission (2010). Final report. 2009 presidential and provincial council elections. Kabul.

Filkins, D. (2009). Threats by taliban may sway vote in afghanistan. *The New York Times*.

Gall, C. (2009). Violence roils afghanistan days before election. *The New York Times*.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the caribbean and central america. *Proceedings of the National Academy of sciences 107*(35), 15367–15372.

MISTI (2013). Measuring impacts of stabilization initiatives (misti). (Available at: http://usaidmisti.com/gis-data).

Tribal Hierarchy and Dictionary of Afghanistan (2007). *Tribal hierarchy and dictionary of Afghanistan: a reference aid for analysts*. Courage Services Inc.