American Economic Journal: Applied Economics 2021, 13(2): 1–51
https://doi.org/10.1257/app.20190443

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

By Robert M. Gonzalez*

This paper examines the impact of cell phone access on election fraud. I combine cell phone coverage maps with the location of polling centers during the 2009 Afghan presidential election to pinpoint which centers were exposed to coverage. Results from a spatial regression discontinuity design along the two-dimensional coverage boundary suggest that coverage deters corrupt behavior. Polling centers just inside coverage report a drop in the share of fraudulent votes of 4 percentage points, while the likelihood of a fraudulent station decreases by 8 percentage points. Analyses of the effect of coverage on citizen participation in election monitoring, election-related insurgent violence, and the tribal composition of villages suggest that the observed declines in fraud are likely attributed to cell phone access strengthening social monitoring capacity. (*JEL* D72, K16, K42, O17, Z13)

Developing countries have experienced a dramatic increase in mobile phone connectivity rates over the past decade. A growing literature has shown that increased access to cell phones can improve market efficiency (Jensen 2007, Aker 2010), literacy (Aker, Ksoll, and Lybbert 2012; Aker and Ksoll 2020), access to mobile banking (Jack, Suri, and Townsend 2010; Jack and Suri 2011), and risk sharing (Jack and Suri 2014; Blumenstock, Eagle, and Fafchamps 2016), among other things. Thus far, however, our understanding of whether this rapid expansion in technology can be effective in combating corruption and improving transparency has not moved at the same pace. In theory, mobile phone access can deter corrupt behavior by facilitating collective action and improving the transfer of information among citizens. It can also give rise to a host of interventions that rely on cell phones as mechanisms to monitor and report corrupt behavior.[1] On the other hand, it can

*Department of Economics, Darla Moore School of Business, University of South Carolina, Columbia, SC 29208 (email: robert.gonzalez@moore.sc.edu). Benjamin Olken was coeditor for this article. I thank Patrick Conway, Erica Field, David Guilkey, Klara Peter, and Tiago Pires for their comments and guidance. I am equally grateful to Michael Callen, Tarek Ghani, Monica Garcia-Perez, Donna Gilleskie, Clément Imbert, Sarah Komisarow, Maureen Pirog, Forrest Spence, Helen Tauchen, and participants at the UNC Applied Micro workshop, IPHD workshop at Duke University, NEUDC 2015, and the WEAI Workshop. I'm indebted to Andrew Shaver and Austin Wright for providing access to data. I acknowledge funding provided by the UNC-Chapel Hill Library, and the AEA Mentoring Program.

[†]Go to https://doi.org/10.1257/app.20190443 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] Anticorruption or election fraud hotlines are important examples. For a detailed description of the history and growth of information and communication technologies (ICT)–based monitoring, refer to Schuler (2008).

be detrimental if it also facilitates coordination among corrupt officials or among violent nonstate actors in countries with a fragile security environment.

This paper provides evidence that widespread access to cell phones can indeed deter corrupt behavior. Specifically, it investigates the impact of cell phone access on election fraud in the context of the 2009 Afghan presidential election by exploiting geographic variations in the exposure of polling centers to cell phone coverage. The empirical analysis employs a novel spatial regression discontinuity (RD) design that compares fraud levels for polling centers within a close distance of the two-dimensional boundary formed between coverage and noncoverage areas. The analysis uses several novel data sources, including detailed coverage maps based on the location of cell phone towers for the two largest mobile service providers in Afghanistan, data on the precise location of polling centers collected by International Security Assistant Force (ISAF) inspection teams shortly after the election, and polling-center-level data on various measures of election fraud collected by a UN-sponsored audit shortly after the election.

The results provide compelling evidence that cell phone coverage reduces fraud. For polling centers within a six to seven kilometers (km) bandwidth around the coverage boundary, the share of fraudulent votes in centers inside coverage areas drops by about 4 percentage points, while the likelihood of a fraudulent station goes down by about 8 percentage points. The results are robust to several choices of bandwidth, polynomial order, and a novel design that accounts for selection into coverage. Careful analysis of a rich set of electoral, geographic, socioeconomic, and demographic indicators of polling centers and villages indicates a smooth transition across the coverage boundary and thus little evidence that changes in fraud are explained by changes in these indicators. Additional results modifying a boundary RD design that recovers a distribution of treatment effects along the two-dimensional coverage boundary suggest a high degree of spatial heterogeneity. Significant drops in fraud are mostly restricted to segments of the boundary located in the eastern and southern regions of the country. Lastly, we find no evidence of fraud displacing into noncoverage areas or of coverage leading to positive spillovers into polling centers near the coverage boundary.[2]

Fraud may respond to cell phone coverage through several mechanisms. I combine contextual evidence with the findings from an illustrative theoretical model to distinguish three key channels: citizen participation via social monitoring, election-related political violence, and candidate-voter affinity along ethnic or tribal lines.

First, I explore whether cell phone access can foster citizen participation and strengthen social monitoring capacity. Specifically, I investigate the impact of coverage on election monitoring in the context of a UN-led initiative that created hotlines and other channels to facilitate election monitoring and reporting during the 2009 Afghan election. This constituted the first time in Afghan history that a formal channel for individuals to address electoral complaints was created. The nature of the monitoring technology indicates that drops in fraud can potentially be explained

[2] A positive coverage spillover can take place if, for instance, polling centers near the coverage boundary but on the no-coverage side can benefit somehow from coverage (e.g., voters in polling centers near coverage walk over the boundary to report fraud using their cell phones).

by corrupt officials expecting increased accountability in coverage areas. Using data on the universe of election complaints received, I show that access to mobile technology can be highly effective in promoting citizen participation. Polling centers in coverage areas report a significantly higher number of reported complaints, while the share of those complaints submitted by individuals, as opposed to government officials, also increases with coverage. We document a significant increase in the share of female complainants as well, suggesting that coverage can have a positive effect on citizen enfranchisement.

Second, I examine whether political violence by insurgent groups, which is strongly related to both cell phone coverage (Shapiro and Weidmann 2015) and electoral fraud (Weidmann and Callen 2013), can explain the drops in fraud at the coverage boundary. For instance, violence can directly impact turnout, which can then lead corrupt officials to engage in fraud to offset the low turnout.[3] In the case of Afghanistan, political violence is a potentially important channel since there is clear evidence that insurgent attacks surged on election day: the number of attacks surpassed the yearly average by a factor of eight, while the number of improvised explosive device (IED) explosions more than tripled. I replicate the spatial RD analysis using recently declassified significant actions (SIGACTs) data collected by Afghan and International Security Assistant (ISAF) forces on daily insurgent attacks. The data include, among other things, time (by the hour), geolocation (within meters), and the type of incident (direct fire, IED, etc.). Results from the RD exercise suggest that the documented drops in fraud at the boundary cannot be explained by similar drops in insurgent violence. Instead, the results provide evidence of a weak relationship between the number of insurgent attacks, IEDs, and direct fire incidents and coverage. With this in mind, a secondary, yet important, contribution of this paper is to advance our understanding of the relationship between cell phone access and insurgent violence.

Third, ethnicity-driven affinity for a candidate may change sharply at the coverage boundary if mobile providers give preference to certain ethnic groups by expanding coverage into their areas. In such cases, the coverage effect on fraud may be explained by sharp changes in ethnic composition at the coverage boundary, as voters with similar backgrounds as the candidate may be willing to vote in spite of violence or be more tolerant of corrupt behavior by the candidate. To explore this possibility, I georeference detailed tribal maps from the Tribal Hierarchy and Dictionary of Afghanistan (2007) containing information on the geographic distribution of more than 50 tribes and ethnic groups across south and east Afghanistan. I then overlay village geolocation data on the georeferenced maps to construct village-level indicators of primary tribe and tribal confederation for almost 18,000 villages in the region. The results replicate the spatial RD analysis using several indicators for village-candidate tribal match as the main outcomes. First, there is no evidence of significant changes in the ethnic and tribal composition of villages near the coverage boundary. Second, additional results on potential coordination between candidates and election officials along tribal lines show that although there is an

[3] See Condra et al. (2018a) for results on the effect of insurgent violence depressing turnout as well as for a detailed description of the violence dataset used in this paper.

increase in fraud in centers with a candidate-official tribal match, this effect does not vary with coverage.

This paper contributes to a growing effort to understand the role of information and communications technologies (ICT) on improving information transfer and social monitoring capacity. This is particularly important considering the rapid expansion in mobile services experienced in the developing world throughout the last decade. More generally, the results in this paper show how commonly available and accessible technology, such as cell phones, can exert a positive externality on institutional development. In that sense, the results add to a rapidly advancing literature on the effectiveness of ICT-based policies on a host of economic development outcomes (e.g., Jensen 2007; Aker 2010; Aker, Collier, and Vicente 2017).

This paper advances our understanding of the effectiveness of technology-based grassroots monitoring on illegal behavior within the realm of election fraud. Callen and Long (2015) implements a field experiment where individuals record photographs of the total vote tally at randomly selected polling centers during the 2010 Afghan parliamentary election. This monitoring technology, however, is conceptually different from the one I studied in this paper, as it does not necessarily rely on cell phone coverage. Aker, Collier, and Vicente (2017) explores the impact of several information interventions during the 2009 Mozambican election, including a short messaging service hotline, a monitoring technology that is very similar to the Afghan setting. However, while they show convincing evidence that the interventions decrease electoral misconduct, it is not clear whether this result would hold in a fragile security environment like Afghanistan or other areas of the world where nearly two billion people's lives are affected by fragility, conflict, and violence (World Bank and United Nations 2018). Prior to the election, the Taliban issued several warnings targeting polling centers and voters, while on election day the number of attacks exceeded the 2009 daily average by a factor of eight. In such cases, election-related violence hampers collective action incentives, as individuals fear retaliation or are simply unable to witness fraud if not present at the polling centers. The fact that I find significant drops in fraud at the coverage boundary suggests that monitoring technologies that rely on cell phone access may offer some degree of plausible deniability to potential whistleblowers and can be effective even in settings characterized by extreme political violence.[4]

Lastly, from an empirical standpoint, this paper contributes to the literature on spatial and, more specifically, geographic RD (e.g., Imbens and Zajonc 2011, Keele and Titiunik 2013), a literature that is rapidly growing as microlevel geospatial data become more available. More importantly, however, is that it illustrates an empirical framework that can be used by other studies trying to uncover heterogeneous effects of mobile phone coverage on any outcome variable using a spatial framework. Further, the possibility of obtaining spatially heterogeneous effects along a geographic boundary has key policy implications, as it can guide the design of localized policies.

The paper proceeds as follows: Section I provides a background of the Afghan 2009 presidential election as well as the nationwide audit that followed shortly

[4] See Chassang and Padro-i-Miquel (2014) for a detailed treatment on the importance of plausible deniability to incentivize monitoring and to avoid side contracting between monitors and misbehaving agents.

afterward. Section II describes the empirical method and reports the results of the effect of coverage on fraud. Section III explores the coverage-fraud channels. Section IV concludes.

I. The 2009 Afghan Election

A. *Background*

The 2009 Afghan presidential election marked the second election after the toppling of the Taliban regime in 2001. Fraud allegations during the 2004 presidential election led to the creation of the Electoral Complaints Commission (ECC) precisely to investigate and adjudicate fraud-related complaints for the upcoming 2009 presidential election. This constituted the first time in Afghan history that a formal channel for individuals to report electoral complaints was created. In addition to adjudication of complaints, the ECC was given the power to issue audits, recounts, and runoff elections if necessary (Electoral Complaints Commission 2010). To improve transparency and guarantee independence from the executive power, three of the five appointed ECC commissioners (including the chairman) were international experts directly appointed by the United Nations representative of the secretary-general. The two Afghan commissioners were selected from the Afghanistan Independent Human Rights Commission and the Supreme Court (National Democratic Institute 2010).

Allegations of fraud during the 2009 Afghan election were widespread. According to ECC's chairman Grant Kippen, the agency received more than 3,300 complaints, with close to 80 percent of these complaints received during the polling and counting period (Electoral Complaints Commission 2010). Most complaints—about 47 percent—dealt with polling and counting irregularities, followed by complaints on intimidation and violence at the center (about 26 percent). The remaining types of complaints were distributed between access to stations (11 percent), missing election materials at the center (4 percent), and other types (12 percent). This degree of citizen participation represented a major improvement from the 2004 presidential election, when no formal channel to file claims existed. A direct implication of this was the implementation of a nationwide audit that is discussed in detail in the following section.

B. *The Audit and Recount*

Election day took place on August 20, 2009. Eighteen days after the tallying of votes began, the ECC ordered a nationwide audit of polling stations after initial investigations of the received complaints revealed clear evidence of widespread fraud.[5] The audit called for the investigation of polling stations reporting unusually high turnout and an unusually high majority of votes for a single candidate. Specifically, the audit-triggering criteria were (i) stations in which 600 or more votes were cast, (ii) stations in which one candidate received 95 percent or more of

[5] For reference, a polling station is a physical location within a polling center. In the sample studied, the average number of polling stations per center is about 4, with some centers having up to 20 stations.

the total votes cast, and (iii) stations satisfying both (i) and (ii). The ECC referred to these categories as Category A, B, and C, respectively.[6]

The motivation for these criteria lay primarily in the particular design of the election and the unusual pattern of reported total votes per station. In particular, polling station managers were provided with a ballot book containing exactly 600 empty ballots (Electoral Complaints Commission 2010). However, a significant number of stations reported totals of exactly 600 or more votes cast. This was particularly unusual given the overall low turnout resulting from expectations around election-related violence (Khadhouri 2010). Such discrepancies in reported turnout can be clearly seen in Figure 1, panel A, which shows a histogram of total votes cast per station for the top two candidates. Notice the pronounced jump in the frequency of total votes cast at exactly 600 for then-candidate Hamid Karzai in particular.[7] The incidence of stations where a candidate obtained more than 95 percent of the total vote share was equally unusual. Note in Figure 1, panel B that a substantially high number of stations (with more than 100 total votes cast) had exactly 100 percent vote share for a single candidate (particularly Karzai).

The ECC classified 3,376 stations, or nearly 15 percent of all stations, as potentially fraudulent (i.e., falling in 1 of the 3 fraud categories mentioned above). Ultimately, the ECC performed a partial audit of all suspect stations given the need to determine in a timely manner whether a runoff election was needed. Particularly, 10 percent of the qualifying stations were randomly selected for a thorough investigation. From the inspected stations, the ECC created a "fraud coefficient" for each of the three categories described above. In essence, the fraud coefficients are the percentage of votes found to be fraudulent out of the total votes inspected within the category. Some indicators of fraud were ballot boxes with broken or tampered seals, uniform markings in most ballots, discrepancies in tally sheets and box totals, etc.

On October 18, nearly two months after election day, the ECC released the results of the audit. Once suspect votes were eliminated from the count, Hamid Karzai's vote share dropped from 54.6 to 49.67 percent, while the vote share of his main challenger, Abdullah Abdullah, went from 27.8 to 30.59 percent. In lieu of the results, the ECC ordered an immediate runoff election. However, the runoff election did not take place, as main challenger Abdullah withdrew from the race.

II. The Effect of Cell Phone Coverage on Fraud

A. *Data and Variables*

Measures of Fraud.—I use the list of polling stations that were subject to the audit and the ECC fraud categories to define the primary measure of fraud in this

[6] To be more specific, the ECC defined a total of six categories; however, given the similarity between some of the categories, I combine them into three aggregate categories. Refer to online Appendix A for a more detailed explanation of the audit categories.

[7] Also notice similar, although not as pronounced, peaks at various multiples of 50, starting with 200. See Beber and Scacco (2012) for a treatment on last-digit-based measures of electoral fraud. In the case of the 2009 Afghan election, the relatively high number of stations with a last digit of zero in their total votes provide, according to Beber and Scacco (2012), a sign of electoral manipulation. For a more specific treatment of the 2009 Afghan election that looks precisely at the measures developed in Beber and Scacco (2012), refer to Weidmann and Callen (2013).

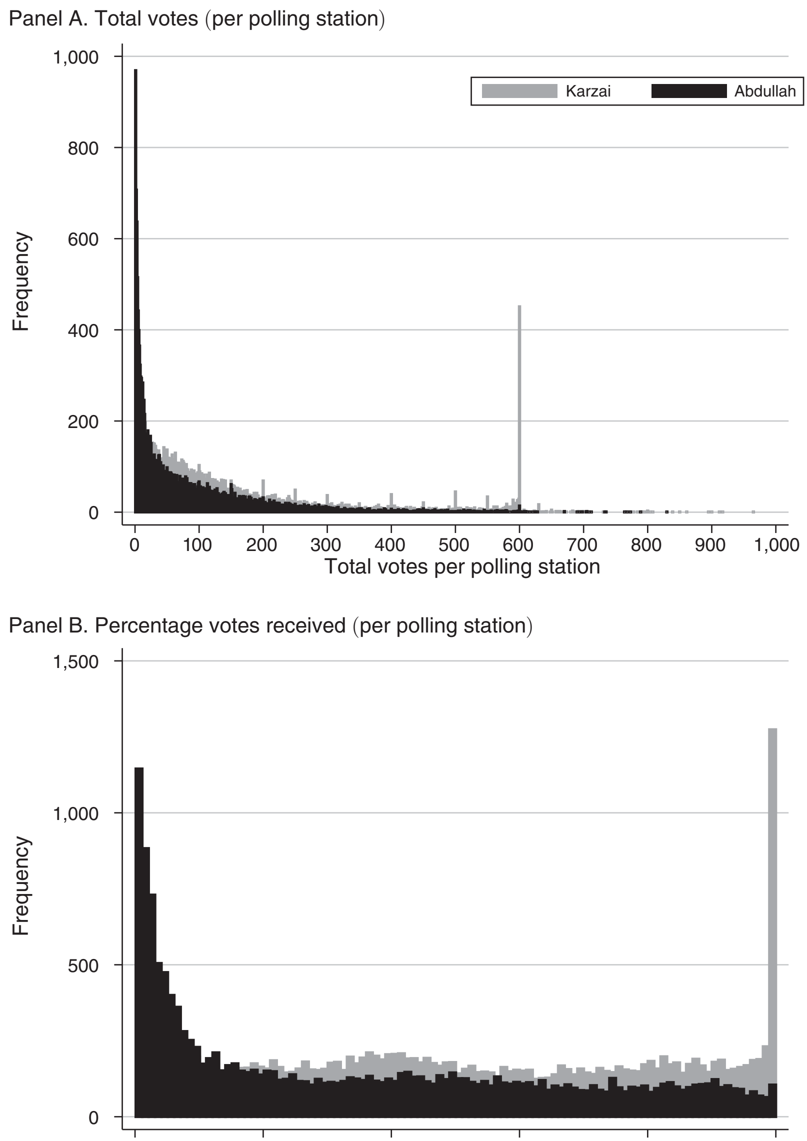

Figure 1. Total Votes and Vote Percentage Received by Top Two Candidates

Notes: Frequency of total votes received by the top two candidates at the station level. In panel A, sample restricted to stations where candidates obtained a nonzero number of votes. Bar width is 1. In panel B, sample is restricted to stations where a candidate obtained a positive share and stations where 100 or more total votes were cast. Bar width is 0.01.

paper. Specifically, I first aggregate the 6 categories used by the ECC to trigger an audit into 3 broader categories: *Category A* (stations with 600 or more votes cast), *Category B* (stations in which 1 candidate received 95 percent or more of the total votes cast), and *Category C* (stations satisfying categories *A* and *B* above).[8] I define two fraud outcomes in this paper. The first is (i) the polling-center-level vote share of stations that were under *Category C*.[9] More specifically, given a polling center c with a total of s stations of which $n \leq s$ qualify as category C, the measure of fraud at center c is given by the total number of votes in the n suspect stations divided by the total votes cast in center c. I refer to this measure as the *Share of votes under Category C*. The second is (ii) an indicator for whether polling center c has at least one disqualified or under *Category C* fraud station.[10]

Note that although the measures defined above are referred to as "measures of fraud," the fact that a station qualifies for one of the categories does not necessarily imply that fraud was committed in this station. One may have, for instance, stations with unusually high voter turnout rates or unusually strong preferences for one specific candidate. With this in mind, one should interpret this measure as a proxy for fraud. These proxies, however, provide a precise signal on actual fraud given the context and design of the election. For instance, in the case of *Category C* fraud, more than 96 percent of the ballots inspected in stations satisfying this category had credible evidence of tampering (Electoral Complaints Commission 2010). Furthermore, this particular measure has been previously used in the economics literature and cross validated with alternative measures of fraud in the same Afghan context (Weidmann and Callen 2013).

Cell Phone Coverage.—Cell phones are the primary medium of communication in Afghanistan, since fixed line phones are relatively scarce.[11] To determine areas with cell phone coverage, I use Global System for Mobile communications (GSM) second-generation (2G) coverage maps directly provided by cell phone operators to the GSM Association (GSMA) and distributed by Collins Bartholomew.[12] The coverage maps indicate areas receiving 2G coverage based on the spatial distribution of cell phone towers across Afghanistan. Specifically, coverage data are in the form of a map raster or grid file indicating cells where signal strength is at least -100 decibel-milliwatts (dBm). This is the typical *minimum received signal power* in GSM wireless networks, or broadly speaking, the minimum signal strength needed to be able to make a call (Figueiras and Frattasi 2010). Figure 2, panel A shows the 2G coverage raster file overlaid on a topographical map of Afghanistan. Shaded areas indicate areas with a signal strength

[8] Refer to online Appendix A for a detailed description of the construction of the three categories.

[9] The *Category C* measure used in this paper also includes a small number of stations that were disqualified prior to the audit due to evidence of tampering.

[10] The appeal of the indicator variable is that, while measurement error in vote counts can affect the *Share of votes under Category C* variable, it is not an issue for this dichotomous outcome.

[11] The number of mobile subscribers in Afghanistan rose from 1.7 million in 2006 to around 24 million by 2017. This translates into a mobile penetration (number of subscribers per 100 inhabitants) of 6.27 in 2006 and 67 in 2017. To put these numbers in perspective, the number of fixed-line phone subscribers, for example, was only 110,000 in 2012, even less than the number of internet users (Hamdard 2012).

[12] GSM is the type of cellular technology used by the Afghan cell phone companies in my sample and most mobile providers in the developing world. The GSM Association is a group comprising most GSM cell phone providers around the world. 2G stands for second generation, and this is the cellular network technology allowing mostly voice calls only (i.e., technology preceding smart phone, or third generation (3G), technology). The dataset is called the *Collins Coverage Explorer*, and more information can be found at https://www.collinsbartholomew.com/mobile-coverage-maps/mobile-coverage-explorer (GSMA 2009).

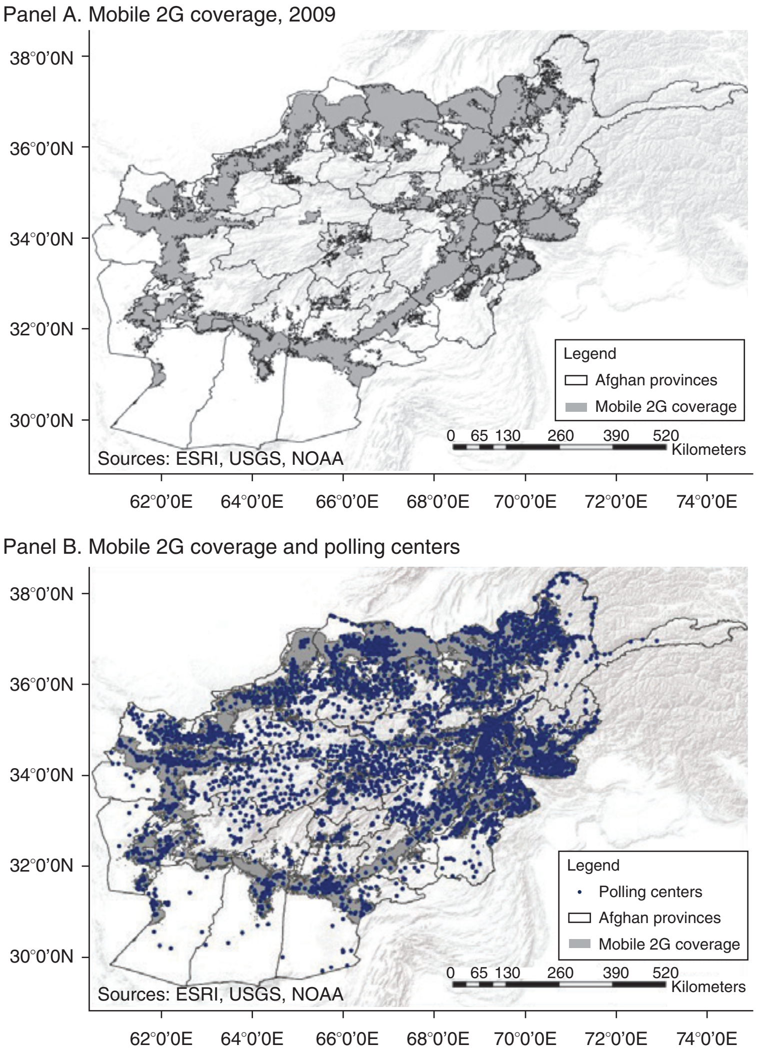

Figure 2. Mobile 2G Coverage and Polling Centers

Notes: Shaded areas represent availability of 2G GSM cell phone coverage for the year 2009. Dots give the location of polling centers during the 2009 Afghan presidential election. Lines demarcate the provinces of Afghanistan. Map overlaid on USGS topographic basemap.

of at least -100 dBm. Within coverage areas, however, Collins Bartholomew does not provide information on how the strength of coverage varies. In the case of Afghanistan, Collins Bartholomew provides information on two of the largest operators, Mobile Telecommunications Network (MTN) and Afghan Wireless (AWCC). These two operators encompass about 46 percent of all cell phone subscriptions in Afghanistan, with more than 8 million subscriptions combined (Hamdard 2012).

The lack of data on other providers may be a source of concern, since some covered areas may be wrongly classified as not covered. However, a detailed inspection of cell phone tower locations in 2012 using maps provided by the Afghan Telecommunication Regulatory Authority (ATRA) suggests significant overlap in tower locations and service areas between MTN, AWCC, and other operators. For the case of smaller operators (Etisalat, Wasel, and Afghan Telecom), the tower locations are entirely contained within the coverage areas of MTN and AWCC (Afghan Telecommunication Regulatory Authority 2012). However, to be careful in our assessment, Section IID presents results that account for the inclusion of Roshan—the largest operator in Afghanistan—in the coverage measure.

An additional source of concern may be the possibility of operators overreporting coverage areas. This might result if operators desire to overstate their service areas for marketing purposes, for example. However, coverage data submissions by operators are considered to be a service to the GSM Association. Operators provide data at no cost, which the GSM Association then uses to assess the state of the technology and sell it as a way of raising funds for the agency to operate. With this in mind, the data are restricted to the general public and require a contractual agreement to purchase and use for research. Therefore, it is unlikely that operators have an incentive to misreport coverage in such cases.

Polling Center Characteristics.—I obtain data on the latitude and longitude of polling centers from an Independent Elections Commission (IEC)–led nationwide inspection of each polling center that took place less than a year after the 2009 presidential election. The purpose of the inspection was to assess the security status and accessibility of designated polling centers for the upcoming September 2010 parliamentary election. The assessments were conducted jointly by ISAF and Afghan National Security Force teams. Each assessment included four pieces of information: a polling center name and code, an MGRS grid providing the exact geographic location of the polling center, and a road accessibility status. Using the coordinates, I overlay the centers on the cell phone coverage map to determine each center's coverage status. Figure 2, panel B depicts the spatial distribution of polling centers along the coverage areas.

To create a sample containing the fraud measures per center along with the geographic location of the centers, I merge the 2010 center assessment data described above with the polling-center-level data on fraud outcomes described at the beginning of this section. The data are merged based on the polling center code and name. In cases where the codes matched but the names did not (100 cases), the match was done based on the names only. The total sample consists of 6,160 polling center observations, for which 5,904 (95.8 percent) have coordinates obtained directly from the 2010 assessment. For the remaining 256 centers, coordinates were imputed as follows: 169 (2.7 percent) used the centroid coordinates of the village or settlement where the center was located, 81 (1.3 percent) used the coordinates of the center with the identifier code closest to it, and, lastly, 6 (0.1 percent) simply used the coordinates of the district capital where the center was located.[13]

I use the released electoral results to obtain additional election-related outcomes: the number of expected voters prior to election day, the total votes cast at the center, the total number of stations per center, the voter turnout rate, and the

[13] Online Appendix Table B1 provides a detailed description of the sample and imputations used.

percentages received by the two main candidates.[14] These data are complemented with preelection data published by the IEC on polling center type (school, mosque, or other) along with the share of stations within a center designated to women and Kuchis, a minority ethnic group. Geographic and economic development characteristics of the area where each center is located are extracted using geographic information system resources. Specifically, I calculate distances from polling centers to primary and secondary roads, district hospitals, basic health centers, and primary and secondary rivers using vector files collected by the Afghanistan Information Management Service (AIMS) and obtained from the Empirical Studies of Conflict Project (AIMS 1997–2005). Information on exogenous geographic characteristics such as polling center elevation and slope is obtained from NASA's Shuttle Radar Topography Mission (SRTM30) (National Aeronautics and Space Administration and the National Geospatial Intelligence Agency 2000).

Lastly, demographic data on the population and ethnic composition around the location of the polling center comes from the Measuring Impacts of Stabilization Initiatives (MISTI) project sponsored by the US Agency for International Development (USAID). The MISTI project (MISTIs 2013) includes geographic coordinates and compiles demographic data from various data sources between the years 2012 and 2013 for more than 37,000 villages across Afghanistan. Using these data, I create variables indicating the population size and the language spoken ("Pashto," "Dari," and "Other") in the village or settlement closest to the polling center.

B. *Regression Discontinuity Design*

This section presents the spatial RD framework used to estimate the effect of cell phone coverage on fraud. It also describes and tests the validity of the identifying assumptions.

Note from Figure 2, panel A that (i) cell phone coverage is a discontinuous function of latitude and longitude and (ii) changes from coverage to noncoverage areas define a two-dimensional boundary along the latitude-longitude space.[15] With this in mind, I employ a spatial regression discontinuity (RD) design that takes advantage of the discontinuity in polling centers' cell phone access to estimate the effect of coverage on various election fraud outcomes. I present results using two approaches. First, I follow the usual approach in the literature by specifying a one-dimensional forcing variable, namely the distance to the closest point in the coverage boundary.[16] This is the equivalent of subtracting the cutoff value from the forcing variable in the one-dimensional design and then using this transformed forcing variable to estimate a single, boundary-wide average effect. Second, I exploit the

[14] The results data are publicly accessible at https://www.iec.org.af/results_2009 (IEC 2009).

[15] We acknowledge that the coverage boundary might slightly oscillate throughout the day due to day-to-day variations in temperature and atmospheric conditions, among other things. Results using a sharp RD design are not affected by the potential fuzziness of the boundary. However, they should be interpreted as an intent-to-treat effect. Section IID addresses this in further detail.

[16] See Holmes (1998); Black (1999); Kane, Riegg, and Staiger (2006); Lalive (2008); and Dell (2010) for examples.

two-dimensional nature of the coverage boundary to estimate *boundary treatment effects* at various points along the treatment boundary following Imbens and Zajonc (2011).[17]

In the case of the one-dimensional or scalar approach, I estimate various specifications of the equation below:

$$v_{f,ij} = \gamma + \beta D_{ij} + g(\mathbf{X}_{ij}) + \Omega_i + \epsilon_{ij}, \tag{1}$$

where $v_{f,ij}$ denotes a fraud measure for polling center j in boundary segment i, D_{ij} is an indicator equaling one if the center lies within the coverage area, and $\mathbf{X}_{ij}$ is the geographic coordinate of center j in segment i. The term Ω_i is a boundary segment fixed effect that ensures that we are comparing polling centers that are within the same segment of the coverage boundary. In addition, to satisfy the boundary positivity assumption described in Imbens and Zajonc (2011), we drop any boundary segments with polling centers on only one side of the coverage boundary. This ensures that, within each segment, there are polling centers on the noncoverage side to serve as counterfactuals for treated polling centers.[18]

The RD polynomial $g(\mathbf{X}_{ij})$ and sample restrictions vary with different specifications of (1). First, I use $g(\mathbf{X}_{ij}) = \alpha \cdot dist_{ij} + \delta D_{ij} \times dist_{ij}$, where the forcing variable $dist_{ij}$ denotes the Euclidean distance between polling center j and the closest point on the coverage boundary. This specification estimates separate lines on each side of the coverage boundary, with the estimation sample restricted to polling centers falling within a bandwidth around the coverage boundary that is chosen optimally following Calonico, Cattaneo, and Titiunik (2014). Second, I follow a more parametric approach that uses all observations on either side of the coverage boundary. However, I allow a more flexible form for the RD polynomial by using higher order polynomials in distance to the boundary. For instance, the RD polynomial of order K is given by $g(\mathbf{X}_{ij}) = \sum_{k=1}^{K} \alpha_k \cdot dist_{ij}^k + \delta_k D_{ij} \times dist_{ij}^k$. The optimal order of the chosen polynomial specification is determined using Akaike's criterion as in Black, Galdo, and Smith (2007) and suggested in Lee and Lemieux (2010). RD coefficient β gives the causal effect of cell phone coverage on fraud for areas in close proximity to the coverage boundary.

The second approach estimates treatment effects using observations within a neighborhood of a specific point in the treatment boundary. This exercise is then repeated for various points along this boundary, thus providing a distribution of these effects along this dimension. However, since there are not enough observations within several neighborhoods to allow for consistent estimation of the boundary

[17] Although there are multiple studies exploring RD methods with a multidimensional forcing variable (e.g., Wong, Steiner, and Cook 2013; Keele and Titiunik 2013), we mostly follow the notation and terminology in Imbens and Zajonc (2011).

[18] Refer to Keele and Titiunik (2013) for a description of the importance of considering segment fixed effects in the context of geographic RD design. To create the segments, I split the boundary into segments of 50 km. I then calculate the distance between polling centers and the segment on this boundary. All the centers closest to one specific segment form the neighborhood around that segment. These neighborhoods are the spatial unit for the segment fixed effects. In all, there are 947 segments that are matched to their closest polling centers. After dropping segments to ensure boundary positivity, we are left with 241 segments. Online Appendix Figure B9 shows an example of these segments with the corresponding neighborhood of polling centers.

treatment effects at each point, I propose a modification that uses all available observations. More specifically, let $\mathcal{C}$ and $\mathcal{B} = bd(\mathcal{C})$ denote the cell phone coverage area and its boundary, respectively. Polling center j receives treatment assignment (i.e., coverage) if its corresponding coordinate vector $\mathbf{x}_j = (longitude_j, latitude_j)$ falls within the coverage area $\mathcal{C}$. Let $\mathbf{b}_i$ with $i = 1, \ldots, I$ denote the coordinate vector of point i on the treatment boundary $\mathcal{B}$. Furthermore, let $N_h(\mathbf{b}_i)$ denote a neighborhood of size h km around this point, with $N_h^+(\mathbf{b}_i)$ and $N_h^-(\mathbf{b}_i)$ denoting the subset of this neighborhood that falls on the coverage and noncoverage sides of the boundary, respectively. As shown in Imbens and Zajonc (2011), the boundary treatment effect at point $\mathbf{b}_i$, denoted as $\tau(\mathbf{b}_i)$ is therefore given by

$$\tau(\mathbf{b}_i) = \lim_{\mathbf{X} \to \mathbf{b}_i} E\left[v_f | \mathbf{X} \in N_h^+(\mathbf{b}_i)\right] - \lim_{\mathbf{X} \to \mathbf{b}_i} E\left[v_f | \mathbf{X} \in N_h^-(\mathbf{b}_i)\right], \tag{2}$$

where v_f is a measure of electoral fraud. I obtain an estimate of $\tau(\mathbf{b}_i)$ by estimating separate lines on each side of the coverage boundary. More specifically, I estimate

$$v_{f,ij} = \gamma + \beta D_{ij} + \mathbf{X}_{ij}'\alpha + D_{ij}\mathbf{X}_{ij}'\delta + \Omega_i + \epsilon_{ij} \tag{3}$$

for centers within h km of the coverage boundary[19] and where $v_{f,ij}$ denotes a fraud measure for polling center j in neighborhood i, D_{ij} is an indicator equaling one if the center lies within the coverage area, $\mathbf{X}_{ij}$ is the geographic coordinate of center j in neighborhood i, and Ω_i is a neighborhood fixed effect. The inclusion of neighborhood fixed effects ensures that I am comparing centers that are within a neighborhood of the boundary point. I choose h optimally, as in Calonico, Cattaneo, and Titiunik (2014). Lastly, to comply with the boundary positivity assumption discussed in Imbens and Zajonc (2011), I restrict the sample to only neighborhoods with at least one polling center on each side of the coverage boundary.[20]

From equation (3), and under certain conditions, a consistent estimator for $\tau(\mathbf{b}_i)$ is given by

$$\hat{\tau}(\mathbf{b}_i) = \hat{\beta} + \mathbf{b}_i'\hat{\delta}. \tag{4}$$

Such conditions are discussed in detail in the following section. In order to evaluate the treatment effect at various points in the boundary, I follow Imbens and Zajonc (2011) by choosing a number of evenly spaced boundary points $\mathbf{b}_i$ that cover the boundary reasonably well. I highlight a point regarding the modification proposed above: estimation of the boundary treatment effects follows from using the actual levels of the forcing variable (i.e., latitude and longitude) rather than the normalized levels (i.e., the distance to the boundary) as it is usually done in the literature and in the scalar RD method described above. Simply put, this is the equivalent of *not*

[19] This is the equivalent of using a rectangular kernel with bandwidth h. Kernel choice has little impact in practice; therefore, simple kernels (i.e., rectangular) can be used for convenience (Lee and Lemieux 2010).

[20] Section IIC describes in detail how the neighborhoods are created. Boundary positivity requires the existence of observations near the boundary in order to identify the treatment effect in the multidimensional RD setting. More specifically, boundary positivity requires that for all $\mathbf{b}_i$ and $\epsilon > 0$, there are polling centers for which $P(\mathbf{x}_j \in N_h(\mathbf{b}_i)) > 0$.

subtracting the treatment threshold from the forcing variable in the one-dimensional case. From the estimation equation (3), this guarantees an estimate of the treatment effect that depends on specific values of the forcing variables at the boundary (i.e, the coordinate of boundary point $\mathbf{b}_i$).

Lastly, I use the estimated boundary treatment effects from equation (4) to estimate a boundary-wide average effect, τ, as

$$\hat{\tau} = \frac{\sum_{i=1}^{I} \hat{\tau}(\mathbf{b}_i) \cdot \hat{f}(\mathbf{b}_i)}{\sum_{i=1}^{I} \hat{f}(\mathbf{b}_i)}, \tag{5}$$

where $\hat{f}(\cdot)$ is the estimated bivariate density of polling centers' coordinate vectors evaluated at boundary points $\mathbf{b}_i$. Following the notation described above, expression (5) provides an estimate of the average effect τ given by $\int_{\mathbf{x} \in \mathcal{B}} \tau(\mathbf{x}) f(\mathbf{x} \mid \mathbf{X} \in \mathcal{B}) d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{B}} \tau(\mathbf{x}) \cdot f(\mathbf{x})\, d\mathbf{x} / \int_{\mathbf{x} \in \mathcal{B}} f(\mathbf{x})\, d\mathbf{x}$. In subsequent discussions of results, I refer to the estimate in equation (5) as the *averaged boundary treatment effect*.

Validity of the RD Identifying Assumptions.—Identification of $\tau(\mathbf{b}_i)$ requires a key assumption: potential outcome functions $E[v_f(1) \mid \mathbf{X}]$ and $E[v_f(0) \mid \mathbf{X}]$ must be continuous at point $\mathbf{b}_i$ in the treatment boundary, where one and zero denote assignment and nonassignment into treatment, respectively. Simply put, polling center characteristics must transition smoothly across the treatment boundary. This assumption allows for centers in the noncoverage side to serve as a valid counterfactual for centers in the coverage side.

Table 1 assesses the validity of the design by comparing electoral outcomes and geographic, economic, and demographic characteristics for centers on each side of the coverage boundary. In addition, it investigates how the primary fraud outcome measure varies across the boundary relative to other polling center characteristics. Columns 1 and 4 report the mean for polling centers within cell phone coverage areas for bandwidths of 10 km and 5 km, respectively. Columns 2 and 5 report the mean for centers in noncoverage areas within the specified bandwidths. Columns 3 and 6 report the clustered standard error of the difference in means between covered and noncovered centers.[21] I highlight two important results. First, note that differences across the boundary for the fraud outcome variable remain economically and statistically significant as the bandwidth decreases. Note that the results exhibit a high degree of spatial variation with centers in the southeast showing significant differences, while differences in the northwest region are indistinguishable from zero. For this reason, the main analysis presented in Section IIC will be performed separately by region.[22] Second, and most importantly in terms of design validity, notice that, unlike the fraud measure, most differences in polling center characteristics become relatively small and statistically insignificant as the bandwidth decreases. To offer a more rigorous assessment, column 7 presents the results from an RD

[21] Standard errors are clustered at the boundary neighborhood level. Refer to Section IIC for a detailed description of how boundary neighborhoods are defined. Online Appendix Table B3 shows the results in Table 1 using Conley (1999) standard errors. Note that the results do not differ greatly in terms of the type of clustering used.

[22] Given the separate analysis by region, online Appendix Table B2 presents a replication of Table 1 using only the southeast region, where most of the differences in fraud measures are found.

Table 1—Mean Comparison for Various Polling Center Characteristics

	Within 10 km of boundary			Within 5 km of boundary			RD estimates	
	Coverage (1)	No coverage (2)	Standard error (3)	Coverage (4)	No coverage (5)	Standard error (6)	RD coefficient (7)	Standard error (8)
Fraud outcomes (share of votes under Category C fraud)								
All regions	0.08	0.11	(0.02)	0.08	0.12	(0.02)	−0.04	(0.019)
East and South	0.14	0.20	(0.03)	0.13	0.20	(0.03)	−0.08	(0.032)
North and West	0.01	0.01	(0.01)	0.01	0.01	(0.01)	0.00	(0.010)
Electoral outcomes								
Number of stations	4.09	3.73	(0.16)	3.86	3.78	(0.18)	0.09	(0.175)
Number of expected voters	2,194.00	1,944.00	(94.04)	2,069.00	1,979.00	(100.00)	94.33	(101.198)
Total votes	871.80	866.60	(56.34)	835.00	863.70	(62.41)	−28.22	(59.992)
Voter turnout	0.43	0.50	(0.02)	0.45	0.49	(0.02)	−0.03	(0.023)
Vote share:								
Karzai	0.50	0.49	(0.03)	0.50	0.49	(0.03)	0.01	(0.030)
Abdullah	0.34	0.33	(0.03)	0.35	0.33	(0.03)	0.01	(0.031)
Polling center characteristics								
Polling center type:								
Mosque	0.24	0.26	(0.03)	0.25	0.26	(0.03)	−0.01	(0.029)
School	0.46	0.37	(0.03)	0.44	0.37	(0.03)	0.06	(0.034)
Other type	0.30	0.37	(0.03)	0.30	0.37	(0.03)	−0.05	(0.032)
Polling center access (2010):								
Road access	0.76	0.78	(0.04)	0.73	0.78	(0.04)	−0.05	(0.039)
Limited access	0.08	0.07	(0.02)	0.10	0.07	(0.02)	0.03	(0.023)
Other access	0.16	0.15	(0.03)	0.17	0.15	(0.03)	0.02	(0.032)
Share female stations	0.44	0.45	(0.01)	0.44	0.45	(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)
Geographic characteristics								
Elevation (meters)	1,570.00	1,782.00	(58.06)	1,617.00	1,756.00	(50.71)	−128.72	(50.617)
Slope (percent)	5.72	7.57	(0.53)	6.46	7.66	(0.58)	−1.01	(0.602)
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)
Secondary road (2005)	44.65	52.60	(3.37)	50.08	49.84	(3.12)	0.49	(3.056)
District hospital (2005)	37.56	45.42	(2.76)	40.25	42.68	(2.73)	−2.34	(2.739)
Basic health center (2005)	20.12	24.31	(1.87)	21.78	22.68	(1.74)	−0.38	(1.837)
Primary river	17.45	18.38	(1.40)	18.00	17.90	(1.29)	0.19	(1.319)
Secondary river	8.11	8.71	(1.12)	8.96	8.03	(1.10)	0.61	(1.040)
Seasonal river	12.66	11.05	(1.54)	13.16	10.30	(1.77)	2.70	(1.669)
District/province capital	0.05	0.03	(0.01)	0.04	0.03	(0.01)	0.01	(0.013)
Demographic characteristics (of closest settlement)								
Population (2012–2013)	1,287.00	957.00	(223.69)	1,004.00	1,018.00	(145.37)	−51.25	(141.521)
Language spoken (2012–2013):								
Dari	0.43	0.46	(0.04)	0.43	0.43	(0.04)	0.00	(0.040)
Pashto	0.44	0.43	(0.04)	0.43	0.45	(0.04)	−0.01	(0.041)
Other	0.13	0.12	(0.03)	0.14	0.12	(0.03)	0.01	(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 parentheses. The sample is restricted to neighborhoods with at least one observation on each side of the boundary. "Number 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 IIA of the text. Year values in parentheses indicate the year the data was collected. Variables without year indication were collected in 2009 or are time-invariant variables. Refer to Section IIB for a description of the RD model used in columns 7 and 8.

analysis that estimates equation (1) within a 5 km bandwidth using each of the specified covariates in Table 1 as the outcome variable.[23] Similar to the mean difference results, the RD exercise shows that, unlike the fraud measure, center characteristics transition smoothly across the boundary for the most part. In all, 24 out of the 28 baseline characteristics tested result in statistically insignificant differences between covered and noncovered centers.

Polling center elevation, slope, and distance to the closest primary road are notable exceptions. Cell phone coverage depends on topographical features; thus, it is plausible that coverage drops in areas with significant changes in elevation and slope. Similarly, primary road access is affected by the ruggedness of the terrain. In spite of these changes across the boundary, robustness checks show that the main RD results in Section IIC are not sensitive to the inclusion of these covariates.[24]

To further assess the validity of the identifying assumption, I perform Cattaneo, Jansson, and Ma's (2019) recent test for breaks in the density of the forcing variable at the treatment boundary and find no evidence of endogenous assignment of polling centers or sorting of villages near the boundary.[25] In the context of this study, however, endogenous sorting of centers close to the boundary is not a cause of concern, since polling center locations were determined primarily by the location of settlements rather than by cell phone coverage. In addition, locations were determined entirely by the UN-led IEC; thus, manipulation of the process by potentially corrupt candidates is unlikely. Similarly, the absence of selective sorting of villages near the coverage boundary is institutionally plausible. Afghanistan experienced a period of rapid expansion in cell phone coverage throughout the second half of the 2000s. With this in mind, the incentives for households to move to a village that has coverage are very low when coverage might soon reach that household's village. Despite this, Section IID provides results that take into account the possibility of endogenous selection into coverage.

C. *Results*

This section begins by describing the results from a graphical analysis of the outcome variables. It then proceeds with a description of the results from the one-dimensional and boundary RD designs described in Section IIB. Given the inherent differences across regions of Afghanistan, this section presents results separately by region.[26]

[23] Column 7 uses a cubic polynomial in distance to the boundary as the specification of equation (1). The bandwidth choice of 5 km is to allow for comparability with the results from columns 4 and 5.

[24] For a graphical depiction of the continuity of baseline covariates across the coverage boundary, refer to online Appendix Figure B1, which presents RD plots for continuous 4 km distance bins for all covariates in Table 1.

[25] Panels A and B of online Appendix Figure B2 show a histogram of the forcing variable, and the results from the Cattaneo, Jansson, and Ma (2019) test for discontinuities in the density of the forcing variable, respectively. Panels C and D repeat the exercise using distance between villages and the coverage boundary.

[26] I define the regions based on the International Security Assistance Forces (ISAF) regional commands classification specified in the Measuring Impacts of Stabilization Initiatives (MISTI) dataset (MISTI 2013). Refer to online Appendix Figure B5 for a depiction of the two regions.

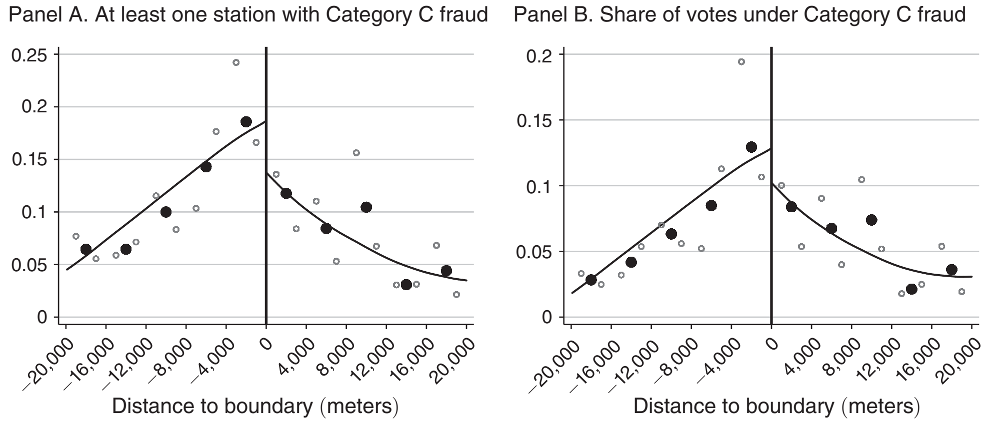

Figure 3. Binned Averages for Category C Fraud (RD plots)

Notes: Solid dots give the average share of votes classifying in Category C fraud for polling centers falling within 4,000-meter distance bins. Hollow dots give the average share of votes classifying in Category C fraud for polling centers falling within 2,000-meter distance bins. Refer to Section IIA 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 noncoverage 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.

Graphical Analysis.—I begin by graphically analyzing the relationship between electoral fraud and cell phone coverage using RD plots of the outcome variables. Figure 3, panel A plots the likelihood that a polling center has at least one station with Category C fraud. Figure 3, panel B plots the average share of Category C fraud votes per polling center. The figures provide two levels of smoothing: solid dots represent the averages of the outcome variables for 4 km distance bins, while hollow dots use 2 km distance bins. Negative values of distance indicate polling centers in noncoverage 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. The window of analysis is 20 km on each side of the boundary, and the regressions are estimated separately on each side.

The RD plots show that, within a narrow window around the coverage threshold, there is a clear drop in the levels of fraud for centers located on the coverage side. The likelihood of a polling center reporting fraud drops by about 5 percentage points (Figure 3, panel A), while the average share of fraudulent votes drops by about 2.5 percentage points (Figure 3, panel B). In relative terms, these are economically significant drops considering the average values in centers on the noncoverage side. Note that fraud levels also decrease gradually with coverage. There is also a decrease in fraud levels on the noncoverage side as one moves away from the boundary. This is likely attributed to fraud being less attractive as one moves into areas with smaller populations.[27]

[27] Online Appendix Figures B3 and B4 present RD plots by region and including confidence intervals, respectively.

Table 2—Effect of Cell Phone Coverage on Category C Fraud

	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
	(0.032)	(0.029)	(0.055)	(0.053)	(0.024)	(0.023)
Observations	1,074	2,039	532	1,087	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	—
Neighborhoods	230	237	95	101	133	137
Panel B. Share of votes under Category C fraud						
Inside coverage	−0.039	−0.041	−0.067	−0.094	0.019	0.006
	(0.021)	(0.020)	(0.040)	(0.040)	(0.012)	(0.015)
Observations	1,064	2,039	528	1,087	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	—
Neighborhoods	228	237	95	101	132	137

Notes: Results use equation (1). Refer to Section IIB for a description. Optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik (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, Galdo, and Smith (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to Section IIC for a description of how boundary neighborhoods are created.

One-Dimensional RD.—Table 2 presents the results from the one-dimensional RD design that estimates the causal impact of coverage on fraud using equation (1). Results are also presented separately by geographic region. Further, all specifications include boundary segment fixed effects and standard errors clustered by boundary segment to account for potential spatial correlation of the error terms within segments. The segments used for the fixed effects are created by splitting the boundary into equal segments. The group of polling centers closest to a specific segment form that segment's neighborhood. This ensures that we are comparing centers within close distance of each other.[28] Table 2 also provides the mean of the dependent variable for centers in the noncoverage side for reference.

Columns 1, 3, and 5 present results using distance to the coverage boundary as the forcing variable and restricting the analysis to an optimal bandwidth around the coverage boundary. Results in column 1 indicate a considerable drop in fraud for polling centers within the coverage area. In particular, for centers within a 7.3 km bandwidth, the likelihood of fraudulent stations drops by about 7.7 percentage points for centers on the coverage side. With respect to the share of votes under Category C fraud, using a bandwidth of 7.2 km, centers on the coverage side report a drop of about 4 percentage points. Notice in columns 3 and 5 that there is a high degree of spatial heterogeneity in the results. While polling centers in the southern and eastern provinces report economically and statistically significant drops in fraud, the RD

[28]The length of segments is 50 km. Refer to Keele and Titiunik (2013) for a description on the importance of within-segment comparisons in the spatial RD setting. Online Appendix Figure B9 presents an example of these segments.

estimates for the northern and western provinces are indistinguishable from zero. This is likely attributed to the fact that, overall, fraud levels in the northwest were much lower relative to the southeast. On average, less than 2 percent of the votes per center classify as potentially fraudulent in the northwest, compared to more than 22 percent in the southeastern region. Hence, any estimates of differences in fraud across the boundary should indeed be relatively small for this region.

Specifications in columns 2, 4, and 6 follow the parametric approach described in Section IIB. The results are quite comparable to the results obtained from the optimal bandwidth specification: the likelihood of fraudulent stations drops by about 8 percentage points, while the share of fraudulent votes drops by about 4 percentage points in covered centers.

We note that results are robust to different choices of bandwidth, polynomial order, clustering of the standard errors, and using district fixed effects instead of the segment fixed effects used in the main analysis. These results are presented in online Appendix Figure B6 and online Appendix Tables B5 and B6, respectively.[29] Results using disqualified stations and stations with 600 or more votes as alternative measures of fraud remain qualitatively similar; however, they tend to be less precise (online Appendix Table B7). In the case of disqualified stations, this is expected given their low number relative to the number of stations under the Category C fraud measure used in the main analysis. In the case of stations with 600 or more votes, one has to be careful in interpreting the results since, in principle, a station with that vote count could just be the result of high turnout in that area and not necessarily fraudulent ballot stuffing. Lastly, consistent with a valid RD design, after the inclusion of a set of baseline covariates, the magnitude of the estimates does not change substantially, while the precision improves to some degree (online Appendix Table B4).

Boundary RD.—To assess the degree of spatial heterogeneity in the impact of cell phone coverage on electoral fraud, I estimate boundary treatment effects at various points along the coverage boundary using equations (3) and (4). As suggested in Imbens and Zajonc (2011), I choose a random number of boundary points $\mathbf{b}_i$ that cover the entire boundary reasonably well. The points have a minimum distance of 50 km between each other. This results in a total of 1,437 boundary points. All specifications of equation (3) include boundary neighborhood fixed effects and standard errors clustered by neighborhood in order to account for spatial correlation of the error terms within neighborhoods. Neighborhoods are determined by first calculating the Euclidean distance between polling centers and boundary points $\mathbf{b}_i$. Centers that are closest to a given boundary point and within the specified bandwidth around the boundary define a neighborhood. To assess the statistical significance of the estimated boundary effects, I calculate the standard errors of the estimates of $\tau(\mathbf{b}_i)$ using the delta method described in Greene (2003). Lastly, as with the one-dimensional RD results, all analyses are done separately by region.

Figure 4 presents the estimated boundary treatment effects on a map of Afghanistan. Panels A and B present results for each of the fraud outcomes. The

[29] Online Appendix Table B5 presents results using clustering of the standard errors at the polling center level. Online Appendix Table B6 presents results using district fixed effects.

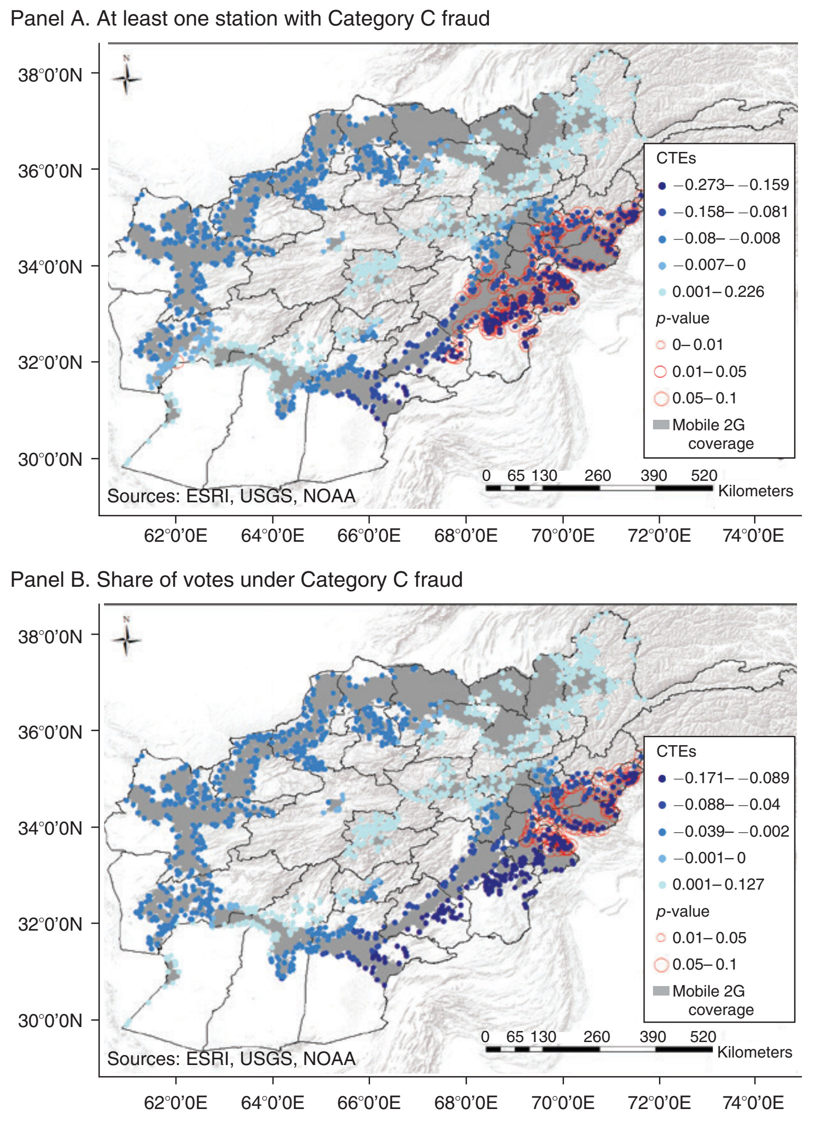

Figure 4. Spatial Distribution of Boundary Treatment Effects (Category C Fraud)

Notes: Shaded areas represent cell phone coverage. Dots indicate the location of $\mathbf{b}_i$ evaluated in equation (4). The color of the dots represents the magnitude of the estimated effect. The effects are estimated using equation (4). Hollow circles of different size around the dots represent the *p*-values of the estimated effects. Standard errors are clustered by neighborhood. Refer to the legend for specific values.

analysis in panel A uses bandwidths of 6.10 and 7.68 km for the southeastern and northwestern areas, respectively, while panel B uses 5.96 and 7.03 km bandwidths for the southeastern and northwestern areas, respectively. Shaded areas represent cell phone coverage. Dots indicate the location of the boundary points $\mathbf{b}_i$. The color of the dots, presented in a monochromatic scale, give the magnitude of the estimated effects. Refer to the legend for specific cutoffs. Statistically significant effects are highlighted with hollow circles representing the 1, 5, and 10 percent significance thresholds of the estimated *p*-values.

Similar to the results from the one-dimensional design, there is clear evidence that the share of fraudulent votes drops significantly at the coverage boundary. The magnitudes of the effects, however, are highly heterogenous both across and within regions of Afghanistan. Note that most of the economically significant effects are in the eastern part of the country. For instance, the magnitude of the drop in the share of fraudulent votes (panel B) in this area ranges between 9 to 17 percentage points. Similarly, most of the statistically significant effects appear in this area. Boundary treatment effects for other portions of the boundary within this region, although lower in magnitude, exhibit a negative sign. In all, about 71 percent (457 out of 642) of the boundary points evaluated in this area indicate a drop in the likelihood of fraudulent stations for centers within the coverage area relative to centers outside. Although some boundary points indicate a positive sign in the effect (and hence an increase in fraud due to coverage), none are statistically significant, and they show an average magnitude that is almost half the average of the boundary treatment effects with negative signs. Specifically, the average of the negative boundary treatment effects for the share of votes under Category C is about 7.3 percentage points, whereas the average magnitude for positive boundary treatment effects is about 4.1 percentage points. As previously shown, the estimated boundary effects for the northwestern region, however, are close to zero.

For a depiction of the variability of the boundary treatment effects, Figure 5 presents histograms of the estimates $\tau(\mathbf{b}_i)$ by region. The solid vertical line gives the arithmetic mean of the estimated boundary treatment effects. Notice that for both outcomes in the southeastern region, the estimated effects are largely negative but with significant variability. Estimates for the northwest region, on the other hand, are mostly clustered around zero.

Table 3 presents the averaged boundary treatment effects estimated from equation (5). I estimate standard errors via bootstrap with 500 replications and resampling within districts. For reference, Table 3 also presents the mean levels of fraud outcomes for centers within the specified bandwidth and outside the coverage boundary. The likelihood of a fraudulent station is, on average, about 5.7 percentage points lower inside coverage (panel A, column 1). The share of fraudulent votes drops by about 5.5 percentage points in the southeastern region from a baseline average in noncoverage centers of about 22 percent (panel B, column 2). Similar to the one-dimensional RD results, centers in the northwest region do not exhibit significant differences in fraud levels across the boundary. More importantly, note that the results are robust to the estimation design used: estimates from the one-dimensional and boundary RD designs are close in magnitude and precision.

D. *Additional Results*

This section examines whether accounting for selection into cell phone coverage has an effect on the results presented in Section IIC. It also examines the possibility of coverage simply leading to a reallocation of fraud to noncoverage areas and the possibility of coverage spillovers into nearby noncoverage areas.

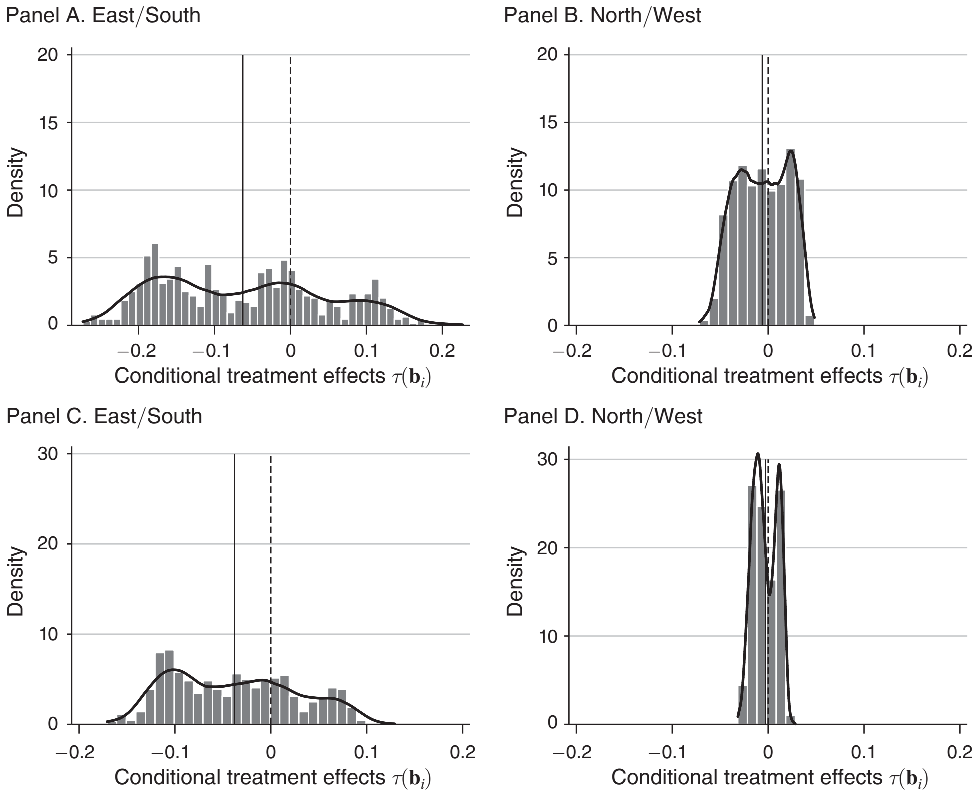

FIGURE 5. DISTRIBUTION OF BOUNDARY TREATMENT EFFECTS (CATEGORY C FRAUD)

Notes: Panels A and B present a histogram of the estimated boundary effects for the outcome "At least one station under Category C fraud." Panels C and D present a histogram of the estimated boundary effects for the outcome "Share of votes under Category C fraud." Refer to Section IIB for description of estimation. The solid vertical line gives the simple average of the estimated effects. The density estimate, presented as a solid line, uses an Epanechnikov kernel function.

Selection into Coverage.—Selection into coverage near the boundary can pose a threat to identification in the regression discontinuity setting. Specifically, one might be concerned that mobile service providers strategically choose their tower locations in a way that settlement characteristics, such as density, urbanization, and education levels, change sharply at the coverage boundary. If this is the case, then the observed drops in fraud levels at the boundary can be the result of corresponding changes in these potential correlates of fraud.

Although the discussion in Section IIB suggest that this is unlikely, this section further addresses this concern by introducing a regression discontinuity differences design that takes into account the possibility of selection near the boundary resulting from endogenous tower location. Specifically, I obtain geocoded data for all cell phone towers in Afghanistan in the year 2016 from the Afghan Telecommunications Regulatory Authority (ATRA). I then use the 2009 coverage maps to determine which towers were installed post-2009 in what were dead zones at the time. I then

Table 3—Averaged Boundary Treatment Effects

	All regions (1)	Southeast region (2)	Northwest region (3)
Panel A. At least one station with Category C fraud			
Inside coverage	−0.057	−0.099	−0.004
	(0.026)	(0.027)	(0.010)
	[−0.114, −0.033]	[−0.147, −0.042]	[−0.022, 0.016]
Boundary points	1,435	640	795
Mean outside coverage	0.183	0.311	0.035
Panel B. Share of votes under Category C fraud			
Inside coverage	−0.031	−0.055	−0.001
	(0.012)	(0.017)	(0.005)
	[−0.056, −0.008]	[−0.091, −0.027]	[−0.008, 0.009]
Boundary points	1,435	640	795
Mean outside coverage	0.124	0.221	0.018

Notes: Results use equations (3) and (4). Refer to Section IIB for description of estimation. Bootstrapped standard errors are in parentheses. Bootstrapped 95 percent confidence intervals are in brackets. Standard errors determined using 500 replications and resampling within districts. Boundary points refers to the total $\mathbf{b}_i$ evaluated within the specified region.

create coverage boundaries around this subset of towers using a 10 km radius around each tower's location.[30]

I perform an RD differences design that proceeds in three steps: First, it replicates the RD results using equation (1) with the post-2009 boundary, omitting from the estimation polling centers that are covered in 2009. Generally speaking, this exercise estimates the selection bias, i.e., the magnitude of any boundary effect on fraud that is solely due to endogenous tower selection, since coverage has not reached these areas in 2009 but will eventually. Second, I reestimate equation (1) using the actual 2009 coverage boundary and omitting the polling centers in the previous step. I then difference out the RD coefficient obtained in the first step from the RD coefficient obtained in the second step. This nets out the portion of the effect that is attributed to precoverage settlement characteristics that may lead to endogenous tower selection by providers at the boundary. Standard errors for this RD differences estimator are obtained using bootstrapping methods with resampling done within the same boundary neighborhoods used in previous results.

Columns 1 and 4 in Table 4 present the estimated RD coefficients that use the actual 2009 coverage boundary (step 2 above). Columns 2 and 5 present the RD coefficients that use the boundary of towers installed post-2009. "Inside coverage," in this case, equals one for polling centers that are inside the coverage area of towers that did not exist in 2009 but were eventually installed. Note that differences in the

[30] Refer to online Appendix Figure B7 for the location of these towers along with the 2009 coverage raster. The choice of 10 km for the coverage boundary follows Shapiro and Weidmann (2015). Shapiro and Weidmann (2015) uses a short and long radius of 4 and 12 km, respectively, depending on urban versus rural locations. I use a value within this range (10 km) for two reasons: (i) Given that the locations of post-2009 towers are mainly rural (note in online Appendix Figure B7 that these towers are well beyond the main cities along Afghanistan's ring road), I use a radius that is closer to the long radius in Shapiro and Weidmann (2015). (ii) However, given the mountainous Afghan terrain, I use a radius that is slightly smaller than the long radius used by Shapiro and Weidmann (2015) in Iraq.

Table 4—RD Differences Design: Effect of Cell Phone Coverage on Category C Fraud

	At least one station with Category C fraud			Share of votes under Category C fraud		
	RD_{2009}	$RD_{Post2009}$	Difference	RD_{2009}	$RD_{Post2009}$	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Inside coverage	−0.077	0.014	−0.091	−0.039	−0.006	−0.032
	(0.032)	(0.054)	(0.051)	(0.021)	(0.016)	(0.028)
			[−0.194, 0.006]			[−0.085, 0.026]
Observations	1,074	275		1,064	270	
Bandwidth (km)	7.278	7.278		7.152	7.152	
Neighborhoods	229	36		227	36	

Notes: Refer to Section IID for a description of the empirical method used. RD_{2009} refers to a regression discontinuity design using the 2009 coverage boundary. $RD_{Post2009}$ refers to a regression discontinuity design using the coverage boundary defined by towers installed post-2009. Bootstrapped standard errors are in parentheses. Bootstrapped 95 percent confidence intervals are in brackets. Standard errors determined using 250 replications and resampling within neighborhoods. Optimal bandwidth for each regression chosen as in Calonico, Cattaneo, and Titiunik (2014). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to Section IIC for a description of how boundary neighborhoods are created.

likelihood and intensity of fraud across covered and comparison centers are negligible and statistically insignificant. This provides strong indication that precoverage settlement characteristics near the coverage boundary cannot explain much of the observed drop in fraud. Columns 3 and 6 present the difference in RD coefficients along with the bootstrapped standard errors and 95 percent confidence intervals. The estimated differences indicate that after differencing out the potential bias due to selection, the likelihood of fraud is still significantly lower in polling centers within coverage areas.

These results suggest that, although providers may choose a tower's location strategically, at the coverage boundary, potential correlates of fraud such as settlement characteristics do not appear to change fraud levels discontinuously. Overall, this shows that endogenous tower selection cannot account for the observed sharp drops in fraud at the coverage boundary.

Mobile Coverage Spillovers and Spatial Displacement of Fraud.—We assess the possibility of coverage leading to positive spillovers or a reallocation of fraud toward nearby noncoverage areas. Spillovers can result from a coverage boundary that, using RD terminology, is not sharp. This would be the case if polling centers close to the boundary on the noncoverage side can benefit from coverage (e.g., voters are able to report fraud by walking to a nearby coverage area). If that is the case, then coverage can deter fraud even in noncoverage areas near the boundary. Overall, this would potentially lead to a downward bias in the main results presented in Table 2.[31]

[31] In principle, these polling centers would be noncompliers, and thus a fuzzy RD design would be the most appropriate approach. However, the lack of data on whether specific centers near the boundary received some sort of positive coverage spillover does not permit to follow such strategy. Using the example of voters in centers near coverage areas walking to covered areas to report fraud, a fuzzy RD design in this case would entail knowing the centers where this occurred and then estimating a joint model where the probability that a center is treated (i.e., individuals are able to report) modeled as a function of the observed coverage status.

In the case of spatial fraud reallocation, one possibility is that corrupt candidates respond strategically to coverage by reallocating fraud to polling centers just outside coverage areas. This in itself is evidence of the deterring effect of coverage. However, it may be problematic from a policy perspective if the objective is to deter overall fraud levels and coverage is simply leading to a spatial displacement of fraud. Lastly, we note that even in the presence of coverage spillovers and fraud reallocation, the main results presented in Table 2 are still valid estimates of the intent-to-treat effect.

To address these concerns, we compare fraud levels across 2 km bands around the boundary on the noncoverage side.[32] Broadly speaking, if fraud is being displaced to centers just outside coverage, we should detect a spike in fraud for bands near the coverage boundary relative to bands farther away. Similarly, if there are positive spillovers, then we should find a sharp drop in fraud levels in bands near the coverage boundary. Specifically, we estimate the following model:

$$v_{f,ij} = \gamma + \mathbf{D}'_{ij}\boldsymbol{\beta} + \Omega_i + \epsilon_{ij}, \tag{6}$$

where $\mathbf{D}_{ij} = \left(D_{ij}^{2km}, D_{ij}^{4km}, D_{ij}^{6km}, D_{ij}^{8km}\right)'$ is a 4×1 vector containing four dummy variables that classify polling centers into four mutually exclusive distance bands around the coverage boundary on the noncoverage side.[33] To ensure comparisons of reasonably similar locations, we restrict the overall window of analysis to 10 km at most. Therefore, polling centers beyond the band used in the analysis but within 10 km are used as the omitted comparison category. As before, Ω_i is a boundary neighborhood fixed effect. We interpret estimates of the coefficient vector $\boldsymbol{\beta}$. In the absence of any response, we should expect these coefficients to be close to zero, suggesting that fraud in near-coverage bands did not significantly change relative to comparison bands. In the case of fraud moving away from coverage and into nearby areas, we should expect these coefficients to be positive. Lastly, if there are positive coverage spillovers in the form of fraud deterrence even in the absence of coverage, then we should expect these coefficients to be negative for the most part.

Table 5 presents the estimates of equation (6). Each column of Table 5 adds a contiguous band to the analysis to see if one can pinpoint a specific band where there is a spike or a significant drop in the likelihood of fraud (columns 1–4) or in fraud levels (columns 5–8).[34] For all bands, we find no evidence of either a significant spike or drop in our fraud measures relative to areas just outside the band of analysis. Furthermore, we perform a test for joint significance of $\boldsymbol{\beta}$, and for all specifications, we cannot reject the hypothesis that the difference in fraud between near-coverage bands and comparison bands is zero.

These results are contextually plausible. First, the fact that individuals are near the coverage boundary does not necessarily imply that they have easy access to the technology so as to "walk to coverage" and report fraud. For instance, in a fragile

[32] The choice of 2 km is to allow for a reasonable number of observations within each band.

[33] Specifically, the categories capture whether in neighborhood j, center i is within 2 km of the coverage boundary (D_{ij}^{2km}), within 4 km—but beyond 2 km—from the coverage boundary (D_{ij}^{4km}), and so on.

[34] Columns 1 and 5, for instance, compare our fraud measures for polling centers within 2 km of the boundary with polling centers in the next band over (within 4 km but beyond 2 km).

Table 5—Cell Phone Coverage Spillovers and the Spatial Reallocation of Fraud

	At least one station with Category C fraud				Share of votes under Category C fraud			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_{2km}	−0.064	0.024	−0.016	−0.012	−0.097	−0.044	−0.049	−0.028
	(0.085)	(0.097)	(0.076)	(0.048)	(0.073)	(0.079)	(0.048)	(0.042)
D_{4km}		0.090	0.036	0.038		0.062	0.043	0.064
		(0.122)	(0.113)	(0.045)		(0.087)	(0.073)	(0.048)
D_{6km}			−0.035	−0.029			−0.012	0.010
			(0.091)	(0.081)			(0.065)	(0.044)
D_{8km}				0.007				0.021
				(0.081)				(0.038)
Observations	366	400	429	453	366	400	429	453
p-value (F-test)	0.45	0.70	0.91	0.90	0.18	0.32	0.49	0.22
Bandwidth (km)	4	6	8	10	4	6	8	10
Neighborhoods	178	185	191	200	178	185	191	200

Notes: Sample restricted to polling centers on noncoverage areas. D_{2km} equals 1 if polling center is between 0 and 2 km from the coverage boundary. D_{4km} equals 1 if polling center is between 2 and 4 km from the coverage boundary. D_{6km} equals 1 if polling center is between 4 and 6 km from the coverage boundary. D_{8km} equals 1 if polling center is between 6 and 8 km from the coverage boundary.

security environment, individuals in noncoverage areas may delay investing in the technology (relatively expensive phone, SIM card, calling plan, etc.) if there is uncertainty on whether violence will hinder the introduction of coverage to their village. This is supported by a growing literature highlighting the negative impact of violence on investment and financial decisions (e.g., Besley, Mueller, and Singh 2011; León 2012; Singh 2013; Blumenstock et al. 2019). Even in the absence of violence, there might be little incentive for individuals to invest in a technology that relies heavily on a network effect if there is no corresponding network within the individual's own village or district. Second, although displacement of fraud away from coverage might take place, this is more likely to occur in district centers where coverage status is common knowledge and displacement coordination between candidate and polling officials is potentially easier. However, at the margin of coverage away from district centers, it is unlikely that corrupt candidates have precise knowledge on whether a certain polling center has coverage or not.

Other Mobile Service Providers.—The results in this paper use coverage data for two of the largest operators in Afghanistan, MTN and Afghan Wireless (AWCC).[35] The lack of data on other providers can introduce measurement error in our coverage indicator if some areas classified as not covered are actually covered by a missing provider. If that is the case, then we are potentially underestimating the effect of coverage. In this section, we gauge the sensitivity of our main results to the inclusion of a third mobile service provider, Roshan, which accounted for about

[35] Together they encompassed about 46 percent of all cell phone subscriptions in Afghanistan around the time of the 2009 election (Hamdard 2012)

32 percent of cell phone subscriptions around this time. To do this, we obtain data on the location of Roshan towers up to the year of study, 2009.[36]

Appendix Figure A1 provides the tower footprint for Roshan Telecom in the year 2009 overlaid on the coverage map for MTN and AWCC. In all, Roshan had 812 towers within Afghanistan. Of those, only 4 percent (31 towers) were outside of our existing coverage map, suggesting that the degree of mismeasurement in our coverage variable is likely low. These towers were located in 25 districts containing 450 polling centers (about 7 percent of all the polling centers in our sample). Of those 25 districts, there were only 3 districts (containing 47 polling centers) where Roshan was the sole provider of coverage in the district.

We present results using three strategies. First, we replicate the main analysis excluding the three districts were Roshan was the sole provider of coverage. Second, we replicate the analysis excluding any district where Roshan has at least one tower outside of MTN and AWCC coverage but it is not necessarily the sole provider. This is a stricter exclusion rule, as it also drops polling centers that are rightly classified as covered by MTN and AWCC; however, it ensures that the remaining sample does not have measurement error from potential Roshan coverage. Third, I impute Roshan's coverage using its tower footprint and combine it with the existing coverage measure to create a single coverage map that includes the three companies. Estimation of Roshan's coverage follows three steps: (i) create a 1×1 km grid of Afghanistan (this roughly matches the grid used by the GSMA to report coverage for the other providers), (ii) assign grid cells within a 10 km radius of each Roshan tower as covered by Roshan, and (iii) merge this grid with the existing coverage map.[37] Appendix Figure A2 presents a sample of this grid with steps (ii) and (iii) and the resulting coverage map.

The last strategy provides a more complete measure of coverage; however, some points are important to note. Although the imputation of coverage for Roshan using the radius follows a conventional strategy (e.g., Shapiro and Weidmann 2015, Jensen 2007), it is not as refined as the coverage measure of MTN and AWCC used in our main analysis. The latter employs proprietary models used by the companies that account for tower characteristics (e.g., transmission power, signal frequency), terrain details, and on-the-ground validations. Therefore, although more inclusive, there is potential for measurement error as well. For this reason, we present results using all strategies in order to gauge the consistency of our findings to different choices.

Table 6 replicates the main results presented in columns 1 and 2 of Table 2 using the two exclusion rules (columns 1–4) and the imputation of Roshan coverage (columns 5 and 6). In columns 1 and 2, we see that our main results do not change significantly when excluding districts where Roshan is the sole provider of services. This is expected, considering that most polling centers in the excluded districts are not in the main analysis in Table 2, since those districts do not have any coverage from MTN or AWCC. Columns 3 and 4 use the strictest restriction possible: excluding

[36] I am deeply grateful to Tarek Ghani for providing original data on the location of Roshan towers. The remaining companies, Etisalat, Afghan Telecom, and Wasel Telecom, account for about 22 percent of the share of subscribers (Hamdard 2012).

[37] Refer to Section IID for a detailed explanation of the choice of 10 km for the coverage radius.

Table 6—Effect of Cell Phone Coverage on Category C Fraud: Roshan Coverage

	Exclusion: Roshan-only districts		Exclusion: At least one Roshan tower in district		Imputing Roshan coverage	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. At least one station with Category C fraud						
Inside coverage	−0.076	−0.082	−0.057	−0.079	−0.071	−0.080
	(0.032)	(0.029)	(0.032)	(0.029)	(0.031)	(0.032)
Observations	1,073	2,014	911	1,808	1,195	2,039
Mean outside coverage	0.183	0.145	0.138	0.115	0.133	0.113
Bandwidth (km)	7.27	—	7.00	—	10.58	—
Neighborhoods	230	237	209	226	233	237
RD type	LLR	Poly	LLR	Poly	LLR	Poly
Panel B. Share of votes under Category C fraud						
Inside coverage	−0.037	−0.041	−0.022	−0.040	−0.044	−0.047
	(0.021)	(0.020)	(0.022)	(0.021)	(0.025)	(0.026)
Observations	1,062	2,014	867	1,808	1,314	2,039
Mean outside coverage	0.124	0.096	0.091	0.075	0.088	0.074
Bandwidth (km)	7.13	—	6.33	—	12.62	—
Neighborhoods	228	237	206	226	236	237
RD type	Optimal	Poly	Optimal	Poly	Optimal	Poly

Notes: Results use equation (1). Refer to Section IIB for a description. Optimal in RD type refers to specification using optimal bandwidth. Optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik (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, Galdo, and Smith (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to Section IIC for a description of how boundary neighborhoods are created. Columns 1 and 2 exclude polling centers in districts where Roshan is the sole provider. Columns 3 and 4 exclude polling centers in districts where there is at least one Roshan tower outside of existing coverage. Columns 5 and 6 use a radius of 10 km for Roshan towers and combine this with existing coverage. Refer to Section IID for more information.

all districts if at least one Roshan tower is outside of the coverage boundary used in the main analysis. Compared to the main results, the estimated RD coefficients are slightly smaller but statistically significant and qualitatively similar to our original results. Lastly, columns 5 and 6 replicate the results after adding the imputed Roshan coverage. The results are qualitatively similar to our main results, with the estimated coefficients in panel B being slightly larger than the original ones.

Note that in principle, the omission of a key service provider points to underestimation of the coverage effect in our main results. However, note that in practice, the results in Table 6 will not necessarily yield an RD estimate that is higher in magnitude, since the exclusion rules either undercorrect the issue by leaving mismeasured observations in the analysis (columns 1 and 2) or overcorrect by dropping accurately measured observations from the analysis (columns 3 and 4). Overall, this section shows that correcting for the omission of an important service provider does not significantly change the interpretation of our results.

Cell Tower Shutdowns.—It is well documented that throughout the eastern and southern regions of the country, the Taliban force cell phone companies to regularly turn off their antennas at dusk to prevent civilians from sharing information on their operations with coalition forces (e.g., Trofimov 2010), while attacks to damage and destroy cell phone towers are not uncommon (e.g., Lakshmanan 2010, Robinson

2013). This is concerning for our empirical strategy if we misclassify some polling centers as being covered when their service tower may, in practice, be shut down by the Taliban. This issue is particularly concerning if the shutdowns occurred around election time or if they systematically occur during the daytime (time during which citizens are likely involved in monitoring/information gathering around election time).

To the best of our knowledge, there is no official dataset documenting tower shutdowns. To address this problem, I compile all available media reports related to tower shutdowns in Afghanistan from both international and local news agencies.[38] In all, I find 41 articles between February 2008 and December 2019 reporting tower shutdowns in Afghanistan; 31 articles specifically reference the timing of the shutdowns. Of these, 29—or about 94 percent—explicitly mention that the shutdowns took place during nighttime, while only 2 articles reference a daytime or complete shutdown.[39] When considering articles up to 2009, and hence up to the election year, 10 out of the 11 articles reference nighttime shutdowns. Only one referenced a daytime shutdown that only lasted ten days (Shalizi 2008). With respect to the reported location of shutdowns, 34 articles report specific locations; 9 articles (about 26 percent) make reference to shutdowns in Helmand province, 3 reference Kandahar (9 percent), 5 reference Ghazni (15 percent), and 8 simply reference southern Afghanistan, while the remaining 9 articles reference other provinces. Out of the 11 articles published up to 2009, 7 have reference to shutdown locations, with Helmand, Kandahar, and Ghazni being the three provinces specifically mentioned.[40]

Although one should be mindful that articles are a selective account of shutdowns, the information obtained points to a clear pattern around the time of the election: shutdowns take place mainly during nighttime and are concentrated in the southern provinces. This boosts our confidence that the shutdowns do not have a significant impact on our main results, as nighttime shutdowns should not affect coverage during regular daytime election operations. Additional details from the articles reinforce this. First, we could not find a single report of shutdowns related to the 2009 election. Second, the shutdowns do not seem to be a permanent, year-round tactic of the Taliban. In fact, the vast majority of reported shutdowns are concentrated in or before March of every year, and very few occur around the month of the election. This suggests that the shutdowns indicate intensifying preparations for the spring offensive (Associated Press 2011).[41]

In order to more formally assess the impact of shutdowns, I replicate the main results in columns 1 and 2 of Table 2 using two exclusion rules: (i) excluding all polling centers from Helmand, Kandahar, and Ghazni, as these were the provinces mentioned

[38] The search parameters used to find the articles using Google are *allintext:taliban phone cell tower shutdown* or blackout* and *allintext:taliban phone cell tower*. For the local news agencies, I focus on TOLOnews and the Pajhwok network. Database collected and prepared by author (Gonzalez 2019). Database can be found on the author's website.

[39] Most articles are specific about the timing: the shutdowns typically start at around 5 PM and end at around 7 AM. In the case of the daytime shutdown, it only lasted 10 days (Shalizi 2008), while the complete shutdown was quite unprecedented (Sameem 2011).

[40] Four articles simply reference southern Afghanistan. Refer to online Appendix Table B14 for more summary statistics on the timing and location of the shutdowns reported in the articles.

[41] Of the 11 articles reporting shutdowns before the election year, 8 report shutdowns in or before March.

TABLE 7—EFFECT OF CELL PHONE COVERAGE ON CATEGORY C FRAUD: TOWER SHUTDOWNS

	Exclusion: Helmand, Kandahar, Ghazni		Exclusion: high violence districts	
	(1)	(2)	(3)	(4)
Panel A. At least one station with Category C fraud				
Inside coverage	−0.069	−0.066	−0.085	−0.083
	(0.028)	(0.029)	(0.033)	(0.029)
Observations	1,238	1,827	1,048	2,015
Mean outside coverage	0.160	0.133	0.180	0.140
Bandwidth (km)	12.82	—	7.14	—
Neighborhoods	214	215	227	236
RD type	LLR	Poly	LLR	Poly
Panel B. Share of votes under Category C fraud				
Inside coverage	−0.022	−0.026	−0.039	−0.042
	(0.019)	(0.020)	(0.022)	(0.021)
Observations	1,007	1,827	1,056	2,015
Mean outside coverage	0.115	0.087	0.125	0.092
Bandwidth (km)	7.44	—	7.23	—
Neighborhoods	208	215	229	236
RD type	Optimal	Poly	Optimal	Poly

Notes: Results use equation (1). Refer to Section IIB for a description. Optimal in RD type refers to specification using optimal bandwidth. Optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik (2014). Columns 2, 4, and 6 use a third-degree polynomial in distance to boundary. Polynomial order is determined using Akaike's criterion as suggested in Black, Galdo, and Smith (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to Section IIC for a description of how boundary neighborhoods are created. Columns 1 and 2 exclude polling centers in Helmand, Kandahar, and Ghazni provinces. Columns 3 and 4 exclude districts where insurgent violence in the year of the election and up to election day was above the ninety-fifth percentile. Refer to Section IID for more details and to Section IIIB for details on violence data.

in the preelection articles as suffering heavily from shutdowns; and (ii) excluding all polling centers in districts with significantly high levels of insurgent violence during the election year (above the ninety-fifth percentile).[42] Given that shutdowns are a key operational tactic of the Taliban, high violence levels are likely a good indicator for areas heavily affected by shutdowns. Table 7 presents these results. After the exclusion of affected provinces (columns 1 and 2) and high violence districts (columns 3 and 4), we find that our main results do not change significantly. This is consistent with the pattern of shutdowns likely having an effect on nighttime civilian activities but not on daytime operations by civilians such as voting during the election.

Signal Strength Fuzziness.—Day-to-day and even within-day variations in factors such as temperature, precipitation, and conductivity of the terrain might cause signal strength near the coverage boundary to become fuzzy or vary slightly. Therefore, one should consider the boundary used throughout the paper as the boundary that is present under typical conditions of the area. Note that results using a sharp RD design in the paper are not necessarily affected by the potential fuzziness of the boundary. However, they should be interpreted as an intent-to-treat effect.

[42] Refer to Section IIIB for a detailed description of the insurgent violence data.

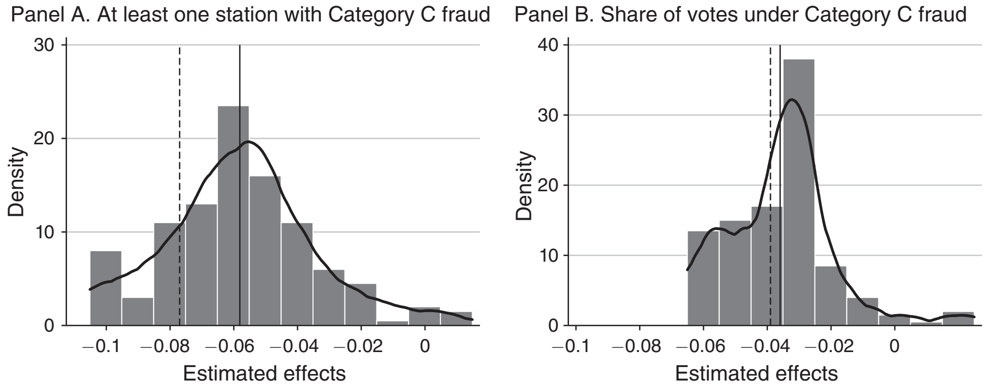

Figure 6. Effect of Coverage on Fraud Using Random Shifts of Boundary

Notes: Histogram of estimated RD coefficients using 200 randomly selected coverage boundaries within a 1 km bandwidth of the coverage boundary used in the main analysis. Coefficients obtained from estimating equation (3). Solid line gives the average of the 200 RD coefficients obtained from each randomly drawn boundary. Dashed line gives the estimated coefficient in the main analysis (column 2 of Table 2). Refer to Section IID for more details.

This section explores whether the results are sensitive to slight changes in the location of the boundary. In practice, one cannot implement a fuzzy RD design in this setting, because that would entail knowing exactly how much the boundary varied on election day. Instead, I replicate the main results using random shifts in the boundary within a 1 km bandwidth of the reported boundary. Specifically, I do 100 random shifts inward (contraction in coverage) and 100 random shifts outward (expansion in coverage). This mimics day-to-day arbitrary variations in the actual boundary. Figure 6 plots a histogram of the RD coefficients using the shifted boundaries. Note that the results are not very sensitive to small shifts in the boundary. The averages of the estimated RD coefficients are around -0.06 and -0.03 when "At least one station with Category C fraud" and the "Share of votes under Category C fraud" are the outcome variables, respectively.[43]

Falsification Tests.—We gauge the likelihood that the coverage effect simply reflects regular fluctuations in fraud levels across space. For this, we randomly choose 100 longitudes from the range of longitudes fully covering Afghanistan and consider each of the north-south boundaries formed by these longitudes as a boundary. We then estimate equation (1) for each of these false boundaries. We then repeat the exercise using 100 randomly chosen latitudes and the east-west boundaries formed by each of these. We omit polling centers in covered areas from both analyses to ensure that we are not comparing centers that are truly covered.

If sharp differences in fraud levels across neighboring polling centers are a common occurrence regardless of coverage, then the distribution of RD coefficients should be centered at a value significantly different from zero. Appendix Figure A3 presents the distribution of estimated coefficients for each of the random longitudes

[43] The corresponding coefficients in the main results are -0.077 and -0.035.

(panels A and B) and latitudes (panels C and D) used as false boundaries. Note that for both fraud measures and both exercises, the estimated coefficients are bunched around zero. Also note that zero is well within the 95 percent confidence interval (dashed vertical lines) of the mean of the estimated coefficients (solid vertical line).

III. Coverage-Fraud Channels

This section explores channels that can potentially explain the relationship between coverage and election fraud documented in Section IIC. To provide an analysis that is both parsimonious and enlightening, the section presents a theoretical model that defines three mechanisms by which coverage can affect fraud: social monitoring, election-related violence, and candidate-voter affinity. The section then proceeds to test these mechanisms empirically.

A. *Theoretical Model*

This model illustrates the link between cell phone coverage and electoral fraud in a fragile security environment. It considers the problem of a candidate determining the purchase of fraudulent and legal votes at a polling center. In the case of fraudulent votes, I follow Callen and Long (2014) by assuming that the candidate pays the price of fraudulent votes to an election official in charge of a polling center. This price takes into account the possibility that the center is audited as a result of complaints received from individuals at the polling center. The price of legal votes takes into account the risk of violence that the voters face at the polling place. Higher levels of expected violence require a higher price.

The Election Official's Problem.—The candidate purchases fraudulent votes from an election official overseeing polling center j. The price of these votes must be high enough that the official's incentive compatibility constraint binds. This price takes into account the probability that the center is audited as a result of reports by ordinary individuals (i.e., social monitoring). Assume the candidate and the official expect the number of submitted fraud complaints r to follow a random process with probability function $H(r;n,\rho)$ that takes ρ and n as parameters. The parameter n is the number of voters at the center, and ρ gives the probability that an individual voter files a complaint.[44] Furthermore, assume that the center is audited if the number of reports r exceeds a predetermined threshold $\bar{r}$ so that the probability that the center is audited is given by $\pi(\bar{r};n,\rho) = 1 - H(\bar{r};n,\rho)$.

Letting v_f, p_f, and F denote the number of fraudulent votes, their price, and a marginal fine, respectively, I assume that the official faces a lottery in which he expects to be caught with probability π, receive fraud revenues $p_f v_f$ net of a total fine $v_f F$, or succeed with complementary probability $1 - \pi$ and pocket all fraud

[44] Refer to online Appendix C for a derivation of ρ using the optimization problem of an individual deciding whether to report fraud.

revenues instead.[45] Assuming that the official is an expected income maximizer, then the minimum price per fraudulent vote that guarantees compliance is given by the expression

$$\pi\left(p_f v_f - v_f F\right) + \left(1 - \pi\right) p_f v_f = 0,$$

$$p_f = \pi F, \tag{7}$$

where it is assumed that the official receives an offer from only one of the candidates (i.e., payoff from noncompliance is zero).[46]

The Candidate's Problem.—The candidate must decide how many votes, 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, the candidate is penalized by losing all fraudulent votes and receiving only legal votes v_l at center j. If the center is not audited, the candidate simply keeps all votes $v_l + v_f$. I consider the price of legal votes v_l to be a function of a parameter a that characterizes each village's affinity toward the candidate. For simplicity, assume that this parameter captures ethnic or tribal similarities between a candidate and the village. Villages where the candidate and voters share a similar ethnic background require a lower legal price per vote to entice individuals to vote. Further, since elections in conflict zones are often characterized by political violence, I also consider the price of legal votes to be a function of an exogenous probability δ that a violent event takes place at polling center j and as a result the individual 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). With this in mind, I define the price of legal votes as $p_l = f\left(\delta, P, a\right)$, with $\partial f(\cdot)/\partial \delta > 0$, $\partial f(\cdot)/\partial P > 0$, and $\partial f(\cdot)/\partial a < 0$.[47]

Given the assessed probability of an audit, π, and assuming that the candidate has quasilinear preferences over votes, then the maximization problem of the candidate is given by

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

subject to

$$p_f v_f + p_l v_l \leq E,$$

[45] This assumes 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). The candidate is penalized by losing the fraudulent votes, as shown in Section IIIA.

[46] I rely on this assumption to simplify the analysis but also because the pattern observed in the data suggests that most fraud took place in the southern and eastern parts of the country that form the backbone of Karzai's coalition. This suggests that fraud offers to polling center officials were likely coming from a single candidate.

[47] Refer to online Appendix C for an extension of the model that derives an expression for the legal price of votes.

where fraudulent votes enter nonlinearly (with $\alpha \leq 1$) to capture the possibility that fraudulent and legal votes are not perfect substitutes and E is the endowment of the candidate.[48] The solution to the problem above provides an optimal relationship between fraudulent votes and their price p_f. Substituting the expressions for prices p_f and p_l in order to obtain the equilibrium level of fraud gives

$$(8) \qquad v_f^* = \left[\alpha \cdot \underbrace{\frac{1-\pi}{\pi}}_{\substack{\text{Social}\\ \text{monitoring}\\ \text{effect}}} \cdot \underbrace{\frac{1}{F}}_{\substack{\text{Penalty}\\ \text{effect}}} \cdot f(\underbrace{\delta, P}_{\substack{\text{Violence}\\ \text{effect}}}, \underbrace{a}_{\substack{\text{Candidate}\\ \text{affinity}\\ \text{effect}}})\right]^{\frac{1}{1-\alpha}},$$

where I separate the expression into the different determinants of fraud in equilibrium. I highlight four key results regarding the relationship between coverage and fraud at equilibrium.

First, the *social monitoring effect* illustrates that increasing election monitoring (π) leads to an unambiguous decrease in fraud levels at equilibrium. In the presence of a monitoring mechanism that relies on cell phone access, it is clear that coverage can increase election monitoring via citizen participation and enfranchisement in general.

Second, higher candidate-voter affinity via ethnic and tribal links lowers fraud as the price of legal votes is reduced. Coverage is unlikely to have an effect on tribal affinity. However, tribal affinity can confound the effect of coverage on fraud if spatial changes in coverage coincide with changes in tribal composition. This may result, for instance, if cell phone providers give preference to certain ethnic groups by expanding coverage into their areas.[49]

Third, the *violence effect* suggests that an increase in the likelihood or magnitude of violence (δ and P, respectively) increases the price of legal votes and as a result increases fraud by making fraudulent votes less expensive relative to legal votes. Coverage can affect fraud through this channel, as coverage may allow insurgents to better coordinate attacks (e.g., Cordesman 2005, Leahy 2005) or use mobile devices to detonate IEDs (e.g., Nolin 2011). Alternatively, coverage can facilitate collective action by citizens and cell phone tracking by counterinsurgency agencies to undermine political violence. Notice that violence can lead to fraud even in polling centers where the candidate is liked, since the price of legal votes may become prohibitively high and thus lead to fraud in order to offset the expected low turnout in "high a" areas.[50]

[48] The quasilinear specification deviates from Callen and Long's (2014) perfect substitutes specification. The appeal of the quasilinear specification is that it avoids a prediction where the candidate simply substitutes to all fraudulent or all legal votes as soon as the relative price deviates from one. The studied sample shows a combination of fraudulent and legal votes for the most part, not corner solutions like the ones obtained from a perfect substitutes specification. For this reason, I also consider the interior solutions only.

[49] Although not included in the model, candidate affinity can also lead to more fraud if candidates and polling center managers of similar ethnic background can better coordinate to engage in fraud. In the context of this paper, this would matter if affinity-driven coordination depends on coverage. Section IIIE addresses this issue.

[50] This is a key result considering that fraud was widespread in areas where Karzai had strong support, which were also the areas with high levels of violence. Furthermore, this can also explain the pattern of fraud observed in the sample. Since the southeast experienced higher levels of violence relative to the northwest, our model would predict that fraud will be more prevalent in the southeast since the candidate has to substitute the lower turnout (due to the higher price legal votes) with fraudulent votes. This is actually consistent with the findings in Weidmann

Fourth, the *penalty effect* shows that fines lower fraud. However, it is unlikely that this depends on coverage, as penalties are a nationwide measure. The following sections proceed to empirically test these mechanisms.

B. *Election-Related Insurgent Violence*

Findings from the literature on conflict and violence suggest a strong relationship between political violence and both cell phone coverage (Shapiro and Weidmann 2015, Pierskalla and Hollenbach 2013) and electoral fraud (Collier and Vicente 2012, Weidmann and Callen 2013). Cell phone coverage may lead to surges in violence, as insurgents can better coordinate attacks (Cordesman 2005, Leahy 2005) or use mobile devices to detonate IEDs (e.g., Nolin 2011). In contrast, collective action by citizens and cell phone tracking by counterinsurgency agencies might undermine the insurgents' actions.

In the case of Afghanistan, political violence is a potentially important channel for two reasons: First, during the preelection period, the Taliban issued several warnings targeting polling centers and voters (Gall 2009, Filkins 2009), while recent research has shown a clear link between insurgent activity and election participation (Condra et al. 2018a). This was followed by a sharp surge in violence on election day as depicted in Appendix Figure A4, which plots the number of daily attacks for the year 2009 and for various attack types. Violence was not directed at a specific candidate but rather at the election process in general. Secondly, media reports suggest that the Taliban have a strong aversion to cell phone coverage and cell phone technology in general. For example, throughout the eastern and southern regions of the country, they have forced cell phone companies to regularly turn off their antennas at dusk to prevent villagers from informing coalition forces of their movements (Trofimov 2010). Attacks to damage and destroy cell phone towers are also well documented (e.g., Lakshmanan 2010, Robinson 2013).

In terms of the theoretical model, if a candidate expects a drop in violence due to coverage (i.e., a drop in δ or P due to the Taliban preferring to operate in areas without coverage), then the price of legal votes $p_l = f(\delta, P, a)$ in covered centers drops relative to fraudulent votes, which then leads the candidate to substitute fraudulent votes for legal votes in coverage areas. Therefore, declines in violence with coverage can actually explain the drops in fraud documented in Section IIC.

I proceed by estimating both scalar RD estimates using equation (1) and boundary treatment effects using equations (3) and (4) for various outcomes on insurgent violence and using the same boundary points $\mathbf{b}_i$ used in the fraud results. The objective of the scalar RD is to assess whether, on average, insurgent violence drops with coverage, while the goal of the boundary RD is to assess whether violence changes discontinuously at boundary points where fraud also drops.

The analysis uses data on the location of polling centers and the type of insurgent attack beginning on January 1, 2009, up to the day of the election on August 20, 2009. Insurgent attacks data are obtained from time-stamped and georeferenced

and Callen (2013) documenting a positive correlation between violence levels and fraud as long as violence is not extremely high, at which point fraud or any election-related operations are not feasible.

records collected by the International Security Assistance Forces (ISAF) and Afghan forces in Afghanistan (Condra et al. 2018b). I define three main outcomes: the total number of attacks, the number of IEDs, and the number of direct fire attacks within a 1 km radius of polling centers.[51] The choice of location and the timing of attacks follows the intuition of the theoretical model. Specifically, the definitions use the location of polling centers and data up to the election day in order to capture individuals' expectations on where violence would potentially occur on that day. On average, about 21 percent of all polling centers experienced at least 1 attack up to election day. The rate is much higher in the southern and eastern provinces, where the Taliban have a stronger presence.[52]

Table 8 presents the scalar RD results. Polling centers just within coverage areas experience about 0.7 more incidents on average (panel A, column 1), about 0.3 more IEDs (panel B, column 1), and about 0.4 more direct fire incidents (panel C, column 1). However, the results are not statistically significant. As expected, the magnitude of the effect is mainly driven by results from the southeast region, where insurgency levels at baseline are much higher. However, as before, the results are not indistinguishable from zero. Appendix Figure A5 presents the distribution of the estimated boundary treatment effects for all three violence outcomes. For the most part, the boundary treatment effects are clustered around zero. In the case of the southeastern provinces, 87 percent, 89 percent, and 85 percent of the estimated boundary treatment effects for all attacks, IEDs, and direct fire incidents, respectively, are positive but statistically insignificant for the most part.[53] Similar to the scalar RD results presented earlier in this section, this provides evidence that insurgent violence does not seem to significantly respond to coverage. More importantly, these findings suggest that the drops in fraud at the boundary described in Section IIC cannot be explained by drops in insurgent violence.

C. *Social Monitoring and Citizen Enfranchisement*

This section explores whether a nationwide election monitoring mechanism that relied on cell phone access can explain the observed drops in fraud at the coverage boundary. Specifically, the 2009 Afghan election saw the creation of the UN-led Electoral Complaints Commission (ECC). Along with the task of investigating and adjudicating fraud-related complaints, the commission also instituted a citizen-monitoring initiative that created an election fraud hotline to facilitate fraud reporting. Specifically, the ECC provided two cell phone hotlines for people to report and find

[51] Condra et al. (2018b) provides a detailed description of this dataset. Condra et al. (2018a) employs the dataset within the Afghan setting. Direct fire refers to attacks using small arms and rocket-propelled grenades. The choice of 1 km radius is to avoid choosing a radius that is too large and would lead to counting incidents on the other side of the coverage boundary for polling centers that are very close to this boundary. Online Appendix Table B13 presents results using a larger 5 km radius, and the results are qualitatively similar. Similarly, the results are robust when excluding election day when constructing the violence outcome (online Appendix Table B11) or when using a longer time frame from January 2006 to election day in 2009 (online Appendix Table B12).

[52] Refer to online Appendix Table B8 for summary statistics on the violence outcome variables and online Appendix Figure B8 for a depiction of the differential rates of insurgent violence across Afghan regions.

[53] When looking at all three outcomes tested, only about 4 percent of the boundary treatment effects are statistically significant at the 10 percent level.

TABLE 8—EFFECT OF CELL PHONE COVERAGE ON NUMBER OF INSURGENT ATTACKS NEAR POLLING CENTERS

	All regions		Southeast region	
	Optimal bandwidth (1)	Polynomial in distance (2)	Optimal bandwidth (3)	Polynomial in distance (4)
Panel A. Number of violent incidents within 1 km radius of polling center				
Inside coverage	0.653	0.781	0.649	1.138
	(0.647)	(0.526)	(1.153)	(0.983)
Observations	1,300	2,118	687	1,163
Mean outside coverage	0.86	0.70	1.60	1.59
Bandwidth (km)	10.5	—	9.77	—
Neighborhoods	236	239	101	102
Panel B. Number of IEDs within 1 km radius of polling center				
Inside coverage	0.228	0.179	0.363	0.260
	(0.198)	(0.164)	(0.316)	(0.306)
Observations	861	2,118	526	1,163
Mean outside coverage	0.14	0.076	0.23	0.17
Bandwidth (km)	4.17	—	4.83	—
Neighborhoods	222	239	96	102
Panel C. Number of direct fire incidents within 1 km radius of polling center				
Inside coverage	0.355	0.423	0.401	0.608
	(0.339)	(0.284)	(0.594)	(0.532)
Observations	1,488	2,118	831	1,163
Mean outside coverage	0.47	0.39	0.90	0.88
Bandwidth (km)	13.9	—	13.8	—
Neighborhoods	238	239	102	102

Notes: Results use equation (1) and outcome variables denoted in panel names. Violence outcomes are measured between January 1, 2009 and August 20, 2009. Refer to Section IIIB for more details on the outcome variables used. Optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik (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, Galdo, and Smith (2007). All specifications use neighborhood fixed effects and standard errors clustered at the neighborhood level. Refer to Section IIC for a description of how boundary neighborhoods are created.

the information needed to file a claim (where to file it, how to file it, deadlines, etc.).[54] The hotlines were widely publicized through a public outreach program that included television and several radio advertisements in both Pashto and Dari. According to ECC guidelines, individuals could file election-related complaints within 72 hours of the election. Typical types of complaints included claims of bribery, intimidation, counting errors, and the theft and manipulation of electoral documents. Private organizations also encouraged the use of cell phones to report instances of fraud. For example, in the weeks prior to the election, Pajhwok News, a major independent news agency in the country, along with other international nongovernmental organizations, enabled several hotlines. In addition, the agency deployed around 80 reporters throughout the country who were instructed to use their mobile phones to text and call in incidents of violence and fraud (Himelfarb 2010). Similarly, cell phone access also facilitated the reporting of instances of violence, intimidation at the polling center, and corruption in

[54] To guarantee some degree of accountability, individuals filing complaints were required to provide their names and addresses to the ECC in case a follow-up investigation would take place. The ECC also guaranteed the anonymity of individuals. More information on the hotlines as well as the public outreach program can be found at the ECC's official website www.iecc.gov.af.

general to the 119 Afghan corruption hotline led by the European Union Police Mission in Afghanistan (EUPOL) and relatively well known by the Afghan population.[55]

The ECC's monitoring intervention led to a surprisingly high degree of citizen participation. According to ECC, the agency received more than 3,300 complaints. To put this number in perspective, there were close to 470 polling centers with reported irregularities, so that translates to a ratio of more than 7 fraud reports per compromised center (Electoral Complaints Commission 2010). The enfranchisement of citizens via the monitoring mechanism, along with the widespread availability of cell phones, suggests that the reported drops in fraud at the coverage boundary can be the result of candidates expecting increased accountability in coverage areas. In the context of the theoretical model, coverage increases the likelihood that a polling official gets caught engaging in election fraud (π). This, in turn, translates into a higher price for fraudulent votes relative to legal votes in order to offset the increased risk.

I proceed by providing results that suggest that coverage does indeed increase enfranchisement and citizen participation in election monitoring. I combine the coverage maps with polling-center-level data on all election-related complaints submitted to the ECC.[56] These data contain information on the date the complaint was submitted, the polling center in question, the type of complainant (e.g., individual, candidate, organization), the gender of the complainant, and a description of the allegation. There are two important caveats. First, the election complaints data come from the 2010 parliamentary election, which took place about a year after the election studied in this paper. The two elections, although held at different dates, were very similar in design: they both used the same polling centers and the same overseeing bodies, and, in fact, they were to be held on the same date to save costs, but disagreements among candidates led to the different dates (Faiez 2008). Second, the GSMA did not collect coverage maps for Afghanistan in the year 2010, and thus I rely on coverage data from 2009. This means that the coverage area used for this analysis is likely smaller than the actual coverage area in 2010. Given these shortcomings, the main results presented in this section do not use a regression discontinuity approach that, by design, relies heavily on observations close to the coverage boundary. Instead, I use an indicator for coverage as the key variable of interest and expand the bandwidth to include polling centers that are farther away from the boundary and hence more likely to be correctly classified as covered/uncovered. Furthermore, the model includes a rich set of covariates and neighborhood fixed effects to account for some of the unobserved heterogeneity across polling centers.[57]

Table 9 presents the results for three key outcomes.[58] Panel A shows considerable evidence that coverage increases citizen participation in election monitoring. The results show that the number of complaints per polling center is significantly higher

[55] According to a 2012 UNDP survey cited by the Ministry of Interior Affairs (MOIA), about 80–90 percent of the Afghan population has some familiarity with the police hotline. This information was obtained by the author through an interview with an MOIA representative.

[56] I am deeply grateful to Michael Callen for sharing these data.

[57] For consistency, the analysis uses the same boundary neighborhoods used in previous results.

[58] Refer to online Appendix Table B9 for summary statistics on variables of interest and by coverage status of the polling center.

TABLE 9—EFFECT OF CELL PHONE COVERAGE ON THE NUMBER AND TYPES OF FRAUD COMPLAINTS

	Baseline (1)	Baseline (2)	Fixed effects (3)	Fixed effects + controls (4)	Fixed effects + bandwidth (5)	Fixed effects + bandwidth (6)
Panel A. Number of complaints						
Inside coverage	0.257	0.144	0.252	0.208	0.231	0.206
	(0.079)	(0.087)	(0.105)	(0.109)	(0.109)	(0.111)
Observations	2,785	2,785	2,785	2,785	691	691
Mean outside coverage	0.33	0.33	0.33	0.33	0.27	0.27
Neighborhoods	384	384	384	384	101	101
Panel B. Share of complaints submitted by individuals						
Inside coverage	0.028	0.048	0.072	0.102	0.060	0.040
	(0.034)	(0.037)	(0.056)	(0.059)	(0.071)	(0.041)
Observations	717	717	717	717	139	139
Mean outside coverage	0.91	0.91	0.91	0.91	0.92	0.92
Neighborhoods	194	194	194	194	57	57
Panel C. Share of complaints submitted by females						
Inside coverage	0.058	0.059	0.128	0.099	0.108	0.083
	(0.026)	(0.028)	(0.043)	(0.041)	(0.043)	(0.046)
Observations	717	717	717	717	139	139
Mean outside coverage	0.080	0.080	0.080	0.080	0.0040	0.0040
Neighborhoods	194	194	194	194	57	57
Geographic controls	Yes	Yes	No	Yes	No	Yes
Demographic controls	Yes	Yes	No	Yes	No	Yes
Economic controls	No	Yes	No	Yes	No	Yes

Notes: Sample restricted to south and east regions only. Fixed effects refers to neighborhoods fixed effects. Refer to Section IIC for a description of how boundary neighborhoods are created. Bandwidth in columns 5 and 6 is 10 km. Panel names indicate the outcome variables used. Individuals refers to citizens, candidates, or candidates' representatives and excludes polling officials, IEC or ECC officials, and other government organizations. Geographic controls include elevation and slope. Demographic controls include population and language indicator for nearest village. Economic controls include distance to nearest basic and district health facility and distance to nearest river. Refer to Section IIIC for more details on the specifications.

in centers with coverage (column 1). The results remain robust when including all sets of geographic, demographic, and economic development controls (column 2) and when adding neighborhood fixed effects (columns 3 and 4). Columns 5 and 6 provide results that are more comparable to the RD results presented before. I restrict the analysis to polling centers within a 10 km bandwidth around the coverage boundary. Note that the effect remains positive and significant; however, the results in columns 5 and 6 should be interpreted with care given the higher potential for measurement error in the coverage variable for centers close to the boundary. Also note that some of the polling centers classified as uncovered in this analysis are potentially covered by 2010, so that means that the results presented here are potentially underestimating the magnitude of the effect.

Panel B presents results on the relationship between coverage and the share of total complaints submitted by individuals rather than government organizations such as the ECC. The estimated coefficients show that coverage increases the share of complaints submitted by individuals. The magnitudes of the results are consistent across specifications; however, they are not statistically significant for the most part. Lastly, panel C further explores the effect of coverage on political enfranchisement by studying how coverage affects the share of complaints submitted by females, a highly disenfranchised group in Afghanistan. The results show that coverage leads to economically and statistically significant increases in the share of complaints submitted by females. This is consistent across all specifications. The findings in this section provide compelling evidence that there is a clear link between cell phone coverage and citizen enfranchisement and participation in election monitoring. Importantly, this suggest that the documented drops in fraud at the boundary can potentially be explained by this mechanism.

D. *Candidate-Voter Affinity*

From the theoretical framework in Section IIIA, note that if parameter a (i.e., a voter's affinity toward a candidate) significantly jumps with coverage, then the price of legal votes $p_l = (\delta, P, a)$ in covered centers drops relative to fraudulent votes, which then leads the candidate to substitute fraudulent votes for legal votes. If that is the case, then the drops in fraud reported in Section IIC can be simply explained by sharp changes at the boundary in a voter's affinity toward the candidate.

In the case of Afghanistan, ethnic and tribal identity are strong predictors of candidate-voter affinity (Lamb and Tarzi 2011). For instance, individuals with the same tribal affiliation as a candidate may be more willing to vote in spite of violence and exhibit a higher tolerance to fraudulent actions by this candidate. In this context, the coverage effect on fraud may be explained by sharp increases at the coverage boundary in tribes that form the voting base of a candidate. This may result, for instance, if cell phone providers give preference to certain ethnic groups by expanding coverage into their locations.

To examine the spatial distribution of ethnic groups and tribes relative to the coverage boundary, I georeference detailed tribal maps of the southeastern region of Afghanistan obtained from the Culture and Conflict Studies program at the Naval Postgraduate School and based on the Tribal Hierarchy and Dictionary of Afghanistan (2007).[59] Georeferenced maps are then combined with village coordinate data from the Measuring Impacts of Stabilization Initiatives project (MISTI 2013) to construct village-level indicators of primary tribe for almost 18,000 villages. I aggregate the more than 50 tribes represented in the sample into 7 tribal confederations using the Tribal Hierarchy and Dictionary of Afghanistan (2007). Confederations are typically formed by groups of tribes with common origin or historical alliances. Appendix Figure A6, panel A presents the spatial distribution of each village's primary tribal confederation. In all, about 66 percent of the villages in the sample are Pashtun villages, with the two largest confederations being the Durrani confederation (37 percent of all Pashtun villages) and the Ghilzai confederation (38 percent of all Pashtun villages). The second-largest ethnic group represented in the sample is the Hazaras, with about 24 percent of the villages belonging to that group.[60]

To explore the possibility of jumps in tribal composition at the coverage boundary, I replicate the scalar RD design using four distinct outcomes: indicators for

[59] For more information on the tribal maps, refer to the Culture and Conflict Studies website (https://my.nps.edu/web/ccs/afghanistan1).

[60] Online Appendix Table B10 presents summary statistics on tribal composition in southeastern Afghanistan.

Table 10—Tribal Landscape and Cell Phone Coverage

	Pashtun (1)	Durrani (2)	Popalzai (3)	Zirak (4)
Panel A. Optimal bandwidth				
Inside coverage	−0.009	−0.019	0.032	0.008
	(0.011)	(0.013)	(0.032)	(0.021)
Observations	7,204	1,676	684	1,930
Mean outside coverage	0.70	0.66	0.024	0.19
Bandwidth (km)	6.07	5.31	6.05	5.10
Neighborhoods	307	96	49	108
Panel B. Polynomial in distance to boundary				
Inside coverage	−0.033	−0.001	0.058	−0.007
	(0.018)	(0.037)	(0.046)	(0.038)
Observations	14,492	3,666	1,660	4,186
Mean outside coverage	0.53	0.66	0.011	0.21
Neighborhoods	316	104	53	116

Notes: Column indicates outcome variable. Outcome variables are an indicator for whether a village is classified as the specified tribe or ethnic group. Classification done using the Tribal Hierarchy and Dictionary of Afghanistan (2007). Refer to online Appendix Table B10 for more details on the classification of tribes. Popalzai is a Pashtun tribe from the Durrani confederation. Durrani confederation is split into Zirak and Panjpai subconfederation. Zirak include the Popalzai, Barakzai, Ashakzai, and Alikozai (Giustozzi 2009). Refer to Section IIID for a description of the choice of these tribes for the analysis.

whether a village belongs to the Pashtun ethnic group, Durrani confederation, Popalzai tribe, and Zirak subconfederation. The choice of outcomes is based on the fact that then-candidate Hamid Karzai is a Pashtun who belongs to the Popalzai tribe in the Zirak group of the Durrani confederation.[61] Therefore, using the voter's affinity argument discussed above, increases at the boundary in the share of villages belonging to any of these categories can potentially explain the observed drops in fraud. Note that the analysis is limited to these groups and then-candidate Karzai because tribal data is only available for southeast Afghanistan, which comprised Karzai's main base of support, and most drops in fraud presented in Section IIC were located in this region.

Table 10 presents the RD estimates. For all categories, there is no evidence of an increase in the likelihood of tribes sympathetic to Karzai at the boundary. For the most part, the estimated RD coefficient is negative and close to zero in magnitude. Although there is an increase, albeit a statistically insignificant one, in the share of villages belonging to Karzai's tribe (column 3), these results are not very reliable. Note in Appendix Figure A6, panel B that almost all Popalzai villages are entirely within the coverage area of Kandahar province, which difficults the identification of a reliable RD coefficient.

E. *Coordination and Collusion*

Coverage can significantly reduce the cost of coordination between a candidate and corrupt election officials. For example, if turnout is low due to violence, it is

[61] The Zirak group includes the Popalzai, Barakzai, Ashakzai, and Alikozai Giustozzi (2009).

much easier and faster for the candidate to direct a polling center official on the other side of the country to inflate vote tallies or engage in ballot stuffing if that official is in a coverage area. Coverage can thus facilitate fraud through a collusion channel. If that is the case, then the effect on fraud documented in Section IIC can be biased downward as a result.

This section explores whether coverage can facilitate collusion. First, I overlay the polling centers on the tribal maps to classify polling centers according to the ethnic and tribal composition of the areas where they are located. This serves as a proxy for the tribal affiliation of the election officials for that polling center. Then I use candidate-polling-center tribal match to determine which polling centers are more likely to engage in collusion. For instance, Karzai, a Popalzai of the Durrani confederation, is more likely to collude with polling centers in Durrani or Popalzai areas than with officials from the Ghilzai confederation, a historic geopolitical rival of the Durranis. I then compare fraud levels across coverage for centers where collusion is likely to take place based on tribal connections. If coverage has an effect on fraud through a collusion channel, then we should expect higher levels of fraud favoring Karzai in polling centers in Durrani villages with coverage than in Durrani centers without coverage. As before, given that the tribal data is only available for southeast Afghanistan, I limit the analysis to the tribal connections of then-candidate Karzai.

Table 11 presents estimates from a Difference-in-Difference (DD) estimator that compares fraud levels across coverage for polling centers in Durrani areas relative to centers in other non-Durrani Pashtun areas. The analysis uses Durrani areas as the measure of tribal kinship (and hence potential collusion) between Karzai and the polling centers, as there were very few centers classified as Popalzai (22 in total). The DD specification also includes controls for the number of stations at the center, the number of female stations, the elevation and slope of the area, and the population of the village where the center is located. Column 2 replicates the DD exercise restricting the sample to centers within a bandwidth around the coverage boundary. Column 3 adds district fixed effects to the restricted DD estimates. As expected from previous results, the coefficient on coverage is consistently negative and highly significant in one specification. Interestingly, the coefficient on Durrani is consistently positive and significant in the most restrictive specification (columns 3 and 6). This provides support for the argument that tribal match between the candidate and an area can lead to higher levels of fraud. The coefficient on the interaction term provides an estimate of the effect on fraud that is potentially due to coverage-driven collusion. However, I find no evidence that fraud levels are significantly higher in Durrani centers with coverage relative to centers outside coverage areas.

IV. Conclusion

The results in this paper provide considerable evidence that cell phone coverage lowers electoral fraud. Specifically, up to 70 percent of the estimated boundary treatment effects along the coverage boundary are negative, implying a drop in fraud for centers just inside coverage areas. It is important to highlight that the estimated effects exhibit a considerable degree of spatial heterogeneity: average impacts in the south and east are economically and statistically significant while nondistinguishable

TABLE 11—EFFECT OF CELL PHONE COVERAGE AND TRIBAL AFFILIATION ON CATEGORY C FRAUD

	At least one station with Category C fraud			Share of votes under Category C fraud		
	Baseline DD (1)	DD-RD (2)	DD-RD (3)	Baseline DD (4)	DD-RD (5)	DD-RD (6)
Inside coverage	−0.077	−0.096	−0.073	−0.002	−0.043	0.003
	(0.050)	(0.047)	(0.052)	(0.034)	(0.030)	(0.036)
Durrani	0.104	0.124	0.234	0.126	0.099	0.236
	(0.118)	(0.098)	(0.111)	(0.101)	(0.069)	(0.103)
Coverage × Durrani	−0.014	0.029	−0.047	−0.055	−0.020	−0.075
	(0.099)	(0.091)	(0.085)	(0.079)	(0.062)	(0.047)
Observations	853	486	486	853	471	471
Bandwidth (km)	—	8.44	8.44	—	7.85	7.85
District fixed effects	Yes	No	Yes	Yes	No	Yes
Districts	93	66	66	93	66	66

Notes: Baseline DD refers to difference-in-difference (DD) specification where the estimation sample is all polling centers located in majority Pashtun villages, omitting Ghilzai-majority villages. DD-RD refers to DD specification that restricts the estimation sample above to observations within an optimal bandwidth of the coverage boundary. Optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik (2014). Columns 1, 2, 4, and 5 include controls for number of stations, number of female stations, elevation, slope, and population. All specifications are clustered by district.

from zero in other parts of the country. These results are potentially explained by a spatial pattern of election-related violence that strongly mimics the observed pattern of fraud.

Given the widespread political violence during the 2009 Afghan presidential election, I test empirically whether this metric may be a potential mechanism explaining the link between coverage and fraud. The results show that there is no clear evidence that changes in insurgent violence at the coverage boundary can explain the observed drops in fraud. Another set of results tests whether changes at the coverage boundary in voters' affinity toward a candidate is the primary channel. Detailed analysis on whether there is a surge in village-candidate tribal match at the boundary—a strong predictor of voters' affinity toward a particular candidate—shows no evidence of significant changes. Similarly, additional results on tribal-based coordination and collusion between candidates and polling officials show that although there is an increase in fraud in centers with a candidate-official tribal match, this effect does not depend on the coverage status of the center.

I find compelling evidence that coverage increased enfranchisement and citizen participation in election monitoring during a second election that followed shortly after the one studied in this paper. Overall, the absence of significant changes at the boundary in both election-related violence and the tribal composition of villages suggests that the impact of cell phone access on citizen enfranchisement and citizen-based monitoring likely explains the observed drops in fraud. From a policy perspective, this paper illustrates that the availability and expansion of cell phone access and usage, along with citizen-based monitoring initiatives, can have positive externalities on institutional development via fraud deterrence and the mitigation of corrupt behavior in general.

To conclude, it is important to keep in mind the 2009 Afghan context if one wants to generalize the role of information-communication technologies to other similar settings. Coverage access can significantly reduce the cost of collective action efforts (e.g., reducing the cost of obtaining/transferring information, shielding informers from retaliation by corrupt officials or violent actors). However, the success of coverage as a fraud-deterring mechanism is also linked to the existence of an institution that can efficiently employ the technology and credibly act on complaints if necessary. In the Afghan context, this took place in the form of the ECC.[62]

This is important for two key reasons: First, a credible, independent institution like the ECC can reduce factors that typically undermine collective action (elite capture of the accountability process or freeriding by citizens if they feel that their complaints are nonpivotal.) This is particularly important in settings with a fragile security environment where levels of institutional trust are low. Second, such institutions increase the perception among candidates that stipulated punishments—likely in the form of lost fraudulent votes if an audit is triggered—are binding. Therefore, a key policy recommendation is that for ICT interventions to be successful, they likely have to be accompanied by credible institutions that can act on the feedback received from the collective action efforts that phones facilitate.

Appendix A

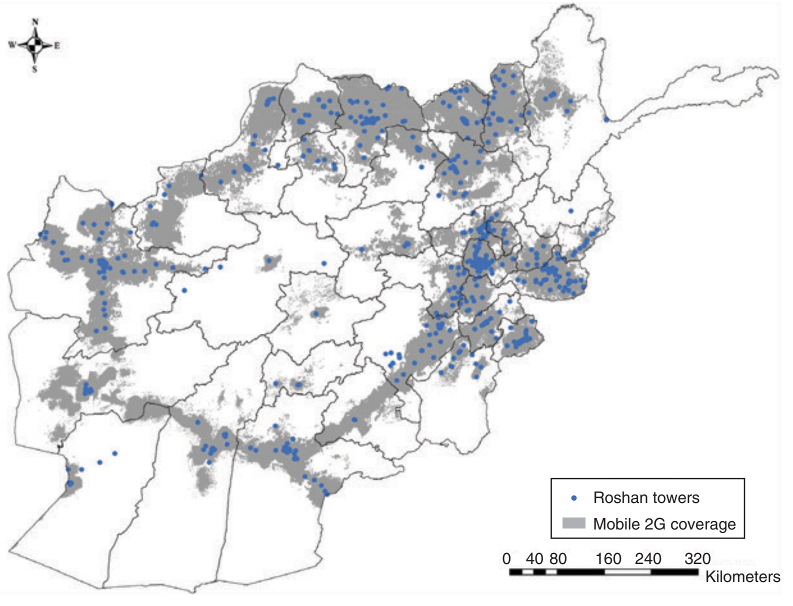

Figure A1. Roshan Tower Footprint (2009)

Notes: Blue dots give the location of Roshan towers. Shaded areas represent cell phone coverage without including Roshan coverage.

[62] Recall that during the 2009 election, the ECC was headed by UN-appointed international experts and was given the power to issue audits and even order a runoff if necessary.

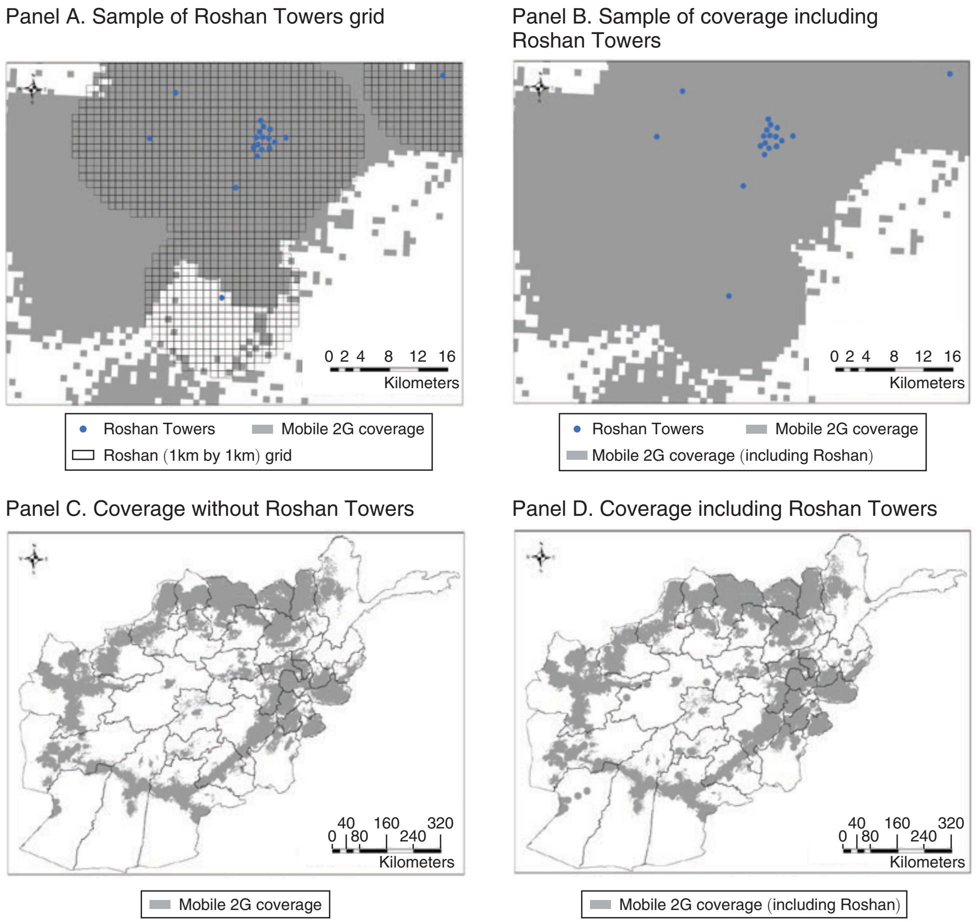

Figure A2. Roshan Towers Grid and Coverage after Inclusion of Roshan Grid

Notes: Blue dots give the location of Roshan towers. Panels A and B present a sample of towers around Lashkar Gah in Helmand province. Panel A uses a 1 km by 1 km grid around Roshan towers. Shaded areas in panel C represent cell phone coverage without including Roshan coverage. Shaded areas in panel D represent cell phone coverage after including Roshan coverage grid.

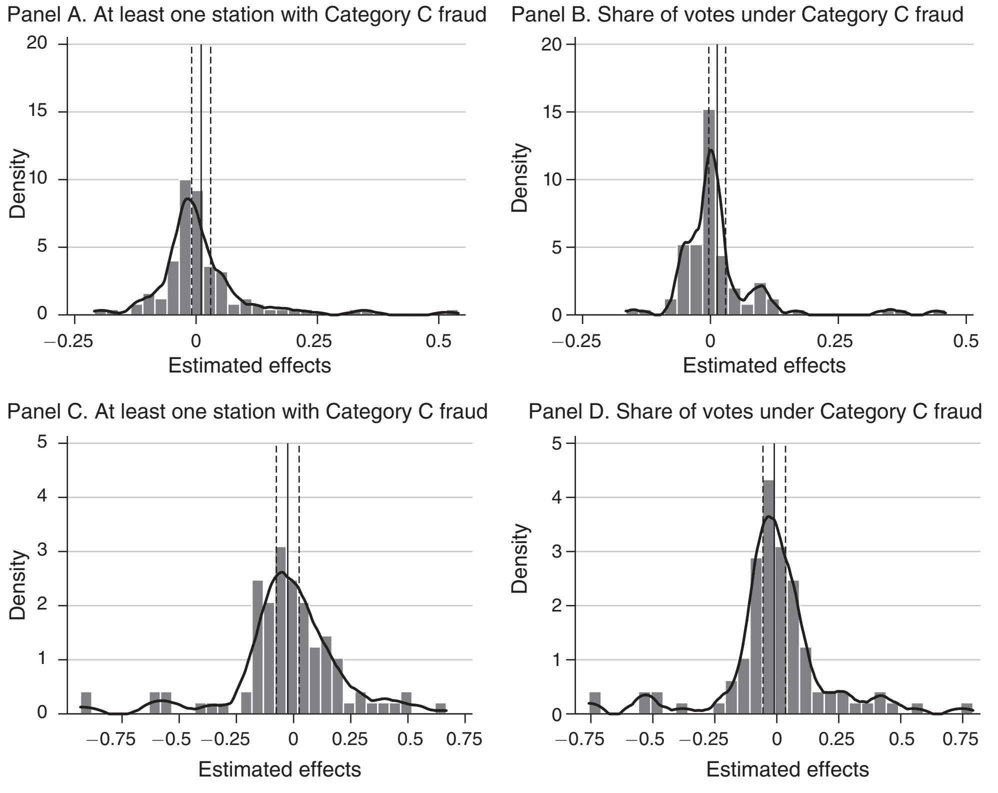

Figure A3. Effect of Coverage on Fraud Using False Coverage Boundaries

Notes: Histogram of estimated RD coefficients using 100 randomly selected longitudes (panels A and B) and latitudes (panels C and D) as false coverage boundaries. Coefficients obtained from estimating equation (3). Polling centers with (true) coverage are omitted from the analysis. Solid and dashed vertical lines give the average and 95 percent confidence intervals of the estimated coefficients, respectively. Refer to Section IIID for more detail on the estimation.

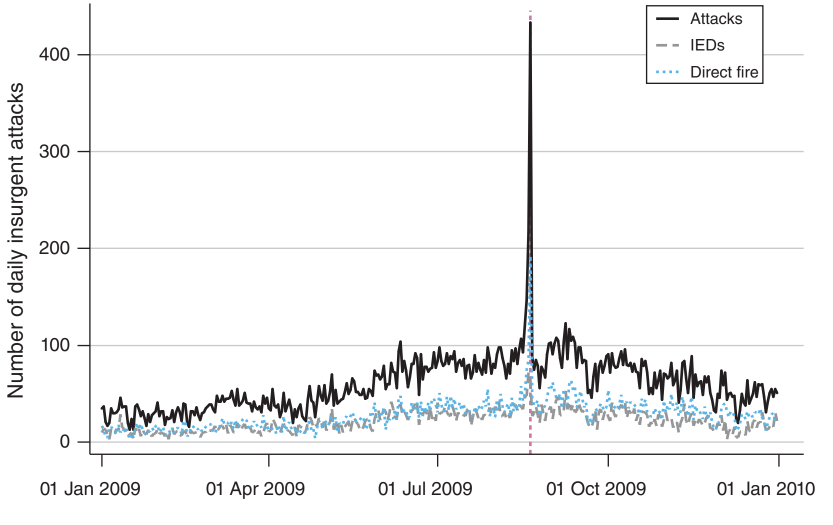

Figure A4. Insurgent Attacks on Election Day

Notes: Vertical line indicates election day: August 20, 2009. Data are significant actions collected by Afghan forces and ISAF. Refer to Condra et al. (2018a) for detailed description of the data.

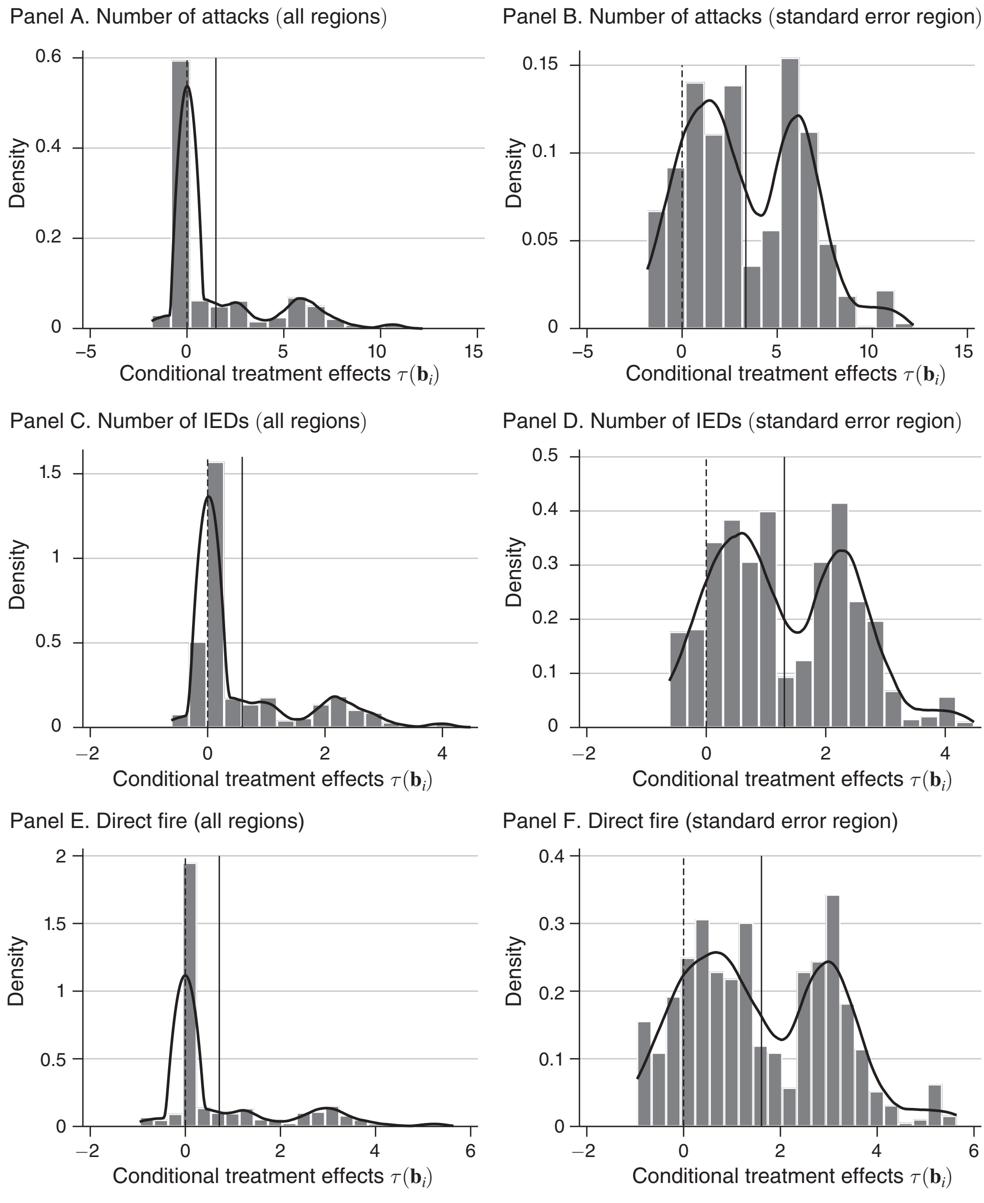

FIGURE A5. DISTRIBUTION OF BOUNDARY TREATMENT EFFECTS (INSURGENT VIOLENCE OUTCOMES)

Notes: Histogram of the estimated boundary treatment effects for outcomes indicated in panel title. Violence outcomes are measured between January 1, 2009 and August 20, 2009. Refer to Section IIB for description of estimation. Solid vertical line gives the simple average of the estimated effects. Solid line represents the estimated density. The density estimate uses an Epanechnikov kernel function.

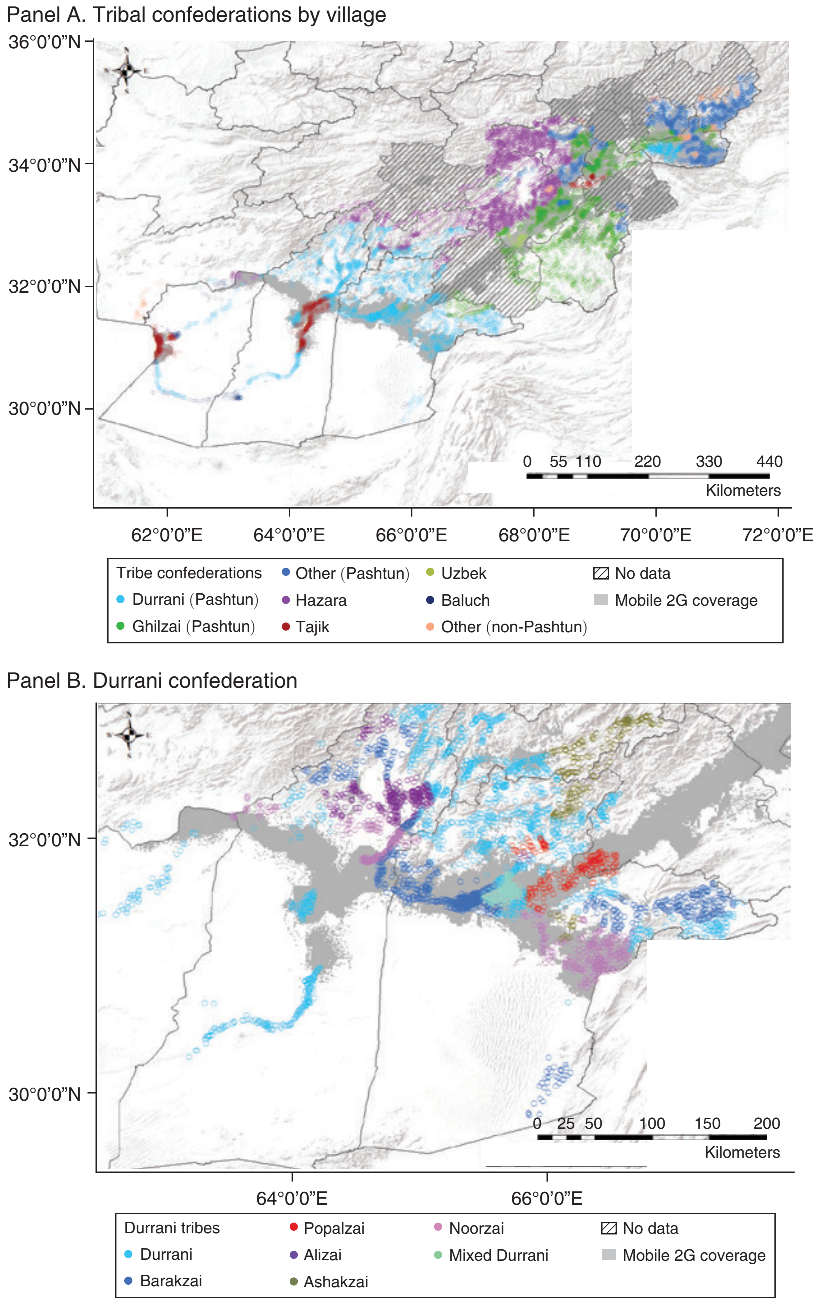

Figure A6. Major Tribal Confederations, Southeast Afghanistan

Notes: Classification of tribes and confederations done using the Tribal Hierarchy and Dictionary of Afghanistan (2007). Refer to online Appendix Table B10 for more details on the classification of tribes. Shaded areas represent availability of 2G GSM cell phone coverage. Dots give the location of villages (MISTI 2013). Lines demarcate the provinces of Afghanistan.

REFERENCES

Afghan Telecommunication Regulatory Authority (ATRA). 2012. "Coverage Footprint (2012)." http://atra.gov.af/en/page/7000/7006/coverage-footpring-2012.

AIMS. 1997–2005. "Roads of Afghanistan. Islamabad, Pakistan." Afghanistan Information Management Services (AIMS).

Aker, Jenny C. 2010. "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger." *American Economic Journal: Applied Economics* 2 (3): 46–59.

Aker, Jenny C., Paul Collier, and Pedro C. Vicente. 2017. "Is Information Power? Using Mobile Phones and Free Newspapers during an Election in Mozambique." *Review of Economics and Statistics* 99 (2): 185–200.

Aker, Jenny C., and Christopher Ksoll. 2020. "Can ABC Lead to Sustained 123? The Medium-Term Effects of a Technology-Enhanced Adult Education Program." *Economic Development and Cultural Change* 68 (3): 1081–1102.

Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert. 2012. "Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger." *American Economic Journal: Applied Economics* 4 (4): 94–120.

Associated Press. 2011. "Taliban Turn Cell Phones Back on in Afghan South." *CTV News*, April 5. https://www.ctvnews.ca/taliban-turn-cell-phones-back-on-in-afghan-south-1.627627.

Beber, Bernd, and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20 (2): 211–34.

Besley, Tim, Hannes Mueller, and Prakarsh Singh. 2011. "Conflict and Investment." https://pdfs.semanticscholar.org/8635/4b0f83919fd50e6520dcddd2d3857c1924d4.pdf.

Black, Dan, Jose Galdo, and Jeffrey Smith. 2007. "Evaluating the Regression Discontinuity Design Using Experimental Data." http://economics.uwo.ca/newsletter/misc/2009/smith_mar25.pdf.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114 (2): 577–99.

Blumenstock, Joshua, Michael Callen, Tarek Ghani, and Robert Gonzalez. 2019. "Violence and Financial Decisions: Evidence from Mobile Money in Afghanistan." Unpublished.

Blumenstock, Joshua E., Nathan Eagle, and Marcel Fafchamps. 2016. "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters." *Journal of Development Economics* 120: 157–81.

Callen, Michael, and James D. Long. 2014. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." https://scholar.harvard.edu/files/michael-callen/files/institutional_corruption_3.pdf.

Callen, Michael, and James D. Long. 2015. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." *American Economic Review* 105 (1): 354–81.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2019. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association* 115 (531): 1449–55.

Chassang, Sylvain, and Gerard Padro i Miquel. 2014. "Corruption, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports." Princeton University Economics, Econometric Research Program Working Paper 062-2014.

Collier, Paul, and Pedro Vicente. 2012. "Violence, Bribery, and Fraud: The Political Economy of Elections in Sub-Saharan Africa." *Public Choice* 153 (1): 117–47.

Condra, Luke N., James D. Long, Andrew C. Shaver, and Austin L. Wright. 2018a. "The Logic of Insurgent Electoral Violence." *American Economic Review* 108 (11): 3199–3231.

Condra, Luke N., James D. Long, Andrew C. Shaver, and Austin L. Wright. 2018b. "The Logic of Insurgent Electoral Violence: Dataset." *American Economic Review*. https://www.openicpsr.org/openicpsr/project/113182/version/V1/view.

Conley, T.G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92 (1): 1–45.

Cordesman, Anthony. 2005. *Iraq's Evolving Insurgency*. Washington, DC: Center for Strategic and International Studies (CSIS).

Cremer, Helmuth, and Firouz Gahvari. 1994. "Tax Evasion, Concealment and the Optimal Linear Income Tax." *Scandinavian Journal of Economics* 96 (2): 219–39.

Dell, Melissa. 2010. "The Persistent Effect of Peru's Mining *Mita*." *Econometrica* 78 (6): 1863–1903.

Electoral Complaints Commission. 2010. *Final Report. 2009 Presidential and Provincial Council Elections*. Kabul. https://www.dropbox.com/s/zq4eo4tlsp6uy0y/ECC%20Final%20Report%20 2009.pdf?dl=0.

Faiez, Rahim. 2008. "Afghanistan to Hold Separate Presidential, Parliamentary Elections." *Associated Press*. https://www.taiwannews.com.tw/en/news/636889.

Figueiras, João, and Simone Frattasi. 2010. *Mobile Positioning and Tracking—From Conventional to Cooperative Techniques*. West Sussex, UK: Wiley.

Filkins, Dexter. 2009. "Threats by Taliban May Sway Vote in Afghanistan." *New York Times*, August 16. https://www.nytimes.com/2009/08/17/world/asia/17taliban.html.

Gall, Carlotta. 2009. "Violence Roils Afghanistan Days before Election." *New York Times*, August 18. https://www.nytimes.com/2009/08/19/world/asia/19afghan.html.

Giustozzi, Antonio, ed. 2009. *Decoding the New Taliban: Insights from the Afghan Field*. New York: Columbia University Press.

Gonzalez, Robert. 2019. "Database of News Articles Referencing Cell Phone Tower Shutdowns in Afghanistan, 2008–2019." Cell Phone Tower Shutdowns in Afghanistan.

Gonzalez, Robert M. 2021. "Replication data for: Cell Phone Access and Election Fraud: Evidence from a Spatial Regression Discontinuity Design in Afghanistan." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E118467V1.

Greene, William H. 2003. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

GSMA. 2009. "Mobile Coverage Explorer." Collins Bartholomew. https://www.collinsbartholomew.com/mobile-coverage-maps/mobile-coverage-explorer (accessed: April 15, 2015).

Hamdard, Javid. 2012. *The State of Telecommunications and Internet in Afghanistan: Six Years Later 2006–2012*. Washington, DC: USAID.

Himelfarb, Sheldon. 2010. "Can You Help Me Now? Mobile Phones and Peace-Building in Afghanistan." United States Institute of Peace. https://www.usip.org/publications/2010/11/can-you-help-me-now.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy* 106 (4): 667–705.

IEC. 2009. "Presidential and Provincial Council Elections, Afghanistan 2009 Elections." Independent Elections Commission.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–59.

Imbens, Guido, and Tristan Zajonc. 2011. "Regression Discontinuity Design with Multiple Forcing Variables." Unpublished.

Jack, William, and Tavneet Suri. 2011. "Mobile Money: The Economics of M-PESA." NBER Working Paper 16721.

Jack, William, and Tavneet Suri. 2014. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review* 104 (1): 183–223.

Jack, William, Tavneet Suri, and Robert M. Townsend. 2010. "Monetary Theory and Electronic Money: Reflections on the Kenyan Experience." *FRB Richmond Economic Quarterly* 96 (1): 83–122.

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics* 122 (3): 879–924.

Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger. 2006. "School Quality, Neighborhoods, and Housing Prices." *American Law and Economic Review* 8 (2): 183–212.

Keele, Luke, and Rosarío Titiunik. 2013. "Geographic Boundaries as Regression Discontinuities." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.363.6915&rep=rep1&type=pdf.

Khadhouri, Sandra. 2010. *A Review of Suspected Electoral Fraud: 2009 Afghan Presidential and Provincial Council Elections*. Democracy International. https://www.dropbox.com/s/a9qevpqv2yxr932/ElectionsReportENGFINAL2.pdf?dl=0.

Lakshmanan, Indira A.R. 2010. "Fighting the Taliban With Cellphones." *New York Times*, March 23. https://www.nytimes.com/2010/03/24/world/asia/24iht-letter.html.

Lalive, Rafael. 2008. "How Do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach." *Journal of Econometrics* 142 (2): 785–806.

Lamb, Robert, and Amin Tarzi. 2011. *Measuring Perceptions about the Pashtun People*. Washington, DC: Center for Strategic and International Studies (CSIS).

Leahy, Kevin. 2005. "The Impact of Technology on the Command, Control, and Organizational Structure of Insurgent Groups." Master's thesis, US Army Command and General Staff College, Fort Leavenworth, KS.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

León, Gianmarco. 2012. "Civil Conflict and Human Capital Accumulation the Long-Term Effects of Political Violence in Perú." *Journal of Human Resources* 47 (4): 991–1022.

MISTI. 2013. "Measuring Impacts of Stabilization Initiatives (MISTI)." http://usaidmisti.com/gis-data.

National Aeronautics and Space Administration and the National Geospatial Intelligence Agency. 2000. "Retrieved from Shuttle Radar Topography Mission 30 Arc Second Finished Data."

National Democratic Institute. 2010. *The 2009 Presidential and Provincial Council Elections in Afghanistan*. Washington, DC: National Democratic Institute.

Nolin, Pierre Claude. 2011. *Countering the Afghan Insurgency: Low-Tech Threats, High-Tech Solutions*. Brussels: NATO Parliamentary Assembly.

Pierskalla, Jan H., and Florian M. Hollenbach. 2013. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107 (2): 207–24.

Robinson, Frances. 2013. "Fewer Cell Towers Are Shut Down in Afghanistan, Minister Says." *Wall Street Journal*, February 28. https://www.wsj.com/articles/BL-TEB-5292.

Roshan. 2009. "Roshan Telecom Tower Footprint, Afghanistan 2003–2009." Roshan Telecom, Afghanistan.

Sameem, Ismail. 2011. "Taliban Stop Cell Phone Signals in Key Afghan Province." *Reuters*, March 24. https://www.reuters.com/article/afghanistan-taliban-phones/taliban-stop-cell-phone-signals-in-key-afghan-province-idUSL3E7EO12O20110324.

Schuler, Ian. 2008. "National Democratic Institute: SMS as a Tool in Election Observation." *Innovations* 3 (2): 143–158.

Shalizi, Hamid. 2008. "Taliban Orders Mobile Shutdown in Afghan Province." *Reuters*, October 21. https://www.reuters.com/article/us-afghan-mobiles/taliban-orders-mobile-shutdown-in-afghan-province-idUSTRE49K2HS20081021.

Shapiro, Jacob N., and Nils B. Weidmann. 2015. "Is the Phone Mightier than the Sword? Cellphones and Insurgent Violence in Iraq." *International Organization* 69 (2): 247–74.

Singh, Prakarsh. 2013. "Impact of Terrorism on Investment Decisions of Farmers: Evidence from the Punjab Insurgency." *Journal of Conflict Resolution* 57 (1): 143–68.

Tribal Hierarchy and Dictionary of Afghanistan. 2007. *Tribal Hierarchy and Dictionary of Afghanistan: A Reference Aid for Analysts*. Courage Services, Inc. http://www.nzdl.org/gsdlmod?e=d-00000-00-...00--off-0areu--00-0----0-10-0---0---0direct-10---4-----dte--0-1l--11-en-50---20-about---00-0-1-00-0-0-11-1-0utfZz-8-00-0-0-11-10-0utfZz-8-00&cl=CL5.17&d=HASH2a85fd9aefbb0309f2bc86&x=1.

Trofimov, Yaroslav. 2010. "Cell Carriers Bow to Taliban Threat." *Wall Street Journal*, March 22. https://www.wsj.com/articles/SB10001424052748704117304575137541465235972.

Weidmann, Nils B., and Micheal Callen. 2013. "Violence and Election Fraud: Evidence from Afghanistan." *British Journal of Political Science* 43 (1): 53–75.

Wong, Vivian C., Peter M. Steiner, and Thomas D. Cook. 2013. "Analyzing Regression-Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods." *Journal of Educational and Behavioral Statistics* 38 (2): 107–41.

World Bank, and United Nations (UN). 2018. *Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict*. Washington, DC: World Bank.