Biometric Identification Machine Failure and Electoral Fraud [1]

Miriam Golden*, George Ofosu‡, and Luke Sonnet†

*European University Institute
‡London School of Economics and Political Science
†University of California, Los Angeles

December 24, 2021

Word Count:
10196 (errors:7)
(including notes, references, tables and figures;
excluding appendices)

Version 6.2

Key words: elections, electoral fraud, election observation, biometric identification, Ghana

[1]Corresponding author: Miriam Golden, Peter Mair Chair in Comparative Politics, European University Institute; email:miriam.golden@eui.eu. Earlier versions of this paper were presented at the 2014 annual meetings of the American Political Science Association, August 28–31, Washington, D.C.; at EGAP: Experiments in Governance and Politics 13, University of California at Los Angeles, February 20–21, 2015; at Columbia University, April 15, 2015; and at the University of Michigan, April 24–25, 2015. We are grateful to Jasper Cooper and Tara Slough for the extensions they suggested. In Ghana, we acknowledge the collaboration of our research partner, the Centre for Democratic Development, as well as the Coalition of Domestic Election Observers. We also thank our 300 field assistants for data collection. Joseph Asunka, Sarah Brierley, and Eric Kramon collaborated on the design and implemention of the original project. Brian Hamel and Bronwyn Lewis provided research assistance. Funding came from the U.K.'s Ghana office of the Department for International Development, a National Science Foundation Grant for Rapid Response Research (RAPID) SES–1265247, and the UCLA Academic Senate, none of which bears responsibility for the results reported here. This research was approved by the University of California at Los Angeles IRB #12 − 001543 on October 26, 2012.

Abstract

Results from a randomized field experiment are used to investigate the operation of newly-introduced biometric identification machines and their possible connection to election fraud in Ghana's 2012 general elections. Evidence uncovers a non-random pattern to frequent equipment breakdowns. Machines were 50 percent less likely to break down in polling stations with a randomly assigned domestic election observer as in those without. We also find that machine breakdown occurred more often in competitive constituencies. Machine malfunction is associated with more election fraud, especially where election observers are not present. Overall, our results suggest that partisan competition may promote election fraud via the opportunity presented by the unexpected breakdown in the equipment used to verify voters' identities. They document that domestic election observers improve election integrity through direct observation and also via second-order effects on election administration, including voter identification procedures. [150 words]

Most countries in the world use elections to select their political leaders but in many countries, the electoral process is compromised by illegitimate strategic manipulation on the part of various actors. Election fraud is commonplace, affecting an estimated fifth of executive branch outcomes around the world (**?**). How fraud occurs, who perpetuates it, and which preventive efforts are effective are inadequately understood. Nonetheless, analysts concur that fraud may distort election outcomes and that accusations of fraud may trigger post-election conflict. Thus, fraud may be politically consequential.

Governments with aspirations to reduce election fraud have responded in at least two ways. The first involves deploying election observers, especially teams from international bodies whose mandates include ensuring election integrity. Eighty percent of elections around the world are observed by monitors specifically to reduce fraud (**?**). Multiple studies show that international election observers operate as anticipated, successfully reducing election fraud (**???????**). A second way to reduce election fraud involves introducing new technologies aimed at exposing or preventing it. Technological solutions, such as electronic voting machines, polling station webcams, and biometric identification equipment, offer the promise of rapid, accurate, and ostensibly tamper-proof innovations. They are expected to reduce fraud in the processes of voter registration, voting, or vote count aggregation. However, little is known about the effectiveness of these technologies, and some claim they may be less useful than human observation (**??**). In practice, even the best technology is neither failure- nor tamper-proof **?**. As the 2020 primary elections in the state of Georgia illustrate, introducing new technologies into the voting process can also introduce unintended obstacles to electoral participation (**?**). Systematic evidence on the impact of anti-fraud voting technologies is thus far scattered and fragmentary, however (e.g. **??**).

In this paper, we contribute to scholarship investigating the effectiveness of technological methods to prevent election fraud, defined as attempts to interfere with election outcomes (**?**). We report results of a study conducted during Ghana's 2012 general elections.

These elections saw the government place newly-acquired biometric identification machines in every polling station in the country with the intention of erecting a technological barrier to election fraud. Unexpectedly, at least 20 percent of biometric identification machines broke down at some point on election day, thereby disrupting the voting process, a fact that was widely reported in the press. Machine failures provoked significant political consequences, contributing to legal appeals by the major losing party and a Supreme Court case that was not decided until eight months after the elections.

Our research was designed as an experiment that randomized the placement of domestic election observers across polling places. Thanks to the random placement of observers in conjunction with the universal deployment of biometric machines, we are able to causally identify a non-random pattern to machine breakdown. Biometric identification machines were twice as likely to break down in polling stations without a domestic election observer as in stations where one was present; in one-third of polling stations that were not under observation, biometric identification machines failed to operate at some point on election day. We interpret this as evidence of some deliberate interference with equipment whose sole purpose was the prevention of illegal voting.

To interpret this result further, we examine theoretically salient observable characteristics of polling station areas. We show that breakdown was more prevalent in electorally competitive areas. Our interpretation of this is that electoral competition encouraged machine interference. We also find that markers of election malfeasance and fraud were more common in polling stations affected by machine breakdown, and that this effect was amplified when an election observer was not present. In particular, data show that when machines failed to operate, more ballot stuffing occurred. We discuss the logistics of ballot stuffing to explain this.

The contribution of this paper is twofold. First, to the best of our knowledge this is the first systematic empirical investigation of how well biometric identification machines operate

at scale in an electoral context. The unexpected breakdown of so many machines in Ghana in their initial rollout provides a context for us to consider how this technology is associated with election integrity. The association of markers of fraud with machine breakdown suggests that biometric identification in the polling place — when it operates properly — is effective in preventing fraud. Second, our work reinforces the suspicion that government crackdowns on election fraud, like crackdowns on other forms of political malfeasance, engender immediate strategic reactions by those affected. When new technologies are put in place to expose or prevent political wrongdoing, the agents who previously were benefitting from illegal activities have incentives to try to evade, reverse, or reduce the impact of such oversight. Our study illustrates this.

Theory and Hypotheses

Competitive elections are the bedrock of democracy (**?**). Election fraud threatens democracy in various ways. It undermines the process of democratic choice, threatens the legitimacy of the outcome, and enlarges the possibility of post-election violence. This gives governments and international democracy sponsors a compelling interest in reducing fraud in electoral processes.

Fraud can occur in many different ways but is typically difficult to observe when underway. Perpetrators usually seek to disguise it when it occurs in democratic settings, because it involves illegal acts subject to prosecution and punishment.[1] Election fraud may take place at any point during the process of electing new political representatives, from the pre-electoral period through aggregation of the final votes. Our study's focus lies with events in the polling station on the day of the election. On election day, the voting process can be manipulated by allowing unregistered or ineligible persons to vote, double voting, ballot swapping, voter intimidation, stuffing the ballot box, and disposing of the contents of the

[1]This contrasts with electoral authoritarian regimes, where election authorities have been known to cheat openly (**?**).

ballot box — to name only some of the myriad and inventive ways that historically have been used.

In the context we study, low-level activists and operatives associated with both major political parties had been implicated in election fraud since the introduction of democratic elections in Ghana in 1992. Fraud prevention is one major goal of the election observation that occurs under the auspices of Ghana's Coalition for Domestic Election Observers (CODEO), a highly-reputable umbrella organization that has been observing elections since 2000 (**?**). Election observation is effective in reducing fraud when the observer organization is credible and legitimate, and is able to call on the press, the international community, and government security forces and legal bodies when fraud is observed. Thus, election observation exposes, publicizes, and denounces activities related to fraud, leaving other bodies to carry out legal enforcement. In practice, election observation is limited by the number of polling stations that can be placed under observation and by its inability to prevent fraud that might appear in the process of voter registration.

Because of the inherent limitations to direct election observation, the Election Commission of Ghana (EC) introduced biometric verification in 2012. The main goal of biometric verification is to reduce fraud in the voter registry (**?**). Biometric verification machinery (BMV) prevents individuals from registering under multiple identities or at multiple polling stations, thereby reducing double voting. During the election itself, biometric identification machines prevent unregistered voters from casting a vote and prevent registered voters from voting more than once.

If biometric verification machines fail to operate, this can facilitate election fraud in a number of possible ways. First, if voting is allowed to continue on the basis of paper-based verification, persons in line to vote may deliberately vote more than once by impersonating other registered voters. Second, if poll workers consult obsolete or inaccurate paper registries in the event of machine breakdown, currently unregistered voters whose names appear on old

registries may be allowed to vote or individuals who were previously illegally registered under multiple identities may vote more than once. Third, breakdown of identification machines may divert the attention of poll workers and allow persons to stuff the ballot box using ballots prepared in advance in favor of a particular candidate or party. All or any of these forms of fraud could affect election outcomes.

Because biometric identification machine breakdown can permit various types of election fraud, individuals or groups who wish to commit fraud have incentives to interfere with and sabotage the operation of this technology. In the context in which we work, all the major political parties are permitted a representative inside the polling station; these individuals have particular incentives to commit fraud. Because they have access to election equipment, some may take advantage of inexperienced election personnel or weak election administration to disrupt the operation of the biometric verification equipment. One way to do this is by pilfering spare batteries that are on site to keep the machines running. Another is deliberately sabotaging the BMV by placing it in direct sunlight, where it is more likely to overheat and then fail to operate.

In the 2012 Ghanian general elections, election observers deployed by CODEO remained on site for the entire day of voting but were deployed to only 4,500 of the country's 26,000 polling stations. In these elections, a newly-purchased biometric identification machine was sent to every polling station in the country, where temporary EC staff were entrusted with its operation. CODEO's observers reported whether the equipment operated properly throughout the day, as well as on all the other aspects of election administration and voting that had reported on in the past. Partnering with CODEO, our field experiment randomized the placement of observers in 60 constituencies in four of Ghana's ten regions. Because we randomize observer placement, we can assess the following hypothesis:

H1 Breakdown: Biometric verification machines break down less often in the presence of an election observer.

We contend that breakdown of identification machines occurred randomly only in some polling stations; in others, it was deliberately instigated or encouraged. Because machine breakdown thus did not always take place randomly, we cannot make causal statements about its effects. However, given that biometric identification machines are designed to prevent election fraud, we should observe the following:

H2 : Fraud should be more prevalent where machines break down.

To study this second hypothesis, we use different analytic techniques than we use to study the first. We use experimental evidence to establish that the presence of an election observer reduces the frequency of biometric verification machine breakdown. This result suggests the interpretation that biometric identification machines may have been deliberately tampered with or their breakdown encouraged in order to facilitate election fraud. To explore the second hypothesis, we next use observational evidence to document patterns in the relationships between incentives and opportunities that are consistent with the theory that some machine breakdown was deliberately instigated in order to commit election fraud.

The Setting

Ghana is one of sub-Saharan Africa's democratic success stories. Home to a population of 25 million, the country has a competitive, stable two-party system. Before 2012, alternation of the political party holding the office of the presidency had twice occurred (2000 and 2008) since adoption of the 1992 constitution and establishment of the country's Fourth Republic. The two major parties — the New Patriotic Party (NPP) and the National Democratic Congress (NDC) — enjoy support from roughly equal numbers of voters, together claiming more than 95 percent of the vote. The two parties hold all of the country's 275 parliamentary constituencies. In the preceding 2008 presidential elections, the NDC won the executive with a margin of $40,000$ votes out of an electorate of 14 million, illustrating the highly competitive nature of national politics. Electoral violence is relatively rare, voter

turnout is high, and the NDC and NPP exhibit modest but genuine programmatic differences as well as partially distinct social bases of support (**?**).

The president is elected by majority vote from a single, nationwide district. The country's unicameral parliament is elected from 275 single-member constituencies, where the main levels of party organization are located. Elections are held simultaneously for parliament and presidency. Partisan competition is not evenly distributed across the country nor its ten regions; each party has stronghold areas (**??**). We include constituencies from stronghold regions of each party in our sample as well as from regions that are highly competitive.

Elections in Ghana have been systematically observed by the Coalition of Domestic Election Observers since 2000. The organization recruits and trains professionals — typically school teachers and college students — in neutral, non-partisan observation of the electoral process. Thanks to their professions, observers benefit from high status in their communities; for this reason, CODEO assigns observers to polling stations in their home areas, where observers are likely to be personally known and to enjoy community respect. CODEO itself is nationally well known, with a strong public reputation for its work in improving electoral integrity. Its observers are recognized and accredited by the Electoral Commission of Ghana and have the legal right to enter and observe election proceedings. Each CODEO observer is assigned a single polling station for the whole election day, including the public counting of votes that occurs at the end of day. Polling places selected for observation are not identified publicly in advance of the election, meaning that officials and voters at every polling station may realistically anticipate an observer. Observers are distinguishable by uniforms and identifying paraphernalia (tee-shirts, hats, etc) and carry official accreditation materials with them.

The thorough and standardized training of election observers includes instructions to observe the EC mandate and not to interfere in election proceedings. CODEO's official

training manual opens with explicit instructions not to interfere in any aspect of the voting processes. The manual's first two rules and regulations are that "An Observer shall not offer advice or give direction to or in any way interfere with an election official in the performance of his or her duties" and "An Observer shall not touch any election material or equipment without the express consent of the Presiding Officer at the polling station or the Returning officer at the constituency center. Observers may not involve themselves in the conduct of the election" (**?**, p. 6). Observers are trained to contact constituency-level CODEO supervisors if election materials, such as ballots, are needed. Observers record administrative or other irregularities on incident forms. For the 2012 elections, CODEO's observers were also trained to use electronic short-message service (SMS) to report irregularities to a national data center. If an incident is serious, CODEO has communication structures in place to immediately alert appropriate legal and security officials. CODEO also releases press statements throughout election day and its Accra election headquarters serves as a major locus of public information about the process. Deliberate election malfeasance committed in front of a CODEO observer is likely to be reported nationally and to elicit a speedy response from government security forces.

During the voting process, the CODEO's observer usually distances herself from other individuals allowed into the polling place. These include the presiding officer, a biometric verification officer, two ballot issuers, a security officer, a representative designated by each of the major political parties, the media, and citizens in the process of voting **?**. No one else is legally permitted to enter the polling station. Polling stations are usually outdoors but are clearly demarcated.

Despite two decades of election observation, fraud was known to have occurred regularly in elections in Ghana. Perhaps thanks to the very effectiveness of election observation during voting, fraud appears to have been especially marked in the pre-election phase, which is also observed by CODEO but less extensively. Implausibly large numbers of names ap-

peared on the voter rolls in the aughts (**?**). Earlier experimental research confirmed extensive registration fraud (**?**).

The decision to adopt biometric identification for voters for the 2012 elections was precipitated precisely because voter registration fraud was believed to be widespread **?**. Using biometric markers, such as fingerprints, that are almost impossible to counterfeit, biometric identification machines authenticate the identity of the individual (**?**). Biometric identification is generally especially useful where governments have not previously established reliable or complete paper-based identification systems for their populations (**?**). Thanks to their supposed in-built capacity to prevent or substantially reduce fraud in the distribution of government allocations or services, biometric identification systems are already in widespread use globally for voter registration. As of early 2013, 34 of the world's low and middle income countries had adopted biometric technology as part of their voter identification system (**?**) and 25 sub-Saharan countries have held elections with biometric voter registers in place (**?**, p. 5). Despite the obvious difficulties in counterfeiting biometric markers, studies of biometric authentication systems have questioned whether they are tamper-proof in the real world, however (**?**).

Ghana's Electoral Commission purchased biometric identification machines from a suite of foreign vendors for the 2012 elections. The entire electorate was reregistered using biometric markers (ten fingerprints) in a six-week period in spring 2012. New voter identification cards were issued featuring head shots. Reregistration was effective in identifying $8,000$ double registrations, of which $6,000$ were judged intentional (**?**). Just prior to December's elections, verification machines were delivered to all 26,000 polling stations in the country. The EC also deployed another 7,500 backup machines in the event of equipment failure. Spare double-A batteries accompanied each machine. Only persons whose identities could be verified biometrically would be legally permitted to vote on December 7.

According to CODEO's reports, approximately 19 percent of polling stations experi-

enced a breakdowns of the verification machine at some point during the day (**?**).[2] Breakdowns appear associated with battery overheating and exhaustion; attempted battery replacement caused the machines to freeze up, reportedly for a minimum of two hours (**?**). The machines had not been properly programmed; in addition, the temporary EC personnel responsible for their operation had not been adequately trained (see (**?**, p. 91) and also (**?**, p. 361)). Batteries were supposed to be changed on a five hour schedule, a schedule which EC personnel did not always observe (**?**, p. 110). Breakdowns thereby delayed voting.[3] By noon, Ghana's President, John Dramani Mahama, appealed to the Electoral Commission to allow individuals with valid voter ID cards to vote at polling stations where biometric verification machines were not functioning.[4] The Electoral Commission officially rejected this suggestion, instructing local officials to permit voting to continue into a second day where necessary instead.[5] This occurred in a few hundred polling stations (**?**).

Research Design, Sample Selection, and Measures

The results reported here follow on those reported in **?**, which documented an experimental study of the impact of domestic election observers on election fraud and violence. That study involved collaboration with CODEO and random assignment of election observers to $1,292$ of Ghana's $26,000$ polling stations in the 2012 general elections. It collected data from these $1,292$ stations and from an additional randomly selected $1,000$ control stations. Identical information was collected from polling stations with and without observers. (Details appear in Appendix B.) Using the same data generated by the experiment reported

[2]In our sample, we find machine breakdowns in 25 percent of polling stations but in 17 percent of polling stations with a CODEO observer. The CODEO figure reflects information collected only from the latter, so our sample result is approximately the same as the national figure for observed polling stations.

[3]Voting was also delayed when the machines failed to authenticate fingerprints, for instance in cases when they had been worn down due to hard manual labor (**?**). For this and related reasons, the election resulted in a large number of rejected ballots (more than two percent of those cast) **?**.

[4]"Let people with valid IDs vote; verification or not — Prez Mahama," myjoyonline.com, (2012, December 7, 15:33 GMT), *http://politics.myjoyonline.com/pages/news/201212/98391.php*; accessed 4 June 2014.

[5]However, there are persuasive reports that voting was permitted in some polling stations even when the biometric verification machine was non-operational (**?**, p. 362).

in **?**, in the present study we consider the effects of election observation on the technology intended to combat fraud and improve election administration.

Sampling and Treatment Assignment

We implement the project in four of Ghana's ten regions.[6] Almost half (46.5 percent) of Ghana's population resides in sampled regions. More relevant is the fact that the party system is similar in the six regions not included in our sample to the four that are. So although the regions where we work were not selected to be statistically representative of the country, we have no reason to believe results would differ significantly had our sample been national.

We randomly sample 60 (out of 122) political constituencies from the four regions.[7] We construct the sample as follows. First, each region is assigned a target number of sample constituencies based on its proportion of the total 122 constituencies.[8] Since each region's number of electoral constituencies is determined by the Electoral Commission on the basis of population, this means the number of constituencies included in the sample from each region makes the regional sample proportional to population.

To select constituencies within regions, we block on electoral competitiveness and urbanization. We construct a sample with roughly equal numbers of constituencies that vary on these characteristics. We block on electoral competitiveness because we hypothesized that election fraud would vary with competitiveness. To generate our indicator of constituency-level electoral competition, we use data from the prior (2008) presidential elections.[9] We define a constituency as *competitive* if the vote margin between the top two presidential

[6]For logistical reasons, we sample only in Ghana's south. We exclude the Greater Accra region, the location of Ghana's capital, because we anticipated that international election observers might focus on the easy-to-reach polling stations there and that their presence could contaminate the treatment.

[7]Sample size was determined on the basis of power calculations and logistical constraints.

[8]For example, the largest region we study is Ashanti, which has 47 constituencies, or about 38 percent of the 122 total. We sample 23 constituencies in Ashanti: 23 is approximately 38 percent of the total sample size of 60.

[9]In order to define the sample prior to the 2012 elections, we necessarily had to define competitiveness on the basis of historical data.

candidates was less than 10 percent and as uncompetitive otherwise. Constituencies that experienced alternations in the party winning a majority in the 2008 presidential elections had a 2004 average margin of victory of 12 percent. Therefore, a 10 percent margin is, in the context in which we operate, easily reversible. We block on urbanization because of the hypothesized relevance of polling stations density to the strategic relocation, or spillover, of election fraud; spillover is not a focus of the present paper, however. We report results in the present paper using this blocking variable although it is not theoretically relevant to the project at hand. We code urban and rural constituencies using a measure of polling station density. We define as *urban* constituencies with a higher-than-the-median number of polling stations per square kilometer (where the median in our sample is 0.14 polling stations per square kilometer) and rural as those with lower-than-the-median.

With blocking variables in place, constituency sampling was performed as follows. Within each region, all constituencies are coded as competitive or stronghold and as urban or rural. We select a random sample of constituencies from each of the four possible combinations (competitive-urban, competitive-rural, stronghold-urban, stronghold-rural) such that the total number sampled from each region equals its target number. To the extent feasible, we sample equal numbers of constituencies within regions from each of the four conditions.[10]

Our units of analysis are individual polling stations, which are nested within the 60 constituencies in our sample. We randomly sample 30 percent of the polling stations in each constituency. We then randomly assign each polling station to either treatment (observer) or control (no observer). Appendix B further details the experimental design. In Appendix C, we provide evidence that treated and control polling station areas are comparable across developmental, political and ethnic characteristics. In our analyses, we report average treatment effects.[11]

[10]In some regions, equal numbers of competitive and stronghold constituencies do not exist, narrowing our choices.

[11]The project experienced no issues with compliance in administering treatment.

Measuring Machine Breakdown

We gathered data at treated and control polling stations on election day using a questionnaire (reproduced in Appendix A). It included the question "Did biometric verification machine fail to function properly at any point in time?"[12] Possible responses were "Yes"/"No." We use responses to measure breakdowns of biometric verification machines at the polling station level. We lack information on whether machines broke down repeatedly, how long breakdowns lasted, what triggered them, or what was done in response.

Enumerators gathered information on machine malfunction in treatment and control areas as follows. At treated stations, CODEO observers collected the information we analyze as part of their official assigned activities. Accordingly, at treated stations, data is a product of direct observation on the part of the data collector. At control stations, data was collected by enumerators who interviewed two party agents, each representing one of the two major political parties, after the polls closed. (We have earlier noted that each party is allowed to designate an official representative as an observer and that individual is permitted to remain in the polling station.) To avoid "observing" control stations, we could not send enumerators to control stations during the election process itself. Enumerators were provided training identical to that of CODEO observers, were accredited by the Election Commission as observers, were uniformed with observer-specific shirts, and used identical data collection questionnaires.

This variation in the data collection processes raises the concern that it may drive the observed causal relation that we report between election observers and machine failures. We have five reasons to believe our data is valid despite the differences in data collection between treatment and control stations.

First, reporting differences are likely to bias results *against* control stations. CODEO

[12]The questionnaire also asked whether a biometric verification machine was present at the polling station. Ten polling stations in our sample did not have machines, and we drop these from the analysis.

observers are trained to document all irregularities that occur at their assigned station, where they remain for the entire day. It seems likely that official observers more often document events such as machine malfunction than do party representatives; in addition, observers record events in real time whereas our enumerators asked party representatives to recollect events that had taken place during the preceding ten to twelve hours. CODEO observers would thereby probably record *more* violations of election integrity than would party agents. In contrast, our data show rates of machine breakdown that are twice as high in control stations as in those observed by CODEO. This suggests that these results reflect genuine not reporting differences.

Second, on other measures of electoral administration, just as we might expect, on-site CODEO observers reported more violations than enumerators sent to control polling stations at the end of the day. These violations include whether election officials were present when the polls were supposed to open, whether voting began on time, whether election materials were missing, and whether party and other officials signed the vote count tabulations. [LUKE: VERIFY ITEMS.] In Appendix **??**, we present data examining ten [LUKE VERIFY ??] measures of administrative integrity of the electoral process and test for significant differences between treated and control polling stations. [LUKE, IS THIS HOW WE WANT TO PRESENT THIS?] These measures were collected using the questionnaire just discussed. If CODEO observers are more assiduous in documenting administrative irregularities, we expect that where there are statistically significant differences between treated and control polling stations, they will show higher rates of administrative irregularities in treated stations. The results of the comparison of means corroborate this. The only one of ten [LUKE VERIFY] variables measuring election administration on which control station enumerators report higher levels is machine malfunction — and the levels are double those at treated polling stations. This suggests to us that reports by enumerators on this particular variable are likely to be accurate.

Third, reporting differences cannot explain the result (presented below) that electoral competition is statistically associated with machine breakdown. CODEO observers are randomly assigned to polling stations *within* constituencies, and every sampled constituency contains both treated and control polling stations. Data gathered by enumerators at control stations should not be correlated with constituency-level variables. In addition, the relationship between electoral competition and machine breakdown is almost identical when we subset the sample into treated and control stations: differences in data collection methods between treated and control stations do not affect this finding. This again suggests that data collection methods did not affect the quality of data collected.

In order to use all the available data, the main analysis reported in this paper includes data from control polling stations that was collected even from a single individual in instances where it proved impossible to collect data from a second person. In Appendix D, we reanalyze using a slightly reduced dataset drawn from only those control polling stations where enumerators collected identical information from at least two different individuals. The reduced dataset drops a maximum of 59 control observations where enumerators collected information from only a single person. [LUKE VERIFY THE NUMBER] In the reduced dataset, identical information was collected from two respondents, who were typically affiliated with different major political parties. Thus, they had competing political incentives. Even using this more conservative dataset, the results reported in the paper continue to hold. [LUKE EXPAND?]

Fifth, data for one of our measures of election fraud (ballot stuffing) was collected after the polls closed in identical fashion by CODEO observers and the enumerators sent to control polling stations. This measure therefore does not exhibit any difference in data collection methods. Rates of ballot stuffing are twice as high in control as in treated polling stations, again validating our data.

Measures of Election Fraud

We construct indicators of election fraud that rely on objective information gathered from sample polling places on election day. By law, ballots must be counted in public at each polling station after the polls close. This makes it possible to collect polling station level information before it is aggregated (and potentially tampered with) at higher levels.

Using data collected during the election process, we construct three measures of fraud. Our first measure, the *overvoting rate*, is the number of votes cast in the presidential race that is above the number of voters officially registered at the polling station. We convert this measure into a share by dividing by the number of registered voters, taken from the Electoral Commission's official figures. Each voter is legally allowed to vote only at the polling station where registered. Overvoting is a marker of potential fraud since it suggests that unregistered voters cast ballots, that double voting occurred, or that vote counts were artificially inflated in some other way.[13] The EC figures on registered voters were released prior to election day. The number of valid votes cast in each polling station is reported on an official form that is filled out at the close of day; we collected these figures in the sampled 2,310 polling stations. The rate of overvoting in these polling stations ranges from 0.3 percent to more than 250 percent. Of the sample polling stations, $1,845$ reported vote totals that did not exceed the number of registered voters, and are therefore coded zero for overvoting.[14] Because there could be fraudulent votes cast that did not cause the total number of votes to exceed the number of registered voters, the overvoting measure understates fraud.

The second measure of fraud we study is what we call *registry discrepancy*. The official number of registered voters reported by the Electoral Commission and the number of voters

[13]Double voting is punishable in Ghana by a jail term, and election officials who permit can be prosecuted (**?**, p. 44).

[14]We coded one polling station in the Ho West constituency as zero for overvoting even though turnout there was over 700 percent. A clear outlier in terms of turnout, this polling station received hundreds of voters from a nearby polling station after its biometric machine broke down. Not making this correction increases the magnitude of the treatment effect and shrinks the p-value but the results we report are not driven by this decision.

on the paper rolls at the polling stations differed in [LUKE GIVE NUMBER] cases, and where these two differed, the paper rolls show systematically higher figures than the EC numbers. Discrepancies may arise from the delivery of false voter rolls to the polling stations or from deliberate manipulation of the voter rolls by persons at the polling station.[15] They may also arise due to administrative error; for instance, if the paper rolls had not been properly updated to incorporate the biometrically-based reregistration that had occurred in the spring of 2012. Differences between the official and the paper rolls were unusually common in polling stations where we identify overvoting, lending plausibility to the interpretation that these differences facilitated election fraud. Indeed, in control stations, local voter rolls are on average 21 voters larger than the Electoral Commission's figures, whereas in treated stations the local voter rolls are only 2 voters larger.[16] In other words, registry discrepancies are significantly greater in polling stations without election observers present.

We code any numerical difference between Electoral Commission figures and those reported by the polling station as a *registry discrepancy.* This takes the value of one; matching figures from the EC and the local rolls are coded zero. In order to account for the possibility that small differences between the EC figures and the figures from the paper registries do not indicate intent to commit fraud but rather are a product of minor administrative error, we also demonstrate in the results section that results are robust to a range of cutoffs for coding *registry discrepancies* as a one. We show that coding *any* discrepancy in the number of registered voters produces results similar to limiting the measure to cases where figures differed by more than 50. We also reproduce our analyses using the continuous measure of differences between the number of registered voters according to the local registry and the official EC figure in Appendix G. Because the continuous nature of this variable allows intentional, administrative, and measurement errors to all inflate it, it exhibits high variation.

[15] *Overvoting* may capture some of this fraudulent behavior, but often turnout will not exceed 100 percent.

[16] The difference between these means has an associated p-value of 0.004 when accounting for clustered standard errors.

As a result, the results reported in Appendix G are consistent with the rest of our findings but are less precisely estimated.

Registry discrepancies, although intimately related to *overvoting*, capture a slightly different fraud mechanism. They highlight the willingness of local officials to change the voter registry rather than indicating a fraudulent outcome (namely, more ballots cast than registered voters). Verifying this distinction, our data show that 78 percent of polling stations with *overvoting* were also locations where there was a *registry discrepancy*, but only 15 percent of the locations where there was a *registry discrepancy* exhibit *overvoting*. We interpret this difference as suggesting more frequent intentions to commit fraud than are captured by the measure of overvoting.[17]

Our third fraud measure captures whether the presidential ballot box appears to have been stuffed. The *ballot stuffing* measure takes a value of one if more ballots were discovered in the ballot box than the number of voters known to have cast ballots and zero otherwise. The data were collected using the questionnaire; enumerators responded "Yes"/"No" to the question "Were more ballot papers found in the presidential ballot box than voters who cast ballots?" (See item CE on the questionnaire in Appendix A.) Because votes are counted at each polling station in public at the end of the day, both observers and enumerators had direct access to this information.

We analyze this suite of indicators separately because they capture different types of irregularities but represent a similar underlying pattern of electoral malfeasance: attempts to alter the electoral outcome through election day vote rigging. They are not the only possible indicators of fraud in Ghana's polling stations. Excess turnout may also proxy election fraud **?**. In this paper, we use indicators that are more obviously and consistently related to fraud

[17]Both *overvoting* and *registry discrepancy* use the number of registered voters reported by the Electoral Commission as a benchmark. This number was fixed before the election took place. As a validation of both of these measures, we test for any treatment effect of the observers on the number of registered voters according to the Electoral Commission. The p-value for the difference in means is 0.99, validating this figure as a benchmark.

than turnout.

[LUKE, I HAVE REVISED THROUGH HERE. I DO NOT KNOW WHAT TO DO WITH THE NEXT SECTION ON COMPETITIVENESS YET. I HAVE NOT TOUCHED THE RESULTS SECTION, THE CONCLUSIONS, OR THE APPENDICES. THUS THE RESULTS STILL HAVE TO BE REALIGNED WITH THE TWO HYPOTHESES]

Measuring Electoral Competitiveness

Our first hypothesis posits that electoral fraud will be concentrated in electorally competitive areas, where the value of an additional vote for a parliamentary candidate will be highest. Since Ghana's president is elected in a single nationwide district, the incentives for parties to win votes in the presidential contest are uniform across parliamentary constituencies. We therefore cannot test whether election competition affects election fraud measuring competitiveness in the presidential race. We instead test this hypothesis with data on electoral competition in the constituency-level parliamentary elections. We label this variable *margin*.

The rationale behind margin is as follows. Political parties in Ghana, as elsewhere, seek to maximize seats in the legislature, as well as to win the presidency. If electoral competition creates incentives for fraud, we are likely to observe more fraud in electorally competitive parliamentary constituencies. The degree of electoral competition in the parliamentary elections is therefore likely to be the primary influence shaping the incentives of party organizations to commit fraud.

We assume a close connection between fraud in the parliamentary and presidential races, but one that differs slightly according to the type of fraud. If an MP's supporter votes twice, that person will do so for both the MP and the presidential candidate for the same party, even though the geographically-specific incentive for fraud relates to the parliamentary race. Each voter normally places two ballots into the two boxes that are set up in each polling station: the parliamentary and the presidential. It is routine for the voter

to vote for both offices. However, if an MP's supporter stuffs the parliamentary ballot box, he may or may not also stuff the presidential ballot box. Doing so requires separate illicit actions in the polling station, one at each box. Theoretically, we would expect to observe more ballot stuffing in the parliamentary races than the presidential. [DO WE NEED TO MOVE INDICATOR-SPECIFIC HYPOTHESES TO A SINGLE PLACE AND PRESENT IN A MORE ORGANIZED FASHION?]

Political parties in Ghana are organized hierarchically, with relatively independent constituency-level organizations guiding the campaign operations of parliamentarians (**?**). These constituency-level organizations are in the first instance creatures of the member of parliament. Party organizational structures thus permit individual members of parliament to weigh the potential costs and benefits of organizing their followers to engage in electoral fraud, and to encourage or permit their followers to do so if they judge it opportune.

We create a continuous measure of the parliamentary vote margin to capture the degree of electoral competition in each constituency. (We remind readers that the measure of electoral competition used in blocking was drawn from the prior (2008) presidential not parliamentary race.) In Ghana, there is very little ticket splitting and a party's margin in the parliamentary race is almost identical to its constituency-level vote share for the presidency.[18] Thus, it hardly matters empirically whether we use the parliamentary margin or presidential vote share as a measure of competitiveness. Theoretically, however, we contend that differences in constituency-level incentives for fraud involve parliamentary races since all votes for president county equally across the country. For that reason, we select to use a measure of margin created for candidates to the parliament.

The average vote margin in the sampled constituencies is 0.31 percent and the median is 0.19. Smaller values on the margin variable indicate higher levels of electoral competition

[18]For the NDC, the correlation in parliamentary and presidential vote shares across constituencies in 2012 is 0.98 and for the NPP it is 0.97.

and higher values indicate the reverse. The constituencies in our sample display a large range of values on this variable, verifying that parliamentary competitiveness is highly variable. There are a noticeable number of constituencies in our sample [GIVE MORE PRECISE INFO] where the margin in 2008 was close to zero, indicating extremely tight parliamentary races. Broadly speaking, nearly 40 percent of parliamentary constituencies house swing seats **?**.

Results

Table 1 presents descriptive statistics about direct rates of machine breakdown and the three measures of electoral fraud in our experimental and blocking conditions. As the data presented in the first row documents, a quarter of the polling stations in our sample experienced machine breakdowns. The three indicators of electoral fraud were much less common. Not shown are the correlations among these four variables, although the patterns are interesting. [PERHAPS WE SHOULD SHOW THEM?] Although overvoting may have been generated by differences between the EC and the paper voter rolls (captured by a registry discrepancy), the two measures are only weakly correlated ($r = 0.30$). Registry discrepancies may induce attempts at overvoting that did not push turnout over 100 percent, which is all that overvoting captures. Nor are overvoting and ballot stuffing correlated in our sample of polling stations ($r = 0.03$). These two types of irregularities almost never occurred in the same polling stations, appearing instead as substitute types of fraud. One possible explanation for this is that these two irregularities may have been committed by different types of individuals and in different ways. Overvoting may have occurred with the complicity of the presiding officer, when individual registered voters were permitted (or encouraged) to vote more than once, perhaps as party activists escorted them back into line after they had voted. Ballot stuffing, by contrast, may have occurred when the presiding officer was inattentive (either deliberately or when distracted), allowing others on the scene

to add more ballots to the presidential box.[19] The data show roughly similar rates of polling station overvoting and ballot stuffing in our sample, both of which are much less frequent than registry discrepancies.

[19]These two types of irregularities were grouped together by the NPP when the party petitioned Ghana's Supreme Court to nullify the election results in a lengthy post-election court case. The NPP labeled both "overvoting." Although the petition was ultimately rejected, four of the nine Supreme Court justices ruled that the NPP's petition was valid, suggesting that our measures of fraud are broadly in line with what legal authorities in Ghana believe true.

Table 1: Descriptive Statistics of Machine Breakdown and Measures of Fraud

	(1) Full Sample	(2) Treatment	(3) Control	(4) Competitive	(5) Stronghold	(6) Urban	(7) Rural
Machine breakdown	0.234	0.172	0.354	0.276	0.201	0.249	0.216
	(0.423)	(0.378)	(0.479)	(0.447)	(0.401)	(0.433)	(0.412)
Overvoting rate	0.027	0.015	0.051	0.035	0.022	0.024	0.032
	(0.194)	(0.152)	(0.255)	(0.210)	(0.182)	(0.180)	(0.209)
Registry discrepancy	0.177	0.134	0.255	0.198	0.163	0.192	0.161
	(0.382)	(0.340)	(0.436)	(0.399)	(0.369)	(0.394)	(0.368)
Ballot stuffing	0.043	0.032	0.062	0.056	0.033	0.043	0.042
	(0.202)	(0.175)	(0.242)	(0.230)	(0.179)	(0.203)	(0.201)
Observations	2047	1281	766	864	1183	1081	966

Notes: Standard deviations in parentheses. Ten polling stations without biometric verification machines removed from sample. Polling stations where enumerators recorded different responses from two party officials on relevant variables are dropped. [IS THIS DIFFERENT THAN OUR DISCUSSION OF ROBUSTNESS CHECKS WHERE WE USE THE REDUCED DATASET? HOW MANY PS ARE DROPPED HERE?] In Appendix D, Table D.1 we report the same table but only include observations where enumerators recorded identical answers from *both* party officials.

The data reported in columns 2 and 3 provide evidence about the causal effects of election observers. Machine breakdown occurred at 17 percent of polling stations with a CODEO observer present but at 35 percent of those without an observer. This is a very large fraction of polling stations and implies an increase in the rate of breakdown of around 100 percent when an election observer was not present. The remaining columns present rates of machine breakdown in blocking environments: competitive and uncompetitive as well as urban and rural constituencies. We find that machines break down more frequently in electorally competitive constituencies: 28 percent of polling stations in competitive areas experience machine breakdown compared with 20 percent in uncompetitive constituencies. (To repeat, here competitiveness is a dichotomous blocking variable drawn from 2008 presidential not parliamentary election data; see the discussion above.) Rates of breakdown are also 3 percentage points higher in urban than in rural areas. The data depict an election in which biometric verification machines broke down more often when an election observer was not present, in competitive constituencies rather than party strongholds, and in urban rather than rural areas.

The next three rows present the same information for the three indicators of fraud that we analyze. Registry discrepancies, which we analyze as a proxy for the intent to commit fraud, characterizes 18 percent of the full sample, and the rate is double at polling stations without an observer compared with those with an observer. However, objectively fraudulent outcomes — proxied by overvoting and ballot stuffing — was much less common than registry discrepancies, occurring at only 3 to 4 percent of polling stations. These too were more likely to occur in the absence of an election observer. Overvoting, ballot stuffing, and registry discrepancies were all more likely to occur in competitive than in stronghold constituencies and registry discrepancies are noticeably more common in urban than rural locations. The data on fraud is thus consistent with our hypotheses that fraud increases when election observers are not present and with election competition.

Causal Impact of Election Observers on Machine Malfunction and on Election Fraud

In Table 2, we report average treatment effects (ATE) for the impact of electoral observers on machine malfunction and the three indicators of fraud. These results come directly out of the experimental design. We use OLS regression analysis in order to incorporate the two blocking variables as covariates and also to display easily interpretable quantities.[20]

[20]In Appendix E, Table E.1, we report logistic regressions on the three dichotomous outcomes (*breakdown*, *registry discrepancy*, and *ballot stuffing*). Results are substantively unchanged.

Table 2: Average Treatment Effects of Election Observers on Machine Breakdown and Fraud

	Breakdown		Overvoting rate		Registry discrepancy		Ballot stuffing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Election observer	-0.182^{***}	-0.181^{***}	-0.036^{***}	-0.037^{***}	-0.122^{***}	-0.121^{***}	-0.031	-0.031
	(0.037)	(0.038)	(0.010)	(0.010)	(0.025)	(0.026)	(0.027)	(0.028)
Competitive		0.078^{*}		0.013		0.039		0.023
		(0.039)		(0.009)		(0.023)		(0.019)
Urban		0.030		-0.008		0.026		0.001
		(0.036)		(0.009)		(0.022)		(0.018)
Constant	0.354^{***}	0.304^{***}	0.051^{***}	0.050^{***}	0.255^{***}	0.224^{***}	0.062^{*}	0.052
	(0.035)	(0.039)	(0.010)	(0.012)	(0.025)	(0.028)	(0.026)	(0.027)
Observations	1,888	1,888	1,864	1,864	1,973	1,973	1,987	1,987
R^2	0.042	0.051	0.008	0.009	0.023	0.027	0.005	0.009

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with robust standard errors clustered by constituency in parentheses. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials on relevant variables are dropped. In Appendix D, Table D.2 we report the same table but only include observations where enumerators recorded identical answers from *both* party officials. Logistic models for dichotomous outcome variables are reported in Appendix E, Table E.1. Results are unchanged in both alternative specifications.

For each outcome variable — machine *breakdown*, *overvoting rate*, *registry discrepancy*, and *ballot stuffing* — we report two specifications. The first column contains unadjusted ATEs whereas the second column adjusts for blocking covariates. Both columns report specifications that use robust standard errors clustered at the constituency level.[21] With or without blocking covariates, results are substantively identical. Column 1 shows that election observers have a negative and statistically significant impact on machine breakdown. The size of the effect is unaltered when we incorporate blocking variables into the model, as shown in Column 2. The estimated average treatment effect is very large, with observers reducing rates of machine breakdown by 18 percentage points.

Results reported in columns 3 through 6 show that election observers also have negative and statistically significant impacts on all three measures of election fraud. Columns 3 and 4 show that the rate of overvoting is 3.6 percentage points lower in observed polling stations and Columns 4 and 5 show that the prevalence of registry discrepancies is about 12 percentage points lower in observed polling stations.[22] Lastly, although observers have a negative impact on incidents of ballot stuffing, the result is not statistically significant.

The effect of election observers on biometric machine breakdown and the three markers of fraud is causally identified as a result of the random assignment of observers to polling stations. However, because we collected data on a wide variety of outcomes and have tested several of them, it is possible that the statistically significant results just reported are ar-

[21] Using nonrobust standard errors yields the same results, although the p-values are smaller, the depressive effect of observers on ballot stuffing becomes significant, and the effect of competitive polling stations on markers of fraud is more clearly positive. This specification can be found in Appendix E, Table E.2.

[22] The reported operationalization of *registry discrepancy* defines any difference between the number of registered voters reported by the Electoral Commission and the rolls at the polling stations as intent to commit fraud. However, it is possible that some discrepancies arise due to simple transcription errors or minor modifications to the voter rolls. For that reason we recode *discrepancies* as one when the difference between the Electoral Commission and the polling station figures is greater than or equal to some cutoff, ranging from one, as it is now, to 500. For example, a cutoff of 100 would indicate that only when the two figures differ by greater than or equal to 100 do we indicate a registry discrepancy. As reported in Figure G.1, the effect of observers on *discrepancies* is negative and statistically significant at the 0.05 level whether the cutoff for fraud versus administrative error is coded as any difference (1), to a very large difference (357).

tifacts of the number of specifications run. To ensure that significant results are not due to multiple comparisons, we use a Holm correction to adjust the p-values. First, we collect over 43 outcomes from the survey instrument (see Appendix A), including some of our own construction. Then we run OLS without blocking covariates and with robust clustered standard errors at the constituency level. Finally, we collect and order the p-values. The analysis requires that we inflate the p-values by dividing by the total number of comparisons made minus the rank of the p-value. In essence, this penalizes the most significant result the most, and the less significant results to a lesser degree. This method strictly dominates the simpler Bonferroni correction (**?**) when it comes to ensuring that the familywise error rate, or the probability of false discovery, is at most an acceptable level of error, usually specified as 0.05.

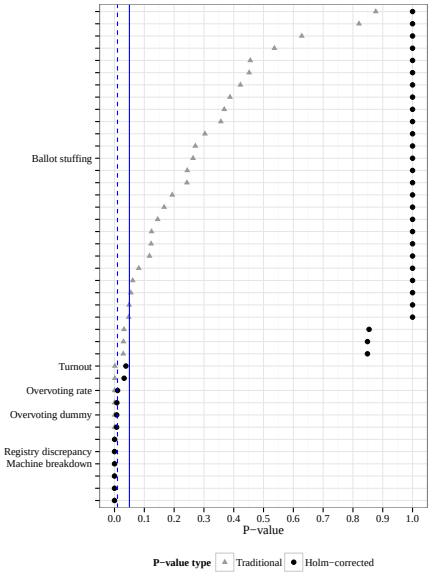
Figure 1 depicts the results of this correction. Highlighted on the y-axis are the key variables of interest and on the x-axis their associated p-value, both before and after correction. None of the results of our main analysis in Table 2 are changed and all remain significant at the 0.05 level.[23] Results are also robust to using the more stringent Bonferroni correction. For a full list of outcomes included in the Bonferroni correction, see Appendix H. This provides ample and robust evidence of a causal depressive effect of election observers on biometric machine breakdown as well as two of the three fraud measures that we employ. [LUKE: COMMENT ON BALLOT STUFFING?]

Do Partisan Incentives Increase Machine Breakdown and Election Fraud?

We now examine tests of Hypothesis 2, that partisan incentives encourage election fraud. We first consider this using the dichotomous blocking measure of competitiveness drawn from the presidential results in the prior election. In Column 2 of Table 2, we see that competitive constituencies are positively associated with machine breakdown. In our

[23]We have added a dichotomous version of *overvoting* that is coded as one if *overvoting* is greater than 0, and as zero otherwise. We also include turnout to show that there is a negative treatment effect of observers on turnout. This suggests that turnout was probably artificially inflated in control stations.

Figure 1: Holm Corrected P-Values



Note: Listed on the y-axis are the four focal outcomes, as well as turnout and a dummy variable for overvoting (1 if it exists, 0 otherwise). The dashed vertical line is at 0.01 and the solid horizontal line is at 0.05. The full list of outcome variables is reported in Table H.1.

sample, polling stations in competitive constituencies are 7.8 percentage points more likely to experience machine breakdown than polling stations in party strongholds. In addition, the estimated effect of competitiveness on the three other markers of fraud is consistently positive, although not statistically significant.

We now consider Hypothesis 2 using *margin*, whose construction is drawn from 2008 parliamentary election results. This measure more accurately captures the incentives that parties face. Table **??** contains identical specifications to those reported in Table 2 but uses the continuous *margin* variable instead of the *competitive* blocking variable. Results are largely the same; an increase in the *margin* of victory is related to a decrease of fraud using any of the four outcomes, including machine breakdown. *Margin* is now a statistically significant predictor of fewer *registry discrepancies*. The results reported in the table are consistent with the hypothesis that competitive constituencies provide incentives for parties to commit

electoral fraud. [CONFUSING TO INCLUDE URBAN AS A CONTROL. DELETE?]

Machine Malfunction and Fraud

Thus far, we have established that trained, neutral domestic election observers causally affected aspects of election administration and integrity in Ghana's 2012 general elections. This includes reducing incidents of biometric machine breakdown and various measures of election fraud.

The causal pathways that we hypothesize characterize the situation make machine breakdown a *mediating* variable, one that intervenes between the presence of an election observer and markers of election fraud. The relationships may be summarized in the following figure.

[HERE WE NEED A DAG. LABEL OBSERVER PRESENCE AS X, FRAUD AS Y, AND MACHINE BREAKDOWN AS M. X GOES TO Y THROUGH M AND ALSO ON ITS OWN.]

The absence of a standard, statistically acceptable method for untangling mediating variables makes these causal relationships difficult to investigate. There is no way to establish conclusively whether the breakdown of biometric verification machines on election day caused activities that in turn generated fraudulent outcomes. Instead, we can document patterns of correlations that suggest a causal role for machine breakdown. In this section, we undertake this exercise.

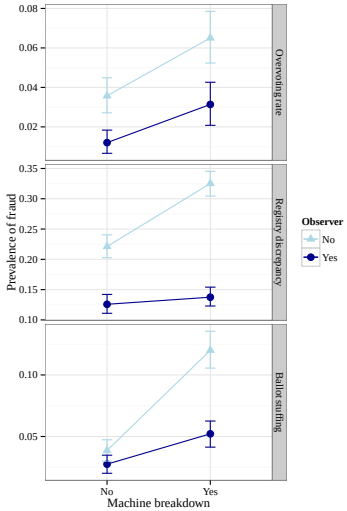
There is no administrative or technical reason that election observers should have any significant impact on the operation of biometric verification machines. CODEO observers were not instructed in the use of the machines nor were they expected to ensure their operation; indeed, as we have already reported, they were instructed in writing (as well as verbally in their training sessions) not to touch or tamper with equipment in polling stations. Thus, we see no reason to believe that observers themselves kept machines operations. Instead,

their presence appears to have affected the behavior of others at the polling station: the presiding officer, for instance, or the biometric verification officer. This raises the question of whether machine breakdown was used strategically as an opportunity to commit election fraud — by encouraging voters to double vote, for instance. To explore this, we now examine the relationship between machine breakdown and election fraud.

Figure 2 shows the various relationships graphically using bootstrapped means by group. The figure is broken into three panels, one for each measure of fraud considered in this paper. The y-axis represents the proportion of polling stations with fraud for *registry discrepancies* and *ballot stuffing* and it represents the mean level for the *overvoting rate*. The lighter line (with circles) corresponds to the control polling stations that were unobserved while the darker line (triangles) corresponds to treatment — that is, to observed polling stations. In all three panels, the bootstrapped mean level of fraud for unobserved stations is above the mean level for observed stations, indicating that observers have a depressive effect on the three markers of fraud independent of machine breakdown. Furthermore, the slope of the lighter lines is steeper in all three panels (although almost imperceptibly in the *overvoting rate* panel). This indicates that when machines break down, fraud goes to higher levels in the absence of observers. For example, around 3 percent of all stations, whether observed or not, exhibit ballot stuffing when there is no machine breakdown. When there is machine breakdown, the percentage of stations exhibiting ballot stuffing jumps to around 12 percent when there is no observer but only to 5 percent when an observer is present at the station.

These results highlight that the most extreme outcomes consistently obtain when the biometric verification machine breaks down and no election observer is posted at the polling station. Election observers have independent effects in reducing fraud when there is a fully functional biometric identification machine. The breakdowns of biometric identification machines appears important for incresing rates of election fraud, and breakdowns are most

Figure 2: Bootstrapped Mean Levels of Fraud by Machine Breakdown and Observer



Note: Machine breakdown is on the x-axis and the different colors and shapes indicate different observer statuses. On the y-axis are mean levels of fraud; this is the mean level of overvoting for *overvoting rate* and the proportion of polling stations with indicators of fraud for *registry rigging* and *ballot stuffing*. Standard errors and mean values computed using 1,000 bootstrapped samples.

strongly related to fraud in the absence of election observers.

Interpretation

We have established that election observers reduce election fraud and that their presence also reduces rates of malfunction of the equipment designed to identify voters. The research design we use shows that domestic election observers can be conclusively identified as causing these improvements in election practices.

Our results suggest that the operation of biometric verification machines may have been deliberately induced in some instances. Machine breakdowns appear to have been used strategically to increase overvoting and ballot stuffing. They occurred more frequently where registry discrepancies had laid the groundwork, since these discrepancies more easily permitted voters to vote more than once or permitted persons not legally registered to vote. These processes resulted in turnout rates above 100 percent, captured with the measure of overvoting. We also find, consistent with a theory of fraud in the context of democratic political

competition, that machines were significantly more likely to break down in constituencies that were more electorally competitive for the parliamentary seat. This corroborates that fraud increases with electoral pressures on political parties.

We can speculate about how machine breakdowns occurred and how this permitted election fraud to occur. There was a "natural" rate of breakdown, which appears to have been under 20 percent. This was the rate of breakdown when an election observer was present. It probably was chiefly due to battery exhaustion, where unfamiliarity with new equipment made it difficult to keep the machines operational. The additional breakdowns that took place when election observers were not present may have been deliberately induced by pilfering spare batteries, by exposing the machine to excessive heat or sunlight, or by rending any available backup machine non-operational. The occasional machine may have been stolen outright.[24] Breakdowns may also have been induced when presiding officers exhibited (perhaps strategically) unfamiliarity with the machines, despite the fact that temporary technical staff from the EC was supposed to be on site to keep the machines operating. Machine breakdowns could have led to confusion in the polling station, permitting ballot stuffing to occur as presiding officials were distracted trying to restore equipment. More frequently, it seems that breakdowns led some polling station officials to allow voting to continue, despite strict instructions from the EC to the contrary; this was extensively reported by the media even on election day (**??**). In the quarter of polling stations without a CODEO observer and where the paper registry had been tampered with, this facilitated double voting and voting by unregistered individuals.

Aware of some of the problems that occurred in 2012, Ghana's Election Commission subsequently upgraded the biometric machines. The subsequent (2014) upgrade included programming the machines to warn when the batteries were running out (**?**). Much of the fraud in the 2012 elections appears to have occurred when the batteries were exhausted

[24]Reported in **?**.

and the machines froze up, making reprogramming the biometric verification machines a potentially successful prevention technique. The extent of electoral fraud that our research shows was associated with biometric machine failure in 2012 was thus not repeated in the subsequent elections.

Conclusions

Our study cautions that biometric technology is susceptible to manipulation, especially in an initial large scale rollout and even in a genuinely competitive democracy. In this context, breakdown may be deliberately induced when machines are not monitored by neutral, trained election observers. The overall legal and political environment is sufficiently relaxed that political party operatives apparently feel free to take advantage of unmonitored voting to tamper with new and imperfectly designed equipment. These results carry general implications for the use of biometric identification technology. Introduction of such equipment reduces fraud, even if we cannot estimate how much fraud is prevented. However, it remains important to use the technology under the watchful eyes of independent, non-partisan and neutral observers who have no interest in perpetuating fraud and who are professionally committed to the practices of good governance. There is no technical fix to election fraud.

References

A Appendices

A Survey Instrument

Figure A.1: Data Collection Instrument Used in Sampled Polling Stations

CODEO (UCLA) OBSERVER CHECKLIST – 07 December 2012 Elections

Enumerator Name		**Constituency**	AFIGYA SEKYERE EAST
Region	ASHANTI	**Electoral Area**	AMANGOASE
Polling Station Name	WIAMOASE SAL ARMY JHS	**Polling Station Code**	F353001

ARRIVAL

AA Were any election officials present upon arrival? — Yes (1) | No (2)

SETUP

AB At what time did voting commence? — By 7:15 (1) | 7:16 to 8:00 (2) | 8:01 to 10:00 (3) | After 10:00 (4) | Never Opened (5)

AC Were you permitted to observe? — Yes (1) | No (2)

AD Was the polling station set up so that voters could mark their ballots in secret? — Yes (1) | No (2)

AE Was the polling station accessible to persons with disabilities and the elderly? — Yes (1) | No (2)

AF Was a polling agent present for NDC? — Yes (1) | No (2)

AG Was a polling agent present for NPP? — Yes (1) | No (2)

AH Which other political party agents were present? (*Tick one or more*) — None (0) | CPP (1) | GFP (2) | NDP (3) | GCPP (4) | PNC (5) | PPP (6) | UFP (7) | URP (8) | INDP (9)

AJ Were security personnel present? — Yes (1) | No (2)

AK Which items, if any, were missing (*Tick one or more*) — None (0) | Ballot Box (1) | Ballot Paper (2) | Voters' Register (3) | Indelible Ink (4) | Voting Screen (5) | Validating Stamp (6) | Endorsing Ink (7) | Ink Pad (8) | Tactile Ballot (9)

AM Did the polling station have a biometric verification machine? — Yes (1) | No (2)

AN Number of names on the voters register? *(Enter zero if no voters register)*

AP Number of names on the proxy voters list? *(Enter zero if no proxy register)*

AQ Number of presidential ballot papers? *(Enter zero if none)*

AR Were the presidential and parliamentary ballot boxes shown to be empty, sealed and placed in public view? — Yes (1) | No (2)

VOTING

BA Were people's biometric registration information verified? — Yes (1) | No (2)

BB Did biometric verification machine fail to function properly at any point in time? — Yes (1) | No (2)

BC Were people's fingers marked with indelible ink before voting? — Yes (1) | No (2)

BD Were ballot papers stamped with the validating stamp before being issued? — Yes (1) | No (2)

BE Were any unauthorised person permitted to remain in the polling station during voting? — Yes (1) | No (2)

BF Did anyone attempt to harass or intimidate voters or polling officials during voting? — Yes (1) | No (2)

(For Questions BG, BH, BJ, BK and BM: None = 0, Few = 1 to 5, Some = 6 to 15 and Many = 16 or more)

BG How many people with Voter ID Cards were not permitted to vote? — None (0) | Few (1) | Some (2) | Many (3)

BH How many people were permitted to vote without voter ID cards? — None (0) | Few (1) | Some (2) | Many (3)

BJ How many people were permitted to vote whose names did not appear on the voters' register? — None (0) | Few (1) | Some (2) | Many (3)

BK How many people were permitted to vote whose biometric details did not match? — None (0) | Few (1) | Some (2) | Many (3)

BM How many people were assisted to vote (blind, disabled, elderly, etc)? — None (0) | Few (1) | Some (2) | Many (3)

BN Were assisted voters allowed to have a person of their own choice to assist them to vote? — Yes (0) | No (1) | No Assisted Voters (3)

BP Overall, how would you describe any problems that may have occurred during the voting process? — Major (1) | Minor (2) | None (3)

CLOSING & COUNTING

CA How many people were in queue at 5:00 pm? — None (0) | Few (1) | Some (2) | Many (3)

CB	Was everyone who was in the queue at 5:00 pm permitted to vote?		Yes (1)	No (2)
CC	Was anyone who arrived at the polling station after 5:00 pm permitted to vote?	Yes (1) / No (2)	No One Arrived After 5pm (3)	
CD	Did anyone attempt to harass or intimidate polling officials during counting?		Yes (1)	No (2)
CE	Were more ballot papers found in the presidential ballot box than voters who cast ballots?		Yes (1)	No (2)
CF	Did any polling agent request a recount of the presidential ballots?		Yes (1)	No (2)
CG	Did an NDC polling agent sign the declaration of results for the presidential election?	Yes (1) / No (2)	No NDC Agent Present (3)	
CH	Did an NPP polling agent sign the declaration of results for the presidential election?	Yes (1) / No (2)	No NPP Agent Present (3)	
CJ	Did any other polling agent present sign the declaration of results for the presidential election?	Yes (1) / No (2)	No Other Agents Present (3)	
CK	Do you agree with the vote count for the presidential election?		Yes (1)	No (2)
CM	Did all polling agents present sign the declaration of results for the parliamentary election?		Yes (1)	No (2)
CN	Do you agree with the vote count for the parliamentary election?		Yes (1)	No (2)

PRESIDENTIAL VOTE COUNT

DA	Spoilt ballot papers	**DH**	UFP (Akwasi Addai Odike)
DB	Rejected ballot papers	**DJ**	PNC (Ayariga Hassan)
DC	Total Valid Votes	**DK**	CPP (Michael Abu Sakara Forster)
DD	NDC (John Dramani Mahama)	**DM**	INDP (Jacob Osei Yeboah)
DE	GCPP (Henry Hebert Lartey)		
DF	NPP (Nana Addo Dankwa Akufo-Addo)		
DG	PPP (Papa Kwesi Ndoum)		

Who was interview

☐ 1) Party Agent (specify)……..
☐ 2) EC Official
☐ 3) Security Personnel

Enumerator First Name _____ Surname _____ Arrival Time _____ Departure Time _____ Signature _____

B Details of the Experimental Design

We implement a "randomized saturation" experimental design (**?**). The advantage of the randomized saturation design is that it allows us to estimate the causal effect of election observers while including in the estimates their potential "spillover" effects. Spillover effects occur when the treatment status of one unit impacts outcomes at other units (**?**): in our case, when the deployment of an observer to one polling station influences election integrity at other polling stations (because the observer "pushes" election fraud to unobserved polling stations).

The design involves a two-stage randomization process: in our case, first at the constituency level and then at the polling station level. In the first stage, we assign constituencies to an observer "saturation" treatment. Saturation is defined as the proportion of polling stations within a constituency that is monitored by observers. In the second stage, we randomly assign observers to polling stations within the sample of constituencies.

In the first stage, we randomly assign each constituency to one of three saturations: *low*, *medium*, and *high*. In the low condition, observers are deployed to 30 percent of sample polling stations in the constituency. In the medium condition, we treat 50 percent of sample polling stations. In the high condition, we treat 80 percent of sample polling stations.[25] In the second stage of our randomization process, we randomly assign individual polling stations to treated (observed) or control (unobserved) status. The proportion of polling stations randomly assigned to treatment within a constituency is determined by the randomly assigned saturation level in the first stage. The approach yields a 3×2 experimental design. In total, we send observers to $1,292$ polling stations across 60 constituencies in the

[25] In other research (**?**), we seek to identify the spillover effects of observers on fraud in polling stations that are not under observation. The estimation of spillover effects relies on comparisons of control units in each of the three constituency level conditions. Since by definition there are relatively few control stations in the higher saturation constituencies, we assign the constituency treatments with a probability of 20 percent for the low condition and 40 percent for the medium and high conditions. This increases the statistical power to detect spillover effects. Such spillovers are not the focus of the present study.

sample.

In our experimental framework, potential outcomes are determined by the polling station's treatment status and the treatment condition of each station's constituency. Potential outcomes can be written as follows:

$$Y_{ij}(T_{ij}, S_j) \tag{B.1}$$

where Y_{ij} is one of the indices of election integrity (such as ballot stuffing or overvoting) at polling station i in constituency j. T_{ij} indicates treatment status at polling station i in constituency j ($T_{ij} = 1$ if an observer is present, and 0 otherwise). The constituency level treatment status is indicated by S_j, where $S_j = s$ and $s \in \{low, medium, high\}$.

To account for spillover in our estimation of causal effects, we compare outcomes in treated polling stations to outcomes in control polling stations in the low saturation constituencies. Since the saturation of treatment in the low condition constituencies is relatively low, the control polling stations in the low condition constituencies are less likely to be affected by spillover effects. Comparing outcomes in treated polling stations only to these low condition control stations should therefore generate less biased estimates of observers' causal effects.[26] To estimate the average treatment effect of election observers, we therefore define a dummy variable, W_{ij}, which takes a value of 1 if the unit is a control polling station located in one of the medium and high saturation constituencies (following **?**). To estimate the average treatment effect, we estimate the following regression model:

[26]Ideally, we would have implemented the study with "pure" control polling stations. Pure control units are untreated units that are not susceptible to spillover effects because there are no treated units in the same constituency (or local area). In our study, no control units were assigned to this pure control status. This decision was driven solely by practical considerations. Given CODEO's mission, which is to deter electoral malfeasance and enhance the quality of elections across the country, we were unable to create constituencies in which no observers were present. It is an important part of CODEO's mission to be present in all regions and constituencies of the country, in part so that the organization maintains credibility as an impartial observer. Therefore, we use control stations in low saturation constituency as our main comparison set.

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 W_{ij} + \epsilon_{ij} \tag{B.2}$$

Here, β_1 provides the estimate of the average treatment effect. It compares outcomes in all treated polling stations to outcomes in control stations in the low saturation constituencies. We cluster standard errors by constituency to account for the fact that the cluster-level treatments are assigned at that level.

In Table 2, we reported average treatment effects. In Table B.1 we report results of the same regressions including spillover effects.

Table B.1: Average Treatment Effects of Election Observers on Machine Breakdown and Fraud, Incorporating Spillover

	Breakdown		Overvoting rate		Registry discrepancy		Ballot stuffing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Election observer	−0.113*	−0.109*	−0.026	−0.028	−0.087*	−0.084*	−0.005	−0.005
	(0.048)	(0.052)	(0.015)	(0.015)	(0.040)	(0.039)	(0.026)	(0.028)
Competitive		0.078*		0.013		0.039		0.023
		(0.037)		(0.009)		(0.023)		(0.019)
Urban		0.034		−0.008		0.029		0.003
		(0.034)		(0.009)		(0.023)		(0.017)
Spillover	0.095	0.099	0.014	0.012	0.048	0.051	0.035	0.035
	(0.062)	(0.065)	(0.019)	(0.018)	(0.050)	(0.049)	(0.042)	(0.041)
Constant	0.285***	0.230***	0.041**	0.041**	0.221***	0.186***	0.036	0.025
	(0.044)	(0.054)	(0.015)	(0.016)	(0.039)	(0.045)	(0.025)	(0.027)
Observations	1,888	1,888	1,864	1,864	1,973	1,973	1,987	1,987
R^2	0.045	0.054	0.008	0.010	0.024	0.028	0.007	0.011

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with robust standard errors clustered by constituency in parentheses. Eleven polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials on relevant variables are dropped.

C Covariate Balance Tests

In this section, we present data showing balance on various dimensions for treated and control polling stations. We use data from a household survey we conducted in the communities near observed and unobserved polling places during the two days following the elections.[27] As part of the survey, we gathered data on voting behavior in the prior 2008 election as well as measures of socio-economic conditions and ethnic self-identification.

Table C.1 presents means in control and observed communities on a number of pre-election covariates. It also presents the difference in these means and the p-value of a two-tailed difference-of-means test. The first section of the table shows that the partisan voting histories of residents near observed and unobserved polling are comparable. In both sets of communities, about 35 percent report voting for the NPP in the 2008 presidential election, while about 43 percent report voting for the NDC, whose candidate was the winner of that election. The remaining sections of the table examine measures of education, poverty and well-being, and ethnicity. Observed and control polling stations are also similar along these dimensions. The data presented in the table shows that the communities surrounding our observed and control polling stations are comparable across a range of political, ethnic and and socio-economic characteristics that could potentially affect the level of election fraud.

[27] We surveyed over 6,000 Ghanaians. Ideally, we would have randomly sampled individuals from the official voter register. Because this was not available, we employed the random sampling techniques used across Africa by the Afrobarometer public opinion survey. Our enumerators visited each of our approximately 2,000 sampled polling place and then selected four households using a random walk technique.

Table C.1: Polling Station (Unit) Level Covariate Balance

	Mean Treatment	Mean Control	Difference	P-Value
NPP Presidential Vote 2008	0.359	0.368	-0.009	0.562
NDC Presidential Vote 2008	0.425	0.426	-0.000	0.975
NPP Parliamentary Vote 2008	0.359	0.391	-0.032	0.034
NDC Parliamentary Vote 2008	0.408	0.401	0.008	0.614
Poverty index	0.956	0.981	-0.025	0.151
Electricity	1.117	1.156	-0.039	0.104
Medicine	0.896	0.886	0.010	0.659
Sufficient Food	0.840	0.879	-0.039	0.106
Cash Income	0.970	1.002	-0.031	0.143
No Formal Schooling	0.145	0.145	0.000	0.987
Completed Primary Schooling	0.716	0.698	0.018	0.206
Post Primary Schooling	0.543	0.522	0.021	0.184
Formal House	0.181	0.177	0.004	0.737
Concrete Permanent House	0.427	0.416	0.011	0.463
Concrete and Mud House	0.218	0.219	-0.001	0.952
Mud House	0.168	0.181	-0.014	0.242
Akan	0.685	0.699	-0.013	0.350
Ga	0.021	0.018	0.002	0.614
Ewe	0.220	0.203	0.016	0.201
Other, Refuse, or Don't Know	0.074	0.079	-0.005	0.546

Notes: P-values calculated from two-tailed difference-of-means tests. Poverty index constructed by adding responses to the following items and dividing by the total number of items: How often did you go without the following in the past year, where the items are cash income; sufficient food; medicine; and electricity. Responses were: *Never* (0), *Occasionally* (1), and *Most of the time* (2).

D Data Collection and Robustness

In this section, we describe features of the data collection and analysis that led us to undertake specific robustness checks. We present the results of the robustness checks and verify that the results reported in the body of the paper remain unchanged.

Data was collected on election day from sampled polling stations by enumerators using the questionnaire that appears in Appendix A. In treated areas, enumerators acted as election observers for the whole of the day. These observer/enumerators remained in a single randomly selected polling station through the vote count. They recorded the information reported and analyzed here by observing events at the polling station as well as the vote count. Due to logistical challenges and because we wanted to avoid treating control stations by sending enumerators to observe activities throughout the course of election day, other enumerators were assigned to visit three to four control polling stations after the polls had closed at 17:00 and, using identical questionnaires, to collect the same information. They were instructed to collect information from two persons, ideally the two official representatives of the major political parties in each polling place. These representatives typically gather similar information to report to central party offices. If enumerators could not speak with a party representative because the person was unavailable, enumerators were instructed to collect the information from the presiding officer. Members of both groups of enumerators received identical training, were officially designated CODEO election observers, and received appropriate identification materials that permitted them entry into polling stations.

The analysis reported in this paper relies on four pieces of information collected by enumerators: (1) the number of rejected ballots; (2) the number of valid votes; (3) whether more ballot papers were found in the presidential ballot box than had been cast; and (4) whether at any point during the data the biometric verification machine malfunctioned. In addition, we use data supplied by the Election Commission on the number of registered voters at each polling station.

The analysis reported in the main body of the paper uses all observations where enumerators only were able to collect data from one party official as well as all observations where the enumerator was able to reach both party officials and the answers were identical along the variables of interest. This means the main analysis drops only observations where the party officials explicitly disagreed. In the robustness tests presented below, we drop observations that fail to meet one of two criteria: the enumerator collected the information through direct observation (i.e. the polling station was a treated unit) or the enumerator collected identical information from two separate respondents. Thus, observations with only one party official's responses are no longer included.

Table D.1 and Table D.2 replicate the main analyses from the results Section using this smaller dataset. All the results remain stable.

Table D.1: Descriptive Statistics of Machine Breakdown and Measures of Fraud, Smaller Dataset

	(1) Full Sample	(2) Treatment	(3) Control	(4) Competitive	(5) Stronghold	(6) Urban	(7) Rural
Machine breakdown	0.223	0.172	0.335	0.264	0.192	0.238	0.206
	(0.416)	(0.378)	(0.472)	(0.441)	(0.394)	(0.426)	(0.404)
Overvoting rate	0.025	0.015	0.045	0.029	0.021	0.021	0.028
	(0.185)	(0.152)	(0.242)	(0.188)	(0.184)	(0.171)	(0.200)
Registry discrepancy	0.172	0.134	0.248	0.190	0.159	0.185	0.157
	(0.377)	(0.340)	(0.432)	(0.392)	(0.366)	(0.389)	(0.364)
Ballot stuffing	0.041	0.032	0.060	0.053	0.033	0.043	0.039
	(0.199)	(0.175)	(0.238)	(0.224)	(0.178)	(0.203)	(0.194)
Observations	1988	1281	707	831	1157	1051	937

Notes: Standard deviations in parentheses. Ten polling stations without biometric verification machines removed from sample. Polling stations where enumerators recorded different responses from two party officials on relevant variables or responses from only one party official are dropped.

Table D.2: Average Treatment Effects of Election Observers on Machine Breakdown and Fraud, Smaller Dataset

	Breakdown		Overvoting rate		Registry discrepancy		Ballot stuffing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Election observer	−0.163***	−0.163***	−0.030**	−0.031**	−0.115***	−0.114***	−0.029	−0.029
	(0.040)	(0.041)	(0.011)	(0.011)	(0.025)	(0.025)	(0.029)	(0.030)
Competitive		0.076		0.008		0.035		0.022
		(0.039)		(0.009)		(0.023)		(0.020)
Urban		0.028		−0.008		0.024		0.004
		(0.035)		(0.009)		(0.022)		(0.019)
Constant	0.335***	0.288***	0.045***	0.047***	0.248***	0.221***	0.060*	0.050
	(0.037)	(0.041)	(0.010)	(0.012)	(0.025)	(0.029)	(0.028)	(0.029)
Observations	1,814	1,814	1,786	1,786	1,896	1,896	1,895	1,895
R^2	0.033	0.042	0.006	0.007	0.021	0.024	0.005	0.008

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with robust standard errors clustered by constituency in parentheses. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials on relevant variables or responses from only one party official are dropped.

Table D.3: Effect of Election Observers and Competitiveness on Machine Breakdown and Fraud, Smaller Dataset

	Breakdown	Overvoting rate	Registry discrepancy	Ballot stuffing
	(1)	(2)	(3)	(4)
Election Observer	-0.157^{***}	-0.030^{**}	-0.111^{***}	-0.028
	(0.041)	(0.011)	(0.025)	(0.029)
Urban	0.023	-0.009	0.022	0.002
	(0.035)	(0.009)	(0.022)	(0.020)
Margin	-0.159^{**}	-0.005	-0.082^{*}	-0.035
	(0.056)	(0.017)	(0.039)	(0.024)
Constant	0.368^{***}	0.052^{***}	0.261^{***}	0.069
	(0.054)	(0.014)	(0.028)	(0.042)
Observations	1,814	1,786	1,896	1,895
R^2	0.045	0.006	0.025	0.007

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with robust standard errors clustered by constituency in parentheses. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials on relevant variables or responses from only one party official are dropped.

Table D.4: Interactive Effect of Machine Breakdown and Election Observers on Fraud, Smaller Dataset

	Overvoting rate	Registry discrepancy	Ballot stuffing
	OLS	*logistic*	*logistic*
	(1)	(2)	(3)
Machine breakdown	0.001	0.388	1.290***
	(0.020)	(0.235)	(0.333)
Election observer	−0.025*	−0.674**	−0.342
	(0.010)	(0.206)	(0.564)
Competitive	0.007	0.302	0.500
	(0.010)	(0.164)	(0.392)
Urban	−0.004	0.158	−0.001
	(0.009)	(0.161)	(0.418)
Breakdown X observer	0.018	−0.341	−0.680
	(0.027)	(0.345)	(0.458)
Constant	0.036***	−1.472***	−3.470***
	(0.011)	(0.228)	(0.508)
Observations	1,676	1,759	1,766

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS in Column 1 and logistic regression in Columns 2 and 3, with robust standard errors clustered by constituency in parentheses. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials on relevant variables or responses from only one party official are dropped.

E Alternative Specifications of Average Treatment Effects

Table E.1: Average Treatment Effects of Election Observers on Machine Breakdown and Fraud, Logistic Models

	Breakdown		Registry discrepancy		Ballot stuffing	
	(1)	(2)	(3)	(4)	(5)	(6)
Election observer	−0.971***	−0.974***	−0.798***	−0.796***	−0.711	−0.720
	(0.188)	(0.194)	(0.146)	(0.149)	(0.496)	(0.512)
Competitive		0.452*		0.273		0.561
		(0.212)		(0.163)		(0.402)
Urban		0.180		0.189		0.037
		(0.211)		(0.160)		(0.437)
Constant	−0.600***	−0.903***	−1.071***	−1.295***	−2.711***	−2.997***
	(0.153)	(0.199)	(0.130)	(0.175)	(0.440)	(0.494)
Observations	1,888	1,888	1,973	1,973	1,987	1,987
Akaike Inf. Crit.	1,981.043	1,967.207	1,803.952	1,800.830	695.955	693.672

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Logistic regression with robust standard errors clustered by constituency in parentheses. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials are dropped.

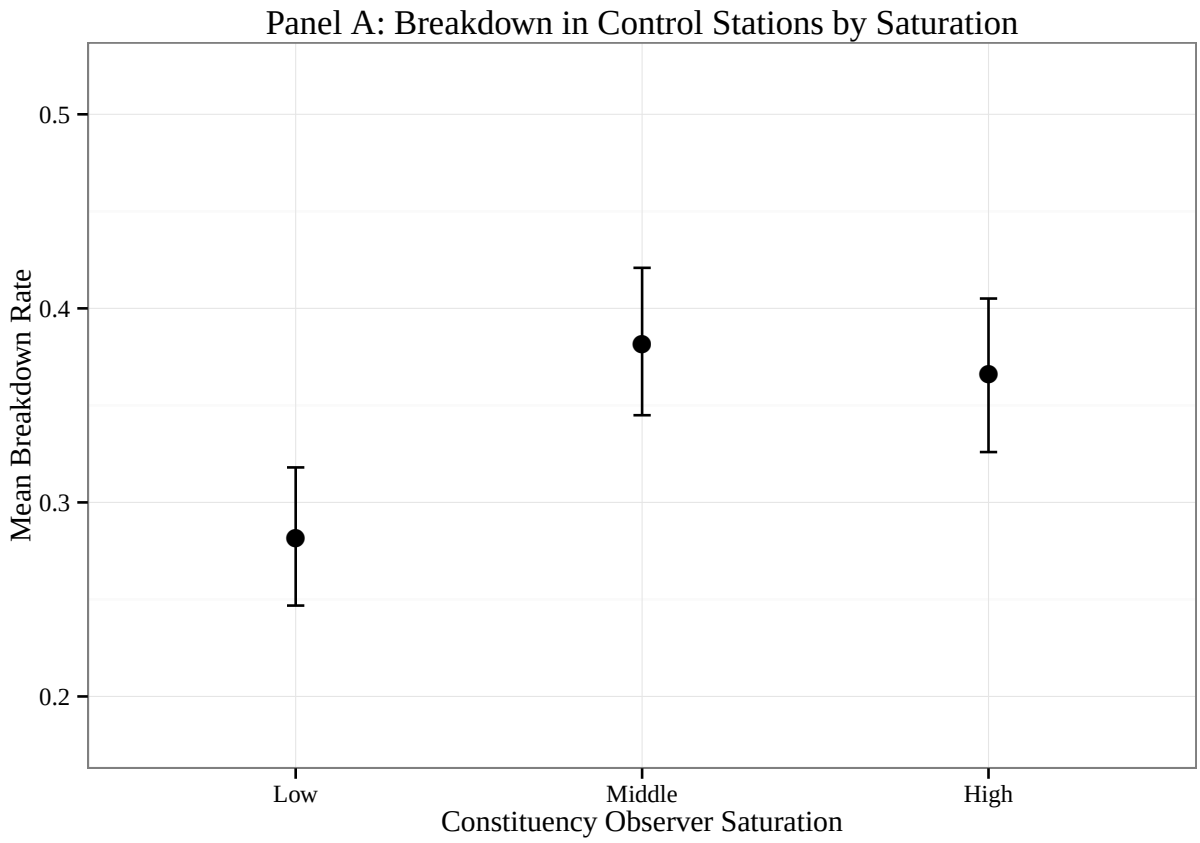
Table E.2: Average Treatment Effects of Election Observers on Machine Breakdown and Fraud, Non-Robust Standard Errors

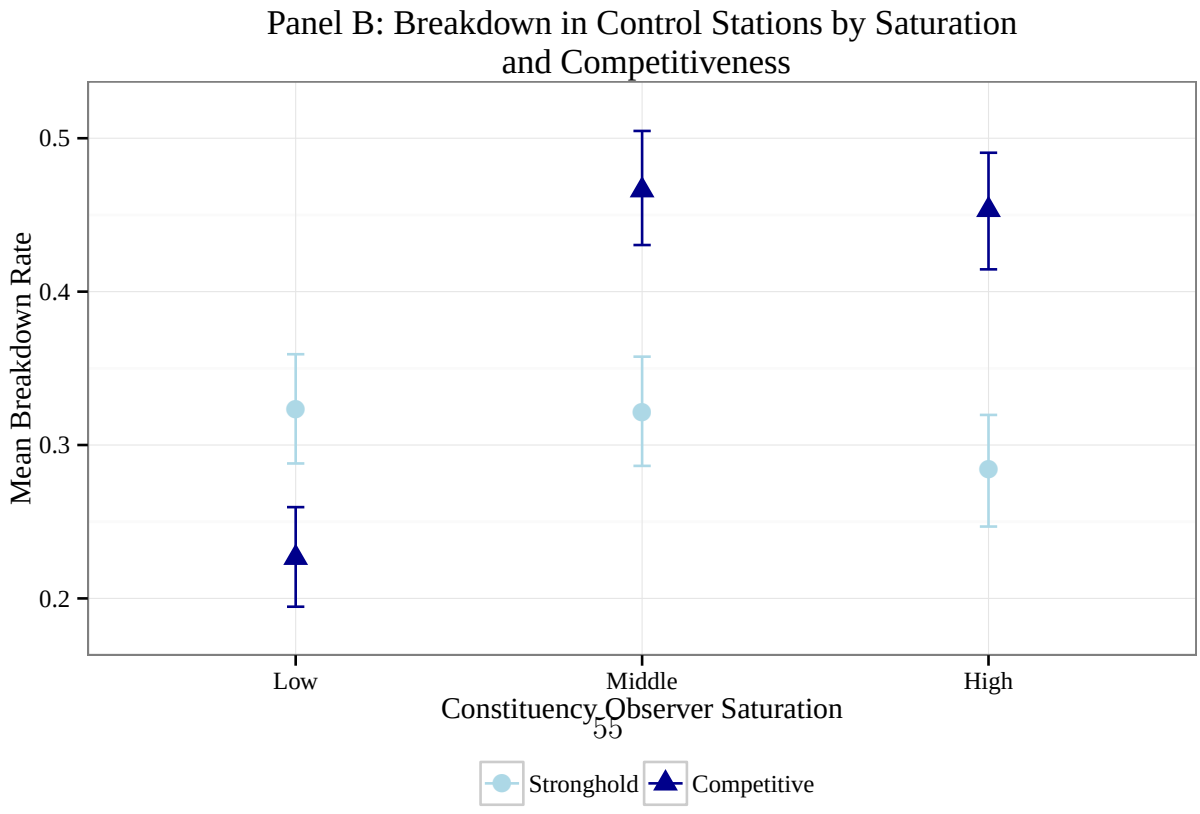
	Breakdown		Overvoting rate		Registry discrepancy		Ballot stuffing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Election observer	−0.182***	−0.181***	−0.036***	−0.037***	−0.122***	−0.121***	−0.031**	−0.031**
	(0.020)	(0.020)	(0.009)	(0.009)	(0.018)	(0.018)	(0.009)	(0.009)
Competitive		0.078***		0.013		0.039*		0.023*
		(0.019)		(0.009)		(0.017)		(0.009)
Urban		0.030		−0.008		0.026		0.001
		(0.019)		(0.009)		(0.017)		(0.009)
Constant	0.354***	0.304***	0.051***	0.050***	0.255***	0.224***	0.062***	0.052***
	(0.016)	(0.022)	(0.008)	(0.010)	(0.014)	(0.019)	(0.008)	(0.010)
Observations	1,888	1,888	1,864	1,864	1,973	1,973	1,987	1,987
R^2	0.042	0.051	0.008	0.009	0.023	0.027	0.005	0.009

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with non-robust standard errors. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials are dropped.

F Estimates of Spillover of Machine Breakdown onto Control Polling Stations

Figure F.1: Spillover of Machine Breakdown onto Control Polling Stations





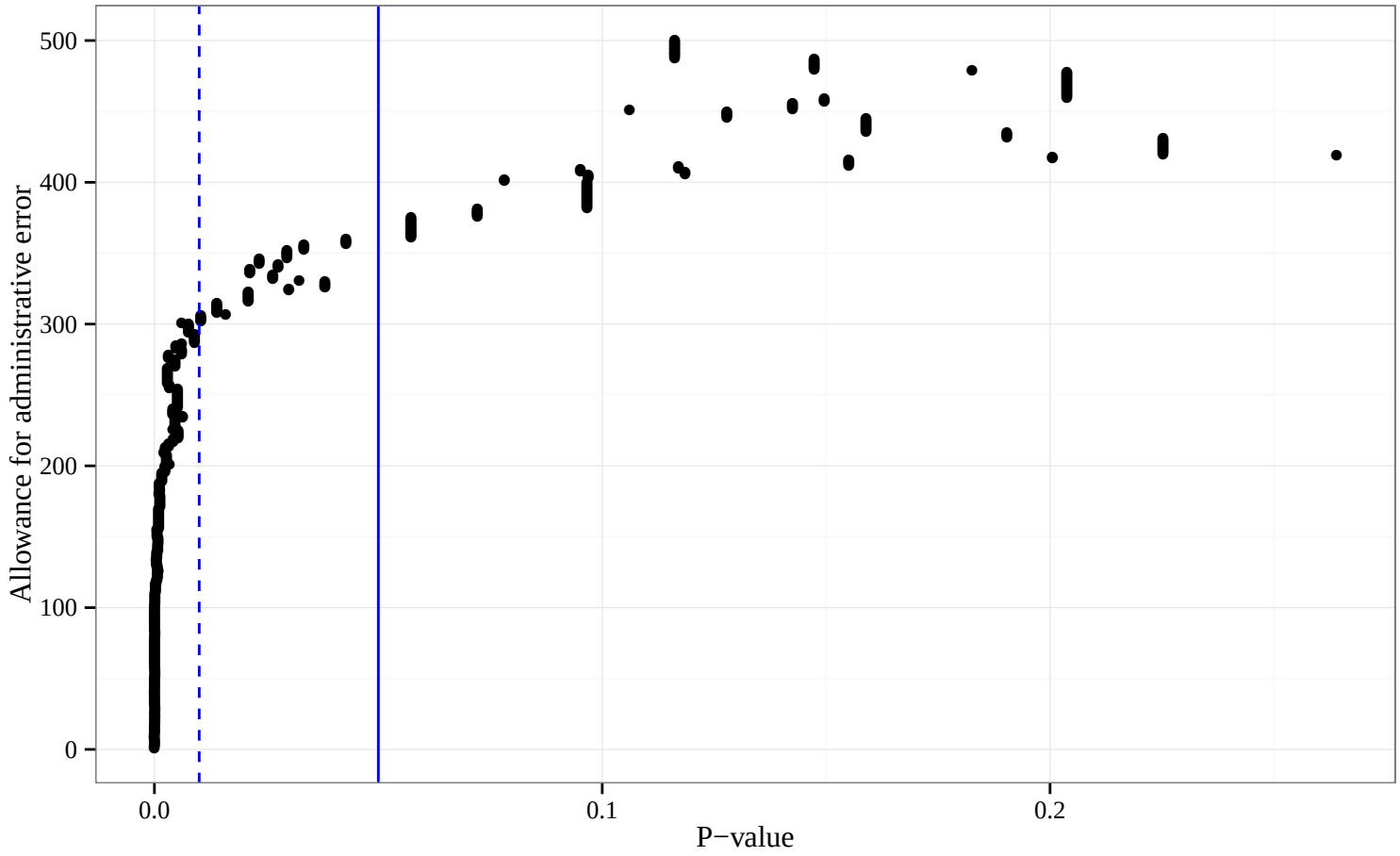
G Cutoffs for Registry Rigging and Registry Difference

Figure G.1 demonstrates that *registry rigging* is robust to different cutoffs that allow for administrative error. Table G.1 replicates the regressions in Table 2 and Table **??** using a continuous rather than discrete measure of registry rigging. The measure *registry difference* is continuous from −923 to 1314, with a large number of values at 0 where the registry number matched the official EC number. The measure is constructed by subtracting the Electoral Commission figures from the number of registered voters according to the local rolls. Positive numbers indicate the number on the local rolls was larger than the Electoral Commission figure. In two polling stations the local rolls showed figures greater than 8000. These appear to be simple transcription errors. In seven polling stations the local roll numbers were not available. These polling stations are all dropped.[28]

The results in Table G.1 show that control stations had significant positive values of *registry difference* while observers reduced that value to almost 0. This would support the story that registry rigging was deliberately done to inflate the voter rolls. Furthermore, although the standard errors are quite large, the second column is consistent with the story that registry rigging to inflate the rolls was encouraged by machine breakdown, although that relationship disappears in the presence of an observer.

[28]Their inclusion only strengthens our findings, but it is inappropriate to include them as they are clear outliers driven by erroneous data collection that drive the regressions below

Figure G.1: Allowance for Administrative Discrepancies When Constructing *Registry Rigging*



Notes: This plot demonstrates the p-values for unequal variance t-tests for a range of cutoffs for the rigging variable. The y-axis describes the allowance for administrative error and the x-axis is the p-value of the ATE effect of observers on registry rigging. The dashed vertical line is at 0.01 and the solid horizontal line is at 0.05. As we move along the y-axis, we only code a polling station with registry rigging if the difference between the number of registered voters according to the Electoral Commission and the local figure at the polling station is greater than or equal to the value of the y-axis — the cutoff. The treatment effect of observers remains significantly negative until the cutoff is 357, meaning until we treat only discrepancies *greater* than or equal to 357 as registry rigging, observers have a significant depressive effect on registry rigging.

Table G.1: Effect of Observers and Machine Breakdown on Registry Difference

	Registry Difference	
	(1)	(2)
Machine breakdown		25.410
		(16.768)
Election observer	−19.651**	−14.302*
	(6.784)	(5.712)
Competitive	2.725	2.621
	(5.690)	(5.671)
Urban	−3.786	−3.126
	(5.471)	(5.636)
Breakdown X observer		−24.981
		(15.768)
Constant	22.329**	16.166**
	(7.159)	(6.052)
Observations	1,973	1,830
R^2	0.008	0.014

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. OLS with non-robust standard errors. Ten polling stations without biometric verification machines removed from sample of machine breakdown. Polling stations where enumerators recorded different responses from two party officials are dropped.

H Holm Correction

For a brief introduction and overview of the Holm correction, see **?**. It is very easily implemented in `R` using the `p.adjust` function. In the table below, we report the average treatment effect of election observers on a host of outcome variables. We include the effect size, the traditional p-value, the Holm corrected p-value, and the Bonferroni corrected p-value. Notice that the key treatment effects that were originally significant remain significant even using the stricter Bonferroni correction.

Table H.1: Effect Sizes and Corrected P-Values of 43 Outcomes

	Control mean	Treatment effect	p-value	Holm p-value	Bonferroni p-value
Time voting commenced (Early[1] - Late[5])	1.23	0.27	0.00	0.00	0.00
Station set up to ensure ballot secrecy (0/1)	0.96	-0.12	0.00	0.00	0.00
Machine breakdown (0/1)	0.35	-0.18	0.00	0.00	0.00
Election official present (0/1)	0.92	-0.13	0.00	0.00	0.00
No. of voters assisted (None[0] - Many[4])	1.05	0.29	0.00	0.00	0.00
Voters w/out bio matches allowed to vote (None[0] - Many[3])	0.14	-0.10	0.00	0.00	0.00
Registry discrepancy (0/1)	0.26	-0.12	0.00	0.00	0.00
Voter not on register allowed to vote (None[0] - Many[3])	0.13	-0.10	0.00	0.00	0.00
Harrassment or intimidation of voters (0/1)	0.12	-0.07	0.00	0.00	0.00
Election materials missing (0/1)	0.03	0.04	0.00	0.00	0.00
Overvoting dummy (0/1)	0.07	-0.05	0.00	0.00	0.00
Security personnel present (0/1)	0.92	-0.06	0.00	0.00	0.00
Was station accessible to elderly/disabled (0/1)	0.98	-0.03	0.00	0.00	0.00
Agents agree on parl. results (0/1)	0.98	0.02	0.00	0.01	0.01
Voters w/ no ID permitted to vote (None[0] - Many[3])	0.56	0.13	0.00	0.01	0.02
Allowed voters arriving after 5 (Yes[1], No[2], NA[3])	2.46	0.10	0.00	0.01	0.02
Registry difference	21.31	-19.40	0.00	0.02	0.03
Agents agree on pres. results (0/1)	0.98	0.02	0.00	0.02	0.04
Voters biometric registration verified (0/1)	0.97	0.02	0.00	0.03	0.06
Overvoting rate	0.06	-0.05	0.00	0.04	0.07
Turnout	87.90	-5.22	0.00	0.05	0.09
Ballot stuffing (0/1)	0.06	-0.03	0.00	0.06	0.12
Unauthorized persons at station (0/1)	0.12	-0.04	0.00	0.06	0.13
Overall problems (Major[1], Minor[2], None[3])	2.71	-0.08	0.00	0.07	0.16
Voters marked with ink (0/1)	0.96	0.02	0.00	0.08	0.19
Count of intimidation events	0.06	-0.03	0.01	0.19	0.47
Other party agents sign pres. results (0/1)	1.43	0.09	0.01	0.24	0.63
NDC agent sign pres. results (Yes[1], No[2], Not Present[3])	1.04	-0.02	0.02	0.28	0.77
Voters queued at 5pm allowed to vote (0/1)	0.80	-0.04	0.05	0.68	1.00
NDC agent present (0/1)	0.99	0.01	0.06	0.84	1.00
NPP votes	217.49	-14.70	0.07	0.89	1.00
Party agent allowed to observe (0/1)	1.00	-0.00	0.25	1.00	1.00
Ballot papers stamped by EC (0/1)	0.98	0.01	0.11	1.00	1.00
No. of voters on proxy list	6.02	3.80	0.20	1.00	1.00
Empty ballot boxes displayed pre-voting (0/1)	1.00	-0.00	0.38	1.00	1.00
Voters queued at 5pm (0/1)	0.78	-0.08	0.14	1.00	1.00
Voters w/ ID not permitted to vote (None[0] - Many[3])	0.30	-0.03	0.29	1.00	1.00
NDC votes	222.92	-5.27	0.46	1.00	1.00
Party agents sign parl. results (0/1)	0.97	0.01	0.36	1.00	1.00
NPP agent sign pres. results (Yes[1], No[2], Not Present[3])	1.02	-0.00	0.60	1.00	1.00
NPP agent present (0/1)	0.99	-0.00	0.76	1.00	1.00
Request for recount (0/1)	0.09	-0.00	0.80	1.00	1.00