MARTA REGALIA

Department of Social and Political Sciences, Università degli Studi di Milano, Italy

The Effect of Election Observation

Evidence from Ukraine 2004 Presidential Elections

ABSTRACT International election observation has become a standard practice in democratizing countries. Doubts have been cast on the ability of electoral observers to accurately judge the freedom and fairness of an electoral process, and the scholarly literature has still not reached a consensus on the unintended consequences of election observation. This article empirically tests the hypothesis that observers can deter election-day fraud through a natural experiment on polling-station-level election results. Using data from the Ukraine 2004 presidential election, it will show that OSCE/ODIHR observation has both immediate and lasting effect on domestic political actors' behavior. Results do support the usefulness of election observation in reducing election-day fraud.

KEYWORDS election observation, election fraud, Ukraine, presidential elections

INTRODUCTION

The study of election observation (EO) is one of the topics of the very large field of study on democratization and, more specifically, democracy promotion and assistance. As elsewhere, in the post-communist world international intervention has been found as one of the driving factors of (attempted or successful) political transitions. For example, in Ukraine, the case examined in this article, the international dimension (or the lack of international involvement) has been recognized by many scholars as a relevant variable in explaining both internal developments (Kubicek, 2005; Žielys & Rudinskaitė, 2014) and voting behaviors (Makhorkina, 2005). Given the importance of elections for democracy, and the rising number of electoral autocracies in post-communist countries, research on election observation is increasingly pivotal. Actually, many post-communist countries have a significant record of rigged elections (among the many, Armenia, Kazakhstan, Kyrgyzstan, Tajikistan, Ukraine, Uzbekistan, etc.) having inherited from Soviet times a sturdy tradition of election manipulation that transforms elections in an instrument of autocratic legitimation instead than an expression of citizens' political preferences. As argued by Andrew Wilson (2005), many post-Soviet countries have developed advanced techniques of manipulating the democratic process. More specifically, many post-communist regimes are used to employing sophisticated strategies to manipulate election processes and their outcomes. Since the aim of Election Observation Missions (EOMs) is to observe and report on the quality of an electoral process, some authors (Bader, 2011) have hypothesized that EOMs may have the unintended consequence of encouraging

innovations in cheating strategies. Among them, Foroughia and Mukhtorova (2017) talk about a "Helsinki's counterintuitive effect," explaining that OSCE/ODIHR election observation in post-communist states with histories of fraudulent elections has become counterproductive, not contributing to deepening democratization and political pluralism but, rather, inadvertently aiding in the consolidation of authoritarianism by helping the propaganda of autocrats sustain that they are interested in and are making progress in democratization. Other authors, on the contrary, assert that EOMs are able to both detect and deter fraud (Hyde, 2007).

While representing only one of the many forms of international actors' involvement in domestic politics, the consequences of EO can be fundamental for the political future of a country. The literature on electoral integrity has emphasized the importance of meeting international standards of free and fair elections in order, for example, to avoid the outbreaks of street protests following a rigged election (Norris, 2014) or, more specifically, to foster satisfaction with the democratic process (Fortin-Rittberger, Harfst & Dingler, 2017). Since EO is meant to assess the quality of elections, understanding its ability to detect and deter fraud is the first step in evaluating its role in democracy promotion. If its potential is confirmed, EO can help improving the quality of election and thus contribute to democratization. Actually, a study of electoral integrity in three post-communist countries (including Ukraine) showed that citizens' perceptions of electoral integrity have a significant impact on satisfaction with democracy (McAllister & White, 2015) and can therefore influence future political trajectories in democratic transition.

In this study, I focus on election observation, a form of international actors' involvement in domestic democratic processes. This article examines the role of OSCE/ODIHR observers as an external agent for democratization in Ukraine. While today up to 80% of elections in developing democracies are monitored by foreign observers (Hyde, 2011b; Kelley, 2012b), the literature has found no agreement on the ability of election observers (EOs) to detect and deter fraud. Empirical research examining EO is still emerging and there is no clear consensus on its effects (Carothers, 1997; Bjornlund, 2004; Alvarez, Hall & Hyde, 2008; Kelley, 2009; Donno, 2010; Darnolf, 2011; Regalia, 2011; Daxecker, 2012; Hyde & Marinov, 2012; Simpser & Donno, 2012; Smidt, 2016).

The scholarly literature on EO documents the role of EOs in enhancing the legitimacy of the election they monitor (Hyde, 2011a; Hyde & Marinov, 2012; Kelley, 2008). This explains why EO has become a standard practice in emerging democracies (Hyde, 2011b), even though doubts have been cast on the ability of EOs to accurately judge the freedom and fairness of an electoral process (Kelley, 2010 2012b). Beyond assessing the quality of an electoral process, scholars often claim observers can influence the quality of the election they monitor by (a) increasing participation of opposition parties (Bjornlund, Bratton & Gibson, 1992; McCoy, Garber & Pastor, 1991) by boosting confidence in the electoral process (Garber & Cowan, 1993; Kelley, 2012a); (b) increasing voter turnout (McCoy, Garber & Pastor, 1991); (c) decreasing the probability of opposition party boycott (Beaulieu & Hyde, 2008) and the level of post-election violence (Daxecker, 2012; Smidt, 2016; von Borzyskowski, 2019); and (d) raising the quality of elections by

reducing fraud and violence (Carothers, 1997; Kelley, 2012b; Donno, 2013; Hafner-Burton, Hyde & Jablonski, 2014).

The ability of EOs to reduce fraud is still debated. Among others, Hyde (2007) analyzes the 2003 presidential elections in Armenia through a natural experiment and finds that EO significantly reduces election-day fraud. She finds similar results in a randomized experiment in Indonesia (Hyde, 2010). Enikolopov et al. (2013) set an experiment in Moscow and find a large observers' effect on turnout and the ruling party's vote share. On the contrary, the analysis of Ichino and Schündeln (2012) shows that fraud in voter registration in Ghana was displaced from districts where observers were present to districts not observed. Similarly, Buzin et al. (2016), analyzing the 2011 Russian parliamentary elections, find little effect either on turnout or on the ruling party vote share.

This article, building on previous research, addresses the unintended consequences of EO on election-day fraud and moves beyond existing works by adding new evidence and extending the set of statistical tests. I focus on election-day fraud, that is, on attempts to unduly influence the outcome of the election taking place in polling stations (PSs) and around them through voter intimidation, ballot box stuffing, factually denying the voting right to particular voters or groups, manipulating the count of votes, and so on (Schedler, 2002). Using data from the 2004 presidential election in Ukraine, I show that EOs have both immediate and lasting effects on the level of election-day fraud.

RESEARCH DESIGN

There are various types of election-day fraud, but all of them share one attribute: the goal of increasing the vote share of the candidate(s) or the party(ies) that is/are cheating (Hyde, 2007). It is exactly this kind of fraud that EOs are best at deterring and discovering: in fact, the behavior of internal political actors may be influenced by the physical presence of EOs inside and around PSs.

The simplest model sees a dependent variable, the percentage vote share for the fraud-sponsoring candidate(s) or party(ies) and an independent dichotomous variable, the treatment, or whether the PS was observed. By comparing election results of PSs visited by observers with results of PSs not visited, we can evaluate the "observers' effect":

> HP: *If international election observation reduces election-day fraud, then the fraud-sponsoring candidate(s) or party(ies) should gain a lower average vote share in polling stations where observers were present than in polling stations where they were not present.*

We can solve the problems of causal inference in assessing if the presence of EOs causes a reduction in election-day fraud through a quasi-experiment in which observers are assigned to PSs in a way that approximates randomization. This micro-level strategy of comparison strengthens the ceteris paribus clause by automatically controlling for a number of country-specific variables in a way that cross-sectional studies are unable to do due to endogeneity problems (Hyde, 2007; Bader & Schmeets, 2014).

The research design I am proposing can be considered a natural experiment because the independent variable (the "treatment" variable) is assigned "as if" it were random.

The distribution of observers to PSs is highly unlikely to be systematically different from a pure randomization because the methodology of international organizations such as the European Union and the OSCE/ODIHR guarantees that there are no geographic or other kinds of bias in the distribution of observers to PSs and, more importantly, that the choice of PSs to visit is not driven by information about polling-station attributes regarding voting patterns; otherwise, the assignment of the treatment could not be considered near random (Bader & Schmeets, 2014).[1] Each short-term team is given an area of deployment where it will carry out its work. Inside this area, observers are free to visit the number of PSs they deem appropriate, to stay in a single PS for as long as they think it is necessary for a thoughtful judgment, and to return to an already visited PS if they find it necessary. Deployment plans are kept confidential, and EU and OSCE/ODIHR EOMs discourage short-term observers from choosing PSs that are either "convenient" (visiting areas near the observers' hotels or near tourist destinations) or "interesting" (visiting areas in which problems with election fraud are expected) (Hyde, 2007, pp. 45–50). Choosing PSs that are problematic or convenient can hamper randomization because the cheating candidate/party may anticipate those behaviors and concentrate their efforts to cheat in other places, where they expect observers would be less likely to go (Ichino & Schündeln, 2012).[2] In the case study analyzed in this article, the Ukrainian 2004 presidential elections, the difference in observers' distribution is mainly due to voters' density, and, more important, it does not follow a clear pattern that would predict Yanukovych and/or Yushchenko's vote share. Therefore, we can consider the assignment of EOs to PSs very close to randomization.[3]

I will perform a difference of means test (*t*-test) comparing the two groups of observations (PSs visited by OSCE/ODIHR observers against PSs not visited by OSCE/ODIHR observers[4]) and testing the hypothesis that the means of the two groups are the same.[5] If the hypothesis is correct—that is, if observers have a measurable deterrent effect on election-day fraud, reducing fraud at the PSs they visited—then, holding all else

1. Even if observers possess information regarding single PSs (but it is quite unlikely) and thus decide to visit stations they believe to be problematic, the (hypothesized) effect of observers on fraud will be strengthened.

2. Again, however, this would strengthen the observers' effect.

3. The problem of selection bias was addressed by Bader and Schmeets (2014), who analyzed original OSCE/ODIHR data and concluded that selection bias in visited polling stations, while present, does not significantly impact the overall percentage of polling stations where observers find significant flaws.

4. For the sake of simplicity, from now on I will only use the term "observed PSs" meaning "PSs observed by an OSCE/ODIHR team." As rightly pointed out by the anonymous reviewer, there were other international EOMs, even if it is not possible to establish how many (OSCE/ODIHR, 2004). The fact that, in this analysis, some PSs observed by other EOMs may be included in the "not observed PSs" category (the control group) strengthens the findings because, on the one hand, if the effect of other EOMs were null, the results would not change, whereas if, as more probable, the effect of the observation is independent of the body that implements it, then, having included in the control group some PSs that actually belong to the treated group decreases the differences between the two groups and results in a moderation of the differences. Since this article shows the presence of a strong observation effect, if it were possible (but there is no data) to include in the treatment group also the PSs observed by other EOMs, the effect of observation would be even more marked.

5. I will use an unpaired two-sample *t*-test with unequal variance.

constant, their presence should decrease the vote share for the fraud-sponsoring candidate. The assignment of international observers, "the treatment," to randomly selected (or "as if" randomly selected) PSs corresponds to holding all else constant.

The 2004 Ukrainian presidential elections saw three rounds of voting: the first round, with 24 candidates; the second round, the runoff between the two candidates with the most votes; and a "third round," which actually was a repetition of the runoff ordered by the Supreme Court of Ukraine, which declared the second round invalid due to widespread election irregularities. We must therefore consider three rounds of treatment (the independent variable, i.e., PSs observed) and a separate voting distribution (the dependent variable) for each round.[6] Specifically, the statistical population of the polling-station-level results[7] can be divided into a number of experimental groups, according to the round considered and to the "treatment" of international observation:

1. Considering round-one vote share, there are two experimental groups: PSs observed ("observed in R1") and PSs not observed ("not observed in R1").

2. Considering round-two vote share, there are four experimental groups: PSs observed only in the first round ("observed only in R1"), only in the second round ("observed only in R2"), in the first and in the second round ("observed in R1R2"), and never observed ("never observed").

3. Considering the repeated second-round vote share, there are eight experimental groups: PSs observed only in the first round ("observed only in R1"), only in the second round ("observed only in R2"), only in the repeated second-round ("observed only in R3"), in the first and in the second round ("observed in R1R2"), in the first and in the repeated second-round ("observed in R1R3"), in the second and in the repeated second-round ("observed in R2R3"), in all three rounds ("observed in R1R2R3") and never observed ("never observed").

Since we are facing three rounds of elections, it is also possible to test if a previous round of observation has any lasting effect in the following rounds. The presence of international election observers can have "immediate" or "lasting" effects: the first term suggests that observers are able to deter fraud, but only during the election they are observing; on the contrary, the second term suggests that the effect of observers lasts also on the actors' behavior in the following round(s).

---

6. Hyde (2007) performs some tests using round-one and round-two average vote share. I'm not convinced by this comparison simply because in round one the number of candidates is higher than in round two (where there are just two candidates), thus distributing the total number of valid votes cast (and the relative percentages) among a higher number of subjects, making them incomparable with round-two vote shares. In our case, also comparisons using the average vote share between second and repeated second-rounds make no sense because in the second round the results were marred by widespread election irregularities, while, in the repeated second, results were "cleaner."

7. Data used during this study are of three main sources: the dependent variable was downloaded from the Central Election Commission website (http://www.cvk.gov.ua/); the independent variable was provided to me by the OSCE/ODIHR under a confidentiality agreement; the control variables are drawn from the 2001 Ukrainian census (http://www.ukrcensus.gov.ua/s).

To avoid biased results, I will control for three variables:

1. polling station size: if observers have systematic difficulties in reaching small PSs and if small PSs systematically support a particular candidate, the mean difference between observed and unobserved PSs can be the result of systematic differences between PSs that are easy to reach and those that are not. PSs were therefore divided in two groups: small PSs, with fewer registered voters than the mean; and big PSs, with more registered voters than the mean;

2. candidate's strongholds: if a candidate gets a particularly strong electoral support in some areas (maybe because it is his area of origin, or among people of the same ethnicity, religion, etc.), and if a high proportion of unobserved PSs is located in those areas, the results of the mean difference can be driven by this bias. In the Ukrainian case, there is a sizable Russian minority (17.3%) in eastern and southern regions. In order to guarantee himself the support of this minority, Yanukovych's campaign platform included the proposal of making Russian the Ukraine's second State language. I therefore divided PSs into two groups according to a dichotomous measure: regions with more Russian native speakers than the national mean and regions with up to 17.3% of Russian minorities;

3. urban/rural divide: if a candidate performs very well in urban areas (or in rural ones) and the sample of visited PSs overrepresents rural (or urban) PSs, then the candidate's disproportionate support in urban (or rural) areas can bias the results of the mean difference comparisons between observed and unobserved PSs. Thus, I divided the population of PSs in two groups: one group contains PSs located in cities with more than 50,000 inhabitants; other PSs are considered nonurban.

THE UKRAINIAN CASE

At the time of the 2004 Presidential elections, the incumbent authorities in Ukraine were President Leonid Kuchma and Prime Minister Viktor Yanukovych. President Kuchma had already served two terms in the office and was precluded from running again due to constitutional term limits; thus, the incumbent candidate supported by the president and by the Russian Federation was Prime Minister Viktor Yanukovych, who stood as the candidate of the Party of Regions. Yanukovych run against the opposition candidate Viktor Yushchenko, leader of the Our Ukraine faction in the Ukrainian parliament and former prime minister, who stood as a "self-nominated" independent candidate and called for Ukraine to turn its attention westward and eventually join the European Union. Before the elections, Yushchenko and Yulia Tymoshenko (Yulia Tymoshenko Bloc) established the Peoples' Power, an electoral coalition aimed to win the 2004 presidential elections. The pact included a promise by Yushchenko to nominate Tymoshenko as prime minister if he won the election. Though 24 candidates contested the election, pre-election polls clearly indicated that only Yanukovych and Yushchenko enjoyed extensive popular support. The election

was held in a largely unfree atmosphere, with Yanukovych and Kuchma using their control of the government and state apparatus for intimidating Yushchenko and his supporters (OSCE/ODIHR, 2004).

The first round was held on 31 October, and the two main candidates achieved very similar results (about 39%). Many complaints were raised regarding voting irregularities in favor of Yanukovych. However, since it was clear that neither Yanukovych nor Yushchenko were able to reach the 50% of the votes and that, therefore, challenging the first-round results would not have prevented the runoff, the complaints were not actively pursued and both candidates concentrated on the upcoming second round. On 21 November, the second round was held, which resulted in the election of Yanukovych by 3 percentage points. Protests began as soon as second-round election results were released, as Yushchenko's team presented many evidence of election fraud witnessed by many local and foreign observers. Beginning on 22 November, massive peaceful protests (later renamed "the Orange Revolution") started in several cities across Ukraine. On 3 December, Ukraine's Supreme Court decided that it was not possible to establish with certainty the results of the presidential elections because of the scale of the electoral fraud, so it invalidated the official results and ordered a revote of the runoff election to be held on 26 December. The revote attracted conspicuous international attention and gave Yushchenko the presidency by about 8 percentage points.

The Ukrainian 2004 presidential election, which was observed by the OSCE/ODIHR, is an excellent case for testing my hypothesis for several reasons. First, one presidential candidate, Yanukovych, orchestrated widespread fraud on election day and EOs were able to witness and to report on them. Second, the Ukrainian Central Election Commission made disaggregated, polling-station-level election results available on its website. Third, EOs were assigned in a way that approximates randomization and this permits the use of a natural experimental design. Fourth, it was a two-round election, thus making it possible to test also for "lasting" effects, but the second round was repeated, allowing further analysis. Fifth, the presence of observers was widespread: during the first round, OSCE/ODIHR EOs submitted 2,578 reports; during the second round, reports were 2,489; finally, during the repeated second-round, the observers submitted 5,920 report forms, thus making this mission the largest in OSCE/ODIHR history (OSCE/ODIHR, 2004, p. 4). OSCE/ODIHR observers visited[8] more than 6% of the PSs active in the country during the first and second rounds, reaching the 14.6% during the repeated second-round. This high percentage makes Ukrainian 2004 elections a good case to study since the high number of "treated" cases increases the power of the analysis.

---

8. A PS is included in the treatment group if it was observed during the opening, polling, and/or counting phase.

TESTS AND RESULTS

The national territory of Ukraine was divided, for the 2004 presidential elections, into more than 33,000 PSs, precisely: 33,101 in the first round, 33,077 in the second, and 33,059 in the repeated second.[9] Among them, OSCE/ODIHR observers visited 2,203 PSs during the first round, 1,998 during the second round, and 4,856 during the repeated second-round.

The analysis involves 19 different tests, each of them in 7 subgroups, for a total of 133 tests, plus one test in each of the 7 subgroups to have a clue on the different effect of observation during the day (opening and polling) vs. observation during counting.[10] To further corroborate the results, I performed all tests using both the vote share of the fraud-sponsoring candidate, Yanukovych, and of the opposition candidate, Yushchenko, thus resulting in a total of 266 tests. Results are about specular, and therefore I will only show tests performed using Yanukovych's vote share.

TEST PERFORMED USING FIRST-ROUND VOTE SHARE

PSs in round one can only be observed or not, limiting the investigation to immediate effects. If the presence of observers reduces election-day fraud, we should expect to find a significantly lower percentage of votes in favor of the candidate that is cheating in observed PSs.

Difference of means tests comparing treatment and control groups performed using round-one vote share are shown in Table 1. Before going to the analysis of results, I will explain how these tables work because they will be the main instruments used to present results. They report the results of unpaired two-sample *t*-tests with unequal variance. Column one ("If polling stations are") specifies to which groups of PSs the test is applied: to all PSs ("All"), to PSs with more or less registered voters than the mean ("Big" or "Small"), to PSs in regions with or without significant Russian minorities ("Without Russian minorities" or "With Russian minorities"), to PSs located in urban or nonurban areas ("Nonurban" or "Urban"). Columns 3 and 4 indicate which subgroups of PSs will be compared by the *t*-test and present, in this order (as reported in column two): the average Yanukovych's vote share of the two subgroups compared (line 1); the mean difference between the percentages of the two subgroups and, in parenthesis, the value of this mean difference compared to Yanukovych's vote share in unobserved PSs[11] (line 2); the value of the Student's *t*-statistic and, in parenthesis, the number of observations involved in the test (line 3); and the *P* value (line 4).

9. PSs outside Ukraine are not included in the analysis because they did not have an equal chance of being visited (no observers were sent there); however, I did perform the analysis also including these PSs and the results did not change.

10. I thank the anonymous reviewer for this suggestion.

11. Let me explain the utility of this further computation with an example. Suppose that Yanukovych gets 20% of the votes in unobserved PSs and 10% in observed ones. The difference between the means is 10 percentage points. However, this 10-percentage-point difference represents the 50% of Yanukovych's vote share in unobserved PSs. Suppose that, then, in another case, Yanukovych gets the 40% of the votes in unobserved PSs and 30% in observed ones. The difference is still 10 percentage points, but, this time, it represents the 25% of Yanukovych's vote share in unobserved PSs. I will consider this aspect only in cases where we are comparing unobserved and observed PSs, not in cases of comparison between two groups of observed PSs.

TABLE 1. Difference of Means Tests Using Yanukovych's First-Round Vote Share

If polling stations are:		Not observed in R1 vs. Observed in R1	Not observed in R1 vs. Observed during counting in R1
All	Average vote share	39.74% vs 35.08%	39.74% vs 33.11%
	Mean difference	Δ 4.66 (11.73%)	Δ 6.62 (16.66%)
	Student's t	$t(33101) = 6.27$	$t(31078) = 2.85$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0049$
Big	Average vote share	42.20% vs 35.48%	42.20% vs 32.63%
	Mean difference	Δ 6.72 (15.92%)	Δ 9.56 (22.65%)
	Student's t	$t(13893) = 9.10$	$t(12331) = 4.13$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0001$
Small	Average vote share	34.04% vs 30.68%	34.04% vs 38.59%
	Mean difference	Δ 3.35 (9.84%)	Δ –4.56 (–13.40%)
	Student's t	$t(19208) = 2.36$	$t(18747) = -0.82$
	P value	$P > |t| = 0.0185$	$P > |t| = 0.4196$
Without Russian minorities	Average vote share	24.15% vs 23.26%	24.15% vs 22.31%
	Mean difference	Δ 0.89 (3.68%)	Δ 1.84 (7.62%)
	Student's t	$t(20559) = 1.05$	$t(19514) = 0.63$
	P value	$P > |t| = 0.2943$	$P > |t| = 0.5302$
With Russian minorities	Average vote share	58.50% vs 46.09%	58.50% vs 44.08%
	Mean difference	Δ 12.41 (21.21%)	Δ 14.42 (24.65%)
	Student's t	$t(12542) = 11.95$	$t(11564) = 4.36$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0000$
Nonurban	Average vote share	35.20% vs 30.82%	35.20% vs 33.92%
	Mean difference	Δ 4.38 (12.44%)	Δ 1.28 (3.64%)
	Student's t	$t(17424) = 3.47$	$t(16687) = 0.28$
	P value	$P > |t| = 0.0005$	$P > |t| = 0.7768$
Urban	Average vote share	43.13% vs 37.08%	43.13% vs 32.78%
	Mean difference	Δ 6.06 (14.05%)	Δ 10.35 (24.00%)
	Student's t	$t(15677) = 6.72$	$t(14391) = 3.83$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0002$

The evidence presented in Table 1 shows that the presence of EOs reduced the vote share for the fraud-sponsoring candidate in the first round by an average of more than 4.6 percentage points (representing the 11.73% of Yanukovych's vote share in

unobserved PSs). This result is statistically significant at the 1% confidence level, allowing a rejection of the null hypothesis that there is no difference between observed and unobserved PSs. Controlling for PSs' size, for the presence of Russian minorities and for the urban divide, the results are in the expected direction and the difference is statistically significant, apart from the case of PSs located in regions without sizable Russian minorities. This can be explained easily: those regions were the ones in the North and in the West of the country, where the opposition candidate, Yushchenko, was stronger. In those regions, in fact, voters (and PSs officials too) did not support Yanukovych, therefore reducing his possibility of manipulating the outcome of the election through election-day fraud. This was also confirmed by the findings of OSCE/ODIHR observers who found more irregularities in eastern regions (OSCE/ODIHR, 2004, p. 25).

It is impressive the fraud-reduction power of observation in PSs located in Yanukovych strongholds: there, the difference between unobserved and observed PSs reaches the remarkable threshold of more than 12.4 percentage points (21.21% if compared to Yanukovych's vote share in unobserved PSs), and it is significant at the 1% confidence level. We can finally notice that the mean difference, while always being statistically significant, is larger in big and urban PSs than in small and nonurban ones. Results can be biased if unobserved PSs "naturally" supported the fraud-sponsoring candidate. But this is not the case; on the contrary, if we compare Yanukovych's performance in unobserved small (34.04%) and big (42.20%) PSs and in nonurban (35.20%) and urban (43.13%) ones, we can easily see that Yanukovych performed better in big and urban PSs (than in small and nonurban ones), exactly where the number of PSs observed was higher. This further supports our initial hypothesis and dismisses any possibility of bias in the results.

As we know from the OSCE/ODIHR Final Report (2004), observers reported many kinds of electoral malpractices and thus they were able to both detect and deter fraud. Why did this happen? It is plausible that PS officials stopped their fraudulent behaviors in front of OEs, but then, as soon as observers left the PS, they went on rigging the election. This makes the effect of observation during opening and polling less strong than observation during counting, when EOs are instructed to follow the entire process of counting in only one PS. To verify if the (enduring) physical presence of observers had a stronger effect than their visit during a short time span at some point in the day, we can rerun the analysis conducted with R1 vote share and consider in the treatment group only those PSs visited during counting and excluding from the analysis PSs observed during opening or polling because using them in the control group would bias the results by decreasing the vote share of the fraud-sponsoring candidate. Results are shown in Table 1 column 4. As we can see, the mean difference in PSs observed during counting is, as hypothesized, bigger than the mean difference computed using a treatment group composed by PSs visited during opening, polling, and/or counting. The only exceptions are small and nonurban PSs. As already emphasized, this does not represent a problem because Yanukovych performed better in big and urban PSs.

TEST PERFORMED USING SECOND-ROUND VOTE SHARE

The presence of a second round allows us to test for immediate and lasting effects. PSs can be divided into four groups: never observed, observed only in round one, observed only in round two, or observed in both rounds. Let us begin by testing the immediate effect of round two observation (Table 2 column 3). In this case, we compare the second-round vote share between PSs observed neither in the first nor in the second round and PSs observed in the second round (but not in the first). We expect the performance of the fraud-sponsoring candidate to be worse in observed PSs.

Table 2 confirms, in general, the presence of an immediate effect of round two observation on round two Yanukovych's vote share. The fraud-reduction effect is not significant in small, "non-Russian," and nonurban PSs where the opposition candidate was stronger (thus, these results do not disconfirm the overall fraud-reduction effect of EO).

Second, we can test if first-round observation has a lasting deterrent effect in the second round by comparing the round-two vote share of PSs never observed and of PSs observed only in the first round. Data presented in Table 2 column 4 show that the lasting effect of observation is reducing Yanukovych's vote share by 4.14 percentage points. This means that PS officials who met EOs in the first round were less likely to commit fraud in the second round. All differences go in the expected direction and are statistically significant, apart from small PSs and PSs located where the presence of Russian minorities is below the national mean. Again, those exceptions do not endanger the validity of the general results since, in those PSs, the cheating candidate gained, in mean, a lower percentage vote share.

Third, by comparing the second-round vote share between unobserved PSs and PSs observed in both rounds, we can measure the "total effect" of observation (the lasting effect of round one observation and the immediate effect of round two observation).[12] Table 2 column 5 provides additional empirical support for the finding that EOs have a strong deterrent effect on of election-day fraud. In fact, Yanukovych received about 9 percentage points more in PSs that were never observed. Results are not significant in small, non-Russian, and nonurban PSs (in the first two subgroups, results even have the opposite sign). This does not represent a problem as long as Yanukovych's vote share is larger in big, Russian, and urban PSs.

Fourth, we can also control whether observing twice the same PSs (thus combining immediate and lasting effects) has a stronger deterrent effect than the immediate effect of round two (only) observation by comparing the second-round vote share between PSs that were observed in both rounds and PSs that were observed only in round two. If the difference is minimal or barely significant, then the immediate effect of round two observation can be considered a sufficient deterrent.[13] Data presented in Table 2 column

12. Hyde (2007) performs this test using, first, the round-two vote share, and, second, the two-round average vote share. As said, I do not agree with the second comparison.

13. Hyde's results (2007) suggest "if first-round monitoring took place then second-round monitoring had only a marginal additional deterrent effect." In my opinion this conclusion can, eventually, be drawn from another

TABLE 2. Difference of Means Tests Using Yanukovych's Second-Round Vote Share

If polling stations are:		Never observed vs. Observed only in R2	Never observed vs. Observed only in R1	Never observed vs. Observed in R1R2	Observed in R1R2 vs. Observed only in R2
All	Average vote share	50.23% vs 47.52%	50.23% vs 46.09%	50.23% vs 41.20%	41.20% vs 47.52%
	Mean difference	Δ 2.71 (5.39%)	Δ 4.14 (8.24%)	Δ 9. 03 (17.98%)	Δ -6.31
	Student's t	t(30874) = 2.39	t(31079) = 3.67	t(29842) = 5.29	t(1998) = -3.18
	P value	P > |t| = 0.0166	P > |t| = 0.0002	P > |t| = 0.0000	P > |t| = 0.0015
Big	Average vote share	54.26% vs 47.75%	54.26% vs 47.01%	54.26% vs 41.19%	41.19% vs 47.76%
	Mean difference	Δ 6.50 (11.98%)	Δ 7.26 (13.38%)	Δ 13.07 (24.09%)	Δ -6.57
	Student's t	t(12224) = 5.86	t(12351) = 6.55	t(11524) = 7.60	t(1566) = -3.33
	P value	P > |t| = 0.0000	P > |t| = 0.0000	P > |t| = 0.0000	P > |t| = 0.0009
Small	Average vote share	40.97% vs 45.23%	40.97% vs 37.56%	40.97% vs 41.59%	41.59% vs 45.23%
	Mean difference	Δ -.26 (-10.40%)	Δ 3.40 (8.30%)	Δ -0.62 (-1.51%)	Δ -3.64
	Student's t	t(18650) = -1.78	t(18728) = 1.87	t(18318) = -0.12	t(432) = -0.65
	P value	P > |t| = 0.0752	P > |t| = 0.0618	P > |t| = 0.9031	P > |t| = 0.5187
Without Russian minorities	Average vote share	29.77% vs 32.37%	29.77% vs 28.55%	29.77% vs 32.98%	32.98% vs 32.37%
	Mean difference	Δ -2.60 (-8.73%)	Δ 1.21 (4.06%)	Δ -3.21 (-10.78%)	Δ 0.61
	Student's t	t(19425) = -1.95	t(19535) = 1.00	t(18866) = -1.38	t(1023) = 0.23
	P value	P > |t| = 0.0519	P > |t| = 0.3170	P > |t| = 0.1691	P > |t| = 0.8183

(continued)

TABLE 2. (*continued*)

If polling stations are:		Never observed vs. Observed only in R2	Never observed vs. Observed only in R1	Never observed vs. Observed in R1R2	Observed in R1R2 vs. Observed only in R2
With Russian minorities	Average vote share	73.09% vs 60.64%	73.08% vs 61.29%	73.09% vs 48.39%	48.39% vs 60.64%
	Mean difference	Δ 12.44 (17.02%)	Δ 11.80 (16.15%)	Δ 24.69 (33.78%)	Δ -12.25
	Student's t	$t(11449) = 8.51$	$t(11544) = 7.84$	$t(10976) = 10.58$	$t(975) = -4.62$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0000$	$P > |t| = 0.0000$	$P > |t| = 0.0000$
Nonurban	Average vote share	43.72% vs 42.11%	43.72% vs 38.79%	43.72% vs 37.92%	37.92% vs 42.11%
	Mean difference	Δ 1.61 (3.68%)	Δ 4.93 (11.28%)	Δ 5.81 (13.29%)	Δ -4.20
	Student's t	$t(16625) = 0.80$	$t(16732) = 2.62$	$t(16261) = 1.92$	$t(692) = -1.18$
	P value	$P > |t| = 0.4223$	$P > |t| = 0.0089$	$P > |t| = 0.0567$	$P > |t| = 0.2402$
Urban	Average vote share	55.09% vs 49.75%	55.09% vs 49.48%	55.09% vs 42.68%	42.68% vs 49.75%
	Mean difference	Δ 5.34 (9.69%)	Δ 5.61 (10.18%)	Δ 12.41 (22.53%)	Δ -7.07
	Student's t	$t(14249) = 4.00$	$t(14347) = 4.12$	$t(13581) = 6.05$	$t(1306) = -3.01$
	P value	$P > |t| = 0.0001$	$P > |t| = 0.0000$	$P > |t| = 0.0000$	$P > |t| = 0.0027$

6 show a quite big overall difference (–6.31 percentage points for Yanukovych), which is statistically significant at the 1% confidence level. As in most cases, the relationship does not hold in small, non-Russian, and nonurban PSs, but this, as said, does not invalidate the results.

Fifth, by comparing round-two vote share between PSs observed in both rounds and PSs observed only during the first round, we can test whether observing the same PSs twice brings an additional (marginal) deterrent effect compared to round one only observation. If observing the same PS (thus combining immediate and lasting effects) twice has a stronger deterrent effect than just the lasting effect of round one (only) observation, then the cheating candidate's vote share should be significantly lower in PSs observed in both rounds. Table 3 column 3 shows that, if EOs were present in the first round, round two observation does have a statistically significant effect. Thus, again, combining immediate and lasting effects has a stronger deterrent effect than just the lasting effect of round one (only) observation: *repetita iuvant*. In fact, Yanukovych's vote share is significantly lower (about 4.9 percentage points) in PSs observed in both rounds. Also in this case, the fact that the detected difference in small, non-Russian, and non-urban PSs is not statistically significant (and sometimes in the opposite direction) does not represent a real challenge for the overall results. The fraud-reduction effect in PSs located in areas with sizable Russian minorities is impressive (about 12.9 percentage points).

Sixth, we can check for a "general" effect of observation that does not distinguish between immediate and lasting effects by comparing the vote share in PSs that were not observed in either round with the vote share in PSs observed in one or both rounds. Since we hypothesize that observation always has some effects, we expect that, if observation has taken place in one or more rounds, then the cheating candidate's round-two vote share should be lower. Table 3 column 4 shows that Yanukovych received 4.3 percentage points more in PSs not observed in either round as compared with his average vote share in PSs observed in one or both rounds. However, also this time, the relationship is reversed in small and non-Russian PSs, where it is, nevertheless, not significant. Again, this does not represent a problem for our results, but, quite the opposite, it strengthens them. Before going to the next test, notice the really high fraud-reduction effect of observation in PSs located in Russian areas: about 14 percentage points.

Seventh, by comparing round-two vote share between PSs observed only in round one and PSs observed only in round two, we can also compare lasting effects with immediate effects in order to control which one is stronger. Results shown in Table 3 column 5 are not significant, meaning that there is little difference in overall fraud reduction between the two groups that were observed only once, regardless of whether observation took place in the first round or the second round. Therefore, we can support the hypothesis that immediate and lasting effect are of similar strength.

---

comparison that will be treated later: observed only in R1 vs. observed in R1R2; in this case, in fact, since both comparison groups have gotten round one observation, it is possible to check the "additional deterrent effect" (if any) of round two re-observation.

TABLE 3. Difference of Means Tests Using Yanukovych's Second-Round Vote Share

If polling stations are:	Observed in R1R2 vs. Observed only in R1	Never observed vs. Observed only in R1, or only in R2, or in R1R2	Observed only in R1 vs. Observed only in R2
All			
Average vote share	41.20% vs 46.09%	50.23% vs 45.93%	46.09% vs 47.52%
Mean difference	Δ -4.89	Δ 4.30 (8.56%)	Δ -1.42
Student's t	t(2203) = -2.47	t(33077) = 5.51	t(3235) = -0.94
P value	P > |t| = 0.0137	P > |t| = 0.0000	P > |t| = 0.3466
Big			
Average vote share	41.19% vs 47.01%	54.26% vs 46.39%	47.01% vs 47.76%
Mean difference	Δ -5.82	Δ 7.87 (14.50%)	Δ -0.75
Student's t	t(1693) = -2.95	t(13917) = 10.15	t(2393) = -0.51
P value	P > |t| = 0.0033	P > |t| = 0.0000	P > |t| = 0.6102
Small			
Average vote share	41.59% vs 37.46%	40.97% vs 41.15%	37.56% vs 45.23%
Mean difference	Δ 4.03	Δ -1.85 (-4.51%)	Δ -7.67
Student's t	t(510) = 0.75	t(19160) = -0.13	t(842) = -2.59
P value	P> |t| = 0.4578	P > |t| = 0.8986	P > |t| = 0.0098
Without Russian minorities			
Average vote share	32.98% vs 28.55%	29.77% vs 30.74%	28.55% vs 32.37%
Mean difference	Δ 4.43	Δ -0.97 (-3.26%)	Δ -3.82
Student's t	t(1133) = 1.71	t(20558) = -1.09	t(1692) = -2.19
P value	P > |t| = 0.0872	P > |t| = 0.2750	P > |t| = 0.0283

(continued)

TABLE 3. (*continued*)

If polling stations are:		Observed in R1R2 vs. Observed only in R1	Never observed vs. Observed only in R1, or only in R2, or in R1R2	Observed only in R1 vs. Observed only in R2
With Russian minorities	Average vote share	48.39% vs 61.29%	73.09% vs 59.11%	61.29% vs 60.64%
	Mean difference	Δ -12.89	Δ 13.98 (19.13%)	Δ 0.64
	Student's t	$t(1070) = -4.82$	$t(12519) = 13.34$	$t(1543) = 0.33$
	P value	$P > |t| = 0.0000$	$P > |t| = 0.0000$	$P > |t| = 0.7431$
Nonurban	Average vote share	37.92% vs 38.80%	43.72% vs 39.92%	38.80% vs 42.11%
	Mean difference	Δ -0.88	Δ 3.80 (8.69%)	Δ -3.32
	Student's t	$t(799) = -0.25$	$t(17424) = 2.91$	$t(1163) = -1.25$
	P value	$P > |t| = 0.8015$	$P > |t| = 0.0037$	$P > |t| = 0.2130$
Urban	Average vote share	42.68% vs 49.48%	55.09% vs 48.56%	49.48% vs 49.75%
	Mean difference	Δ -6.80	Δ 6.51 (11.82%)	Δ -0.27
	Student's t	$t(1404) = -2.88$	$t(15653) = 6.85$	$t(2072) = -0.15$
	P value	$P > |t| = 0.0041$	$P > |t| = 0.0000$	$P > |t| = 0.8800$

TEST PERFORMED USING REPEATED SECOND-ROUND VOTE SHARE

The repeated second-round was held under the authority of a newly appointed Central Election Commission, which administered the election process efficiently and with significantly more transparency. Overall, observers assessed the process much more favorably than during the two previous rounds. However, they remarked a clear regional variation with observers in southern, eastern, and central regions assessing the process less favorably than in northern and western regions (OSCE/ODIHR, 2004, pp. 36–37). Therefore, it could be possible that the fraud-deterrence role of EOs will not be detectable simply because fraud was not present or was present at a very low level. This should not be true in southern and eastern regions (i.e., in regions with sizable Russian minorities), where fraud persisted.

Let us start by checking for immediate effects by comparing the repeated second-round vote share of the fraud-sponsoring candidate between PSs never observed and PSs observed only during the repeated second-round. If, notwithstanding the level of election irregularities dramatically decreased, the presence of observers still had a deterrent effect, then we should find a lower percentage of votes sustaining the cheating candidate. From the results shown in Table 4 column 3, we can hypothesize that, since the level of election irregularities dramatically decreased, the presence of observers did not play, in general, a deterrent effect: Yanukovych, in fact, generally gained very similar results in observed and unobserved PSs, and the small differences are not at all statistically significant. However, the difference remained quite big and is statistically significant at the 1% confidence level in Yanukovych strongholds (where fraud was still present): in big, Russian, and urban PSs. In the other cases, the difference goes in the opposite direction and/or does not reach statistically significance.

In order to test for the presence of previous rounds' lasting effects, we can compare never-observed PSs with PSs observed during round one, round two, or both round one and round two. Apart from PSs located in the North and the West of the country, where the relationship is, in any case, not significant, the results go in the expected direction (see Table 4 column 4). They are, as usual, not significant in small and non-Russian PSs, and, since a greater number of PSs used during this mean difference test are in those groups, this can explain why the general test turns out to be not significant. Note, however, the still high (lasting) fraud-reduction effect of observation in PSs located in the South and East of Ukraine, that is, in Yanukovych's stronghold.

We can also divide "lasting effects" in "long-term lasting effects," the ones of round-one observation on repeated second-round vote share, and "medium-term lasting effects," the ones of round two observation on repeated second-round vote share. Comparing the repeated second-round vote share in PSs never observed and in PSs observed only in round one, allows us to check for "long-term" lasting effects (Table 4 column 5). Comparing the repeated second-round vote share in PSs never observed and in PSs observed only in round two, means testing for "medium-term" lasting effects (Table 4 column 6). Data clearly show that these effects are present only in big PSs and in PSs located in regions with considerable Russian minorities. These effects are, moreover,

TABLE 4. Difference of Means Tests Using Yanukovych's Repeated Second-Round Vote Share

If polling stations are:		Never observed vs. Observed only in R3	Never observed vs. Observed only in R1, or only in R2, or in R1R2	Never observed vs. Observed only in R1	Never observed vs. Observed only in R2
All	Average vote share	44.80% vs 44.79%	44.80% vs 43.41%	44.80% vs 43.38%	44.80% vs 44.18%
	Mean difference	Δ 0.01 (0.02%)	Δ 1.39 (3.10%)	Δ 1.42 (3.17%)	Δ 0.62 (1.38%)
	Student's t	t(29411) = 0.01	t(28203) = 1.43	t(26935) = 0.99	t(26861) = 0.45
	P value	P > |t| = 0.9893	P > |t| = 0.1535	P > |t| = 0.3225	P > |t| = 0.6495
Big	Average vote share	51.09% vs 45.89%	51.90% vs 44.85%	51.09% vs 45.48%	51.09% vs 45.22%
	Mean difference	Δ 5.19 (10.16%)	Δ 6.24 (12.02%)	Δ 5.61 (10.98%)	Δ 5.87 (11.49%)
	Student's t	t(11124) = 6.32	t(10403) = 6.51	t(9471) = 4.01	t(9444) = 4.43
	P value	P > |t| = 0.0000	P > |t| = 0.0000	P > |t| = 0.0001	P > |t| = 0.0000
Small	Average vote share	32.42% vs 36.89%	32.42% vs 31.33%	32.42% vs 27.37%	32.42% vs 35.94%
	Mean difference	Δ -4.46 (-13.76%)	Δ 1.09 (3.36%)	Δ 5.05 (15.58%)	Δ -3.52 (-10.86%)
	Student's t	t(18287) = -3.25	t(17800) = 0.69	t(17464) = 2.60	t(17417) = -1.35
	P value	P > |t| = 0.0012	P > |t| = 0.4883	P > |t| = 0.0096	P > |t| = 0.1762
Without Russian minorities	Average vote share	23.07% vs 24.00%	23.07% vs 24.82%	23.07% vs 23.41%	23.07% vs 25.50%
	Mean difference	Δ -0.93 (-4.03%)	Δ -1.75 (-7.58%)	Δ -0.38 (-1.65%)	Δ -2.42 (-10.49%)
	Student's t	t(18671) = -1.03	t(17237) = -1.68	t(17542) = -0.23	t(17486) = -1.67
	P value	P > |t| = 0.3016	P > |t| = 0.0926	P > |t| = 0.8198	P > |t| = 0.0955

(continued)

TABLE 4. (*continued*)

If polling stations are:		Never observed vs. Observed only in R3	Never observed vs. Observed only in R1, or only in R2, or in R1R2	Never observed vs. Observed only in R1	Never observed vs. Observed only in R2
With Russian minorities	Average vote share	72.88% vs 61.09%	72.88% vs 62.62%	72.88% vs 64.77%	72.88% vs 63.46%
	Mean difference	Δ 11.79 (16.18%)	Δ 10.24 (14.05%)	Δ 8.11 (11.13%)	Δ 9.42 (12.92%)
	Student's t	$t(10740) = 10.78$	$t(9966) = 7.94$	$t(9393) = 4.24$	$t(9375) = 5.38$
	P value	$P > \vert t \vert = 0.0000$	$P > \vert t \vert = 0.0000$	$P > \vert t \vert = 0.0000$	$P > \vert t \vert = 0.0000$
Nonurban	Average vote share	36.31% vs 38.02%	36.31% vs 33.13%	36.31% vs 32.23%	36.31% vs 33.30%
	Mean difference	Δ -1.71 (-4.71%)	Δ 3.18 (8.76%)	Δ 4.08 (11.24%)	Δ 3.01 (8.29%)
	Student's t	$t(16122) = -1.30$	$t(15614) = 2.15$	$t(15157) = 1.92$	$t(15086) = 1.53$
	P value	$P > \vert t \vert = 0.1922$	$P > \vert t \vert = 0.0320$	$P > \vert t \vert = 0.0555$	$P > \vert t \vert = 0.1530$
Urban	Average vote share	51.82% vs 47.82%	51.82% vs 48.14%	51.82% vs 49.18%	51.82% vs 48.67%
	Mean difference	Δ 4.00 (7.72%)	Δ 3.67 (7.08%)	Δ 2.63 (5.07%)	Δ 3.14 (6.06%)
	Student's t	$t(13289) = 3.88$	$t(12589) = 3.06$	$t(11778) = 1.48$	$t(11775) = 1.93$
	P value	$P > \vert t \vert = 0.0001$	$P > \vert t \vert = 0.0023$	$P > \vert t \vert = 0.1402$	$P > \vert t \vert = 0.0533$

slightly stronger in the case of round two observation, suggesting that "medium-term lasting effects" played a somewhat stronger role.

The following test is a combination of the previous two and, by putting together "long-term" and "medium-term" lasting effects, it can show if observing the same PSs in round one and round two has a deterrent effect on the repeated second-round. If this effect exists, then the cheating candidate's vote share in PSs never observed should be significantly higher than in PSs observed in both round one and round two. Again, the effect is present and significant only in big, Russian, and urban PSs, where Yanukovych was stronger, and fraud was still present (Table 5 column 3). Impressively, Yanukovych gained 20.91 percentage points more (about 28.7% of his vote share in unobserved PSs) in PSs never observed located in areas with a high concentration of Russian minorities.

Comparing the repeated second-round vote share in PSs never observed with PSs observed during the first, second, and repeated second-round, we can test the magnitude of the lasting effects ("long-term" and "medium-term") of round one and round two observation together with the immediate effect of the repeated second-round observation. Table 5 column 4 shows that the difference is present and quite strong. Notwithstanding the lower level of fraud, the deterrent effect of observation played a role: generally, Yanukovych's vote share was reduced by about 12.7 percentage points (28.3% of his vote share in unobserved PSs). This difference is significant at the 1% confidence level and it is even stronger in big, Russian, and urban PSs, whereas it remains negative and not significant in small and non-Russian PSs.

The following test measures the (marginal) effect of observing the same PSs three times instead of observing them only during the first round. Since Yanukovych's vote share in PSs observed in the first, second, and repeated second-round is about 11.3 percentage points lower than his vote share in PSs observed only in round one, then observing the same PSs three times is useful and brings an additional deterrent effect. This effect is, in general, statistically significant at the 1% confidence level, but it holds only in big, Russian, and urban PSs (Table 5 column 5).

Similarly, we can measure the (marginal) effect of observation in round one and round two. If the fraud-sponsoring candidate's vote share in PSs observed only in the repeated second-round is higher than the same candidate's vote share in PSs observed three times, then, observing the same PSs three times increases the deterrent effect of observation. Table 5 column 6 shows that this is the case in general and in big, Russian, urban, and nonurban PSs. The effect is particularly strong in regions with sizable Russian minorities.

By comparing PSs observed in the first, second, and repeated second-round with PSs observed in round one and round two, we can assess the (marginal) effect of observing the same PSs a third time. If this marginal effect is positive, then the cheating candidate's vote share in PSs observed in the first, second, and repeated second-round should be lower than the same candidate's vote share in PSs observed in round one and round two. From Table 6 column 3 we can say that, generally, this marginal effect is positive and significant at the 1% confidence level. However, as usual, it is not significant (and in one case it goes in the opposite direction) in small, non-Russian, and nonurban PSs.

TABLE 5. Difference of Means Tests Using Yanukovych's Repeated Second-Round Vote Share

If polling stations are:		Never observed vs. Observed in R1R2	Never observed vs. Observed in R1R2R3	Observed in R1R2R3 vs. Observed only in R1	Observed in R1R2R3 vs. Observed only in R3
All	Average vote share	44.80% vs 40.66%	44.80% vs 32.12%	32.12% vs 43.38%	32.12% vs 44.79%
	Mean difference	Δ 4.14 (9.24%)	Δ 12.68 (28.30%)	Δ -11.26	Δ -12.67
	Student's *t*	t(26043) = 1.59	t(26057) = 6.32	t(1356) = -4.68	t(3832) = -6.02
	P value	P > |t| = 0.1132	P > |t| = 0.0000	P > |t| = 0.0000	P > |t| = 0.0000
Big	Average vote share	51.09% vs 41.24%	51.09% vs 31.78%	31.78% vs 45.48%	31.78% vs 45.89%
	Mean difference	Δ 9.85 (19.28%)	Δ 19.30 (37.78%)	Δ -13.70	Δ -14.11
	Student's *t*	t(8904) = 3.80	t(8928) = 9.56	t(983) = -5.74	t(2636) = -6.72
	P value	P > |t| = 0.0002	P > |t| = 0.0000	P > |t| = 0.0000	P > |t| = 0.0000
Small	Average vote share	32.42% vs 29.20%	32.42% vs 43.37%	43.38% vs 27.37%	43.38% vs 36.89%
	Mean difference	Δ 3.22 (9.93%)	Δ -10.95 (-33.77%)	Δ 16.01	Δ 6.49
	Student's *t*	t(17139) = 0.53	t(17129) = -0.97	t(373) = 1.40	t(1196) = 0.57
	P value	P > |t| = 0.5996	P > |t| = 0.3451	P> |t| = 0.1783	P> |t| = 0.5751
Without Russian minorities	Average vote share	23.07% vs 28.26%	23.07% vs 26.10%	26.10% vs 23.41%	26.09% vs 24.00%
	Mean difference	Δ -5.19 (-22.50%)	Δ -3.02 (-13.09%)	Δ 2.69	Δ 2.10
	Student's *t*	t(17019) = -1.45	t(17012) = -0.98	t(744) = 0.79	t(1873) = 0.66
	P value	P > |t| = 0.1491	P > |t| = 0.3291	P > |t| = 0.4285	P > |t| = 0.5100

(continued)

TABLE 5. (*continued*)

If polling stations are:		Never observed vs. Observed in R1R2	Never observed vs. Observed in R1R2R3	Observed in R1R2R3 vs. Observed only in R1	Observed in R1R2R3 vs. Observed only in R3
With Russian minorities	Average vote share	72.88% vs 51.97%	72.88% vs 37.16%	37.16% vs 64.77%	37.16% vs 61.08%
	Mean difference	Δ 20.91 (28.69%)	Δ 35.72 (49.01%)	Δ -27.61	Δ -23.93
	Student's *t*	*t*(9024) = 6.36	*t*(9045) = 13.96	*t*(612) = -8.95	*t*(1959) = -9.01
	P value	*P* > |*t*| = 0.0000	*P* > |*t*| = 0.0000	*P* > |*t*| = 0.0000	*P* > |*t*| = 0.0000
Nonurban	Average vote share	36.31% vs 36.64%	36.31% vs 26.96%	26.96% vs 32.23%	26.96% vs 38.02%
	Mean difference	Δ -0.33 (-0.91%)	Δ 9.34 (25.72%)	Δ -5.26	Δ -11.05
	Student's *t*	*t*(14779) = -0.07	*t*(14782) = 2.47	*t*(531) = -1.23	*t*(1496) = -2.80
	P value	*P* > |*t*| = 0.9413	*P* > |*t*| = 0.0157	*P* > |*t*| = 0.2217	*P* > |*t*| = 0.0062
Urban	Average vote share	51.82% vs 42.34%	51.82% vs 34.46%	34.46% vs 49.18%	34.46% vs 47.82%
	Mean difference	Δ 9.48 (18.29%)	Δ 17.35 (33.48%)	Δ -14.72	Δ -13.36
	Student's *t*	*t*(11264) = 3.03	*t*(11275) = 7.43	*t*(825) = -5.19	*t*(2336) = -5.49
	P value	*P* > |*t*| = 0.0029	*P* > |*t*| = 0.0000	*P* > |*t*| = 0.0000	*P* > |*t*| = 0.0000

TABLE 6. Difference of Means Tests Using Yanukovych's Repeated Second-Round Vote Share

If polling stations are:		Observed in R1R2R3 vs. Observed in R1R2	Observed in R1R2R3 vs. Observed in R2R3	Never observed vs. Observed only in R1, or only in R2, or only in R3, or in R1R2, or in R2R3, or in R1R3, or in R1R2R3
All	Average vote share	32.12% vs 40.67%	32.12% vs 39.62%	44.80% vs 43.01%
	Mean difference	Δ -8.54	Δ -7.50	Δ 1.79 (3.99%)
	Student's t	t(464) = -2.64	t(730) = -3.01	t(33059) = 2.85
	P value	P > |t| = 0.0087	P> |t| = 0.0027	P > |t| = 0.0044
Big	Average vote share	31.78% vs 41.24%	31.79% vs 39.23%	51.09% vs 43.88%
	Mean difference	Δ -9.45	Δ -7.45	Δ 7.21 (14.11%)
	Student's t	t(416) = -2.92	t(632) = -3.00	t(13902) = 11.41
	P value	P > |t| = 0.0037	P > |t| = 0.0028	P > |t| = 0.0000
Small	Average vote share	43.38% vs 29.20%	43.37% vs 46.36%	32.42% vs 35.37%
	Mean difference	Δ 14.17	Δ -2.98%	Δ -2.94 (-9.07%)
	Student's t	t(48) = 1.11	t(98) = -0.24	t(19157) = -2.85
	P value	P > |t| = 0.2779	P > |t| = 0.8145	P > |t| = 0.0043
Without Russian minorities	Average vote share	26.10% vs 28.26%	26.10% vs 26.01%	23.07% vs 24.52%
	Mean difference	Δ -2.17	Δ 0.09	Δ -1.45 (-6.28%)
	Student's t	t(221) = -0.46	t(328) = 0.02	t(20555) = -2.17
	P value	P > |t| = 0.6451	P > |t| = 0.9806	P > |t| = 0.0299

TABLE 6. *(continued)*

If polling stations are:		Observed in R1R2R3 vs. Observed in R1R2	Observed in R1R2R3 vs. Observed in R2R3	Never observed vs. Observed only in R1, or only in R2, or only in R3, or in R1R2, or in R2R3, or in R1R3, or in R1R2R3
With Russian minorities	Average vote share	37.16% vs 51.97%	37.16% vs 49.06%	72.88% vs 58.48%
	Mean difference	Δ -14.81	Δ -11.90	Δ 14.40 (19.76%)
	Student's t	t(243) = -3.63	t(402) = -3.79	t(12504) = 16.82
	P value	P > |t| = 0.0004	P > |t| = 0.0002	P > |t| = 0.0000
Nonurban	Average vote share	26.96% vs 36.64%	26.96% vs 37.69%	36.31% vs 35.67%
	Mean difference	Δ -9.68	Δ -10.73	Δ 0.64 (1.76%)
	Student's t	t(153) = -1.65	t(235) = -2.27	t(17423) = 0.67
	P value	P > |t| = 0.1013	P > |t| = 0.0246	P > |t| = 0.5044
Urban	Average vote share	34.46% vs 42.34%	34.46% vs 40.40%	51.82% vs 46.26%
	Mean difference	Δ -7.87	Δ -5.93	Δ 5.56 (10.73%)
	Student's t	t(311) = -2.06	t(495) = -2.06	t(15636) = 6.95
	P value	P > |t| = 0.0406	P > |t| = 0.0403	P > |t| = 0.0000

By comparing the repeated second-round vote share of PSs observed in first, second and repeated second-round with PSs observed only during the second and the repeated second-rounds, we can check the (marginal) effect of observing the same PSs three times. Since, in general, Yanukovych's vote share is significantly higher in PSs observed in the second and in the repeated second-round, we can infer that observing the same PSs three times is useful and brings an additional deterrent effect (Table 6 column 4). This relationship holds in big, Russian, nonurban, and urban PSs, and it is particularly strong in PSs located in the South and East of Ukraine.

The final test compares the repeated second-round vote share between PSs never observed and PSs that were observed in any round or in any combination of rounds, showing if the presence of observers in at least one occasion has a deterrent effect. This test supports the hypothesis that observers reduce election-day fraud, through both immediate and lasting effects. According to the results shown in Table 6 column 5, in general, there is a positive deterrent effect of observation. Those findings are even stronger and significant in big, Russian, and urban PSs. This adds one additional piece of information supporting the hypothesis that observers reduce election-day fraud, through both immediate and/or lasting effects.

DISCUSSION AND CONCLUSIONS

We have repeatedly claimed that the central measurable effect of observers on election-day fraud should be to decrease the vote share for the fraud-sponsoring candidate. We have seen that, holding all else constant, OSCE/ODIHR observers did in fact reduce fraud at the PSs they visited since Yanukovych performed significantly worse (and Yushchenko significantly better) in observed PSs. This natural experiment design, therefore, offered a good test of whether EOs are able to deter election-day fraud, providing causal evidence of how international actors can influence domestic politics. Its results added an additional argument supporting international observation missions, their role in deterring election-day fraud (and not only in detecting it), and, consequently, the importance of their presence for a free and fair election process that could lead to stronger democratization. More specifically, results of the tests performed above show the following:

1. EO had an immediate effect both in the first and in the second round. The same is not true for the repeated second-round probably because the level of election irregularities dramatically decreased; however, the immediate effect of observation was present and quite strong where fraud was still present: in the South and in the East of the country.

2. Observation during the counting has a stronger effect than observation during opening or polling.

3. EO in the first round had lasting effects on the second round; on the contrary, observation in the first and/or the second round did not have lasting effects in

the repeated second-round if all PSs are considered, but the effect was present and remained significant in PSs located in the South and in the East of the country.

4. Observing the same PSs in all two or three rounds gives positive results confirming the hypothesis that EOs did reduce fraud.

5. Observing the same PSs more than once had a stronger effect than observing them only once.

6. The difference between immediate effect and lasting effect is small and not statistically significant.

7. Observation always played a role where fraud was widespread, that is, in regions with sizable Russian minorities.

One point must be highlighted. Observers were able to reduce the rate of fraud in the PSs they visited. However, this article has also shown that previous observation influences future rounds. We can thus tentatively conclude that observation has many different unintended consequences, among them that it modifies the behavior also of PS officials who were not directly observed, but which simply understood the potential of observation and autonomously decided to reduce fraud. The findings of this article support the strain of literature claiming that international actors may play a role in democratization process. I am not arguing that EOs changed the political future of Ukraine (actually, Yushchenko's presidency failed to reform state institutions, to establish a clear system of checks and balances, and to strengthen the rule of law [Kudelia, 2012]), but, more circumspectly, that they contributed, along with other domestic and international factors, to model the (failed?) Ukrainian transition. Future research may shed light on the differences, only mentioned in this article, among observation during opening or polling and observation during counting; or on the role of domestic observer in comparison with international ones; or, again, on the possible incentives EOs offer to undemocratic leaders to change fraudulent methods in order not to be caught. It is not methodologically possible to extend the findings of this article outside the case considered. Only adding new cases can cumulate knowledge and confirm (or disconfirm) our findings. More importantly, the future of the country shows that reducing fraud at election day is not a sufficient condition of a deterministic process through democratic transition. It could be a necessary one, but more studies are needed even in this direction. ∎

Corresponding author email: marta.regalia@unimi.it

REFERENCES

Alvarez, M. R., Hall, T. E. & Hyde, S. D., eds. (2008) *Election fraud: Detecting and deterring electoral manipulation*. Washington, DC, Brookings Institution Press.

Bader, M. (2011) The challenges of OSCE electoral assistance in the former Soviet Union. *Security and Human Rights*. 22 (1), 9–18.

Bader, M. & Schmeets, H. (2014) Is international election observation credible? Evidence from Organization for Security and Co-operation in Europe missions. *Research & Politics*. 1 (2), 1–6.

Beaulieu, E. & Hyde, S. D. (2008) In the shadow of democracy promotion: Strategic manipulation, international observers, and election boycotts. *Comparative Political Studies*. 42 (3), 392–415.

Bjornlund, E. (2004) *Beyond free and fair: Monitoring elections and building democracy*. Washington, DC, Woodrow Wilson Center Press with Johns Hopkins University Press.

Bjornlund, E., Bratton, M. & Gibson, C. (1992) Observing multiparty elections in Africa: Lessons from Zambia. *African Affairs*. 91 (364), 405–431.

Buzin, A., Brondum, K. & Robertson, G. (2016) Election observer effects: A field experiment in the Russian Duma election of 2011. *Electoral Studies*. 44, 184–191.

Carothers, T. J. F. (1997) The rise of election monitoring: The observers observed. *Journal of Democracy*. 8 (3), 17–31.

Darnolf, S. (2011) International election support: Helping or hindering democratic elections? *Representation*. 47 (4), 361–382.

Daxecker, U. E. (2012) The cost of exposing cheating: International election monitoring, fraud, and post-election violence in Africa. *Journal of Peace Research*. 49 (4), 503–516.

Donno, D. (2010) Who is punished? Regional intergovernmental organizations and the enforcement of democratic norms. *International Organization*. 64 (4), 593–625.

Donno, D. (2013) *Defending democratic norms: International actors and the politics of electoral misconduct*. Oxford, Oxford University Press.

Enikolopov, R., Korovkin, V., Petrova, M., Sonin, K. & Zakharov, A. (2013) Field experiment estimate of electoral fraud in Russian parliamentary elections. *Proceedings of the National Academy of Sciences*. 110 (2), 448.

Foroughia, P. & Mukhtorova, U. (2017) Helsinki's counterintuitive effect? OSCE/ODIHR's election observation missions and solidification of virtual democracy in post-communist Central Asia: The case of Tajikistan, 2000–2013. *Central Asian Survey*. 36 (3), 373–390.

Fortin-Rittberger, J., Harfst, P. & Dingler, S. C. (2017) The costs of electoral fraud: Establishing the link between electoral integrity, winning an election, and satisfaction with democracy. *Journal of Elections, Public Opinion and Parties*. 27 (3), 350–368.

Garber, L. & Cowan, G. (1993) The virtues of parallel cota tabulation. *Journal of Democracy*. 4 (2), 95–107.

Hafner-Burton, E. M., Hyde, S. D. & Jablonski, R. S. (2014) When do governments resort to election violence? *British Journal of Political Science*. 44 (1), 149–179.

Hyde, S. D. (2007) The observer effect in international politics: Evidence from a natural experiment. *World Politics*. 60 (1), 37–63.

Hyde, S. D. (2010) Experimenting in democracy promotion: International observers and the 2004 presidential elections in Indonesia. *Perspectives on Politics*. 8 (2), 511–527.

Hyde, S. D. (2011a) Catch us if you can: Election monitoring and international norm diffusion. *American Journal of Political Science*. 55 (2), 356–369.

Hyde, S. D. (2011b) *The pseudo-democrat's dilemma: Why election monitoring became an international norm*. Ithaca, NY, Cornell University Press.

Hyde, S. D. & Marinov, N. (2012) Which elections can be lost? *Political Analysis*. 20 (2), 191–210.

Ichino, N. & Schündeln, M. (2012) Deterring or displacing electoral irregularities? Spillover effects of observers in a randomized field experiment in Ghana. *Journal of Politics*. 74 (1), 292–307.

Kelley, J. (2008) Assessing the complex evolution of norms: The rise of international election monitoring. *International Organization*. 62 (2), 221–255.

Kelley, J. (2009) D-Minus elections: The politics and norms of international election observation. *International Organization*. 63 (4), 765–787.

Kelley, J. (2010) Election observers and their biases. *Journal of Democracy*. 21 (3), 158–172.

Kelley, J. (2012a) *The good, the bad, and the ugly: Rethinking election monitoring*. Stockholm, International IDEA.

Kelley, J. (2012b) *Monitoring democracy: When international election observation works and why it often fails*. Princeton, NJ, Princeton University Press.

Kubicek, P. (2005) The European Union and democratization in Ukraine. *Communist and Post-Communist Studies*. 38 (2), 269–292.

Kudelia, S. (2012) The sources of continuity and change of Ukraine's incomplete state. *Communist and Post-Communist Studies*. 45 (3–4), 417–428.

Makhorkina, A. (2005) Ukrainian political parties and foreign policy in election campaigns: Parliamentary elections of 1998 and 2002. *Communist and Post-Communist Studies*. 38 (2), 251–267.

McAllister, I. & White, S. (2015) Electoral integrity and support for democracy in Belarus, Russia, and Ukraine. *Journal of Elections, Public Opinion and Parties*. 25 (1), 78–96.

McCoy, J., Garber, L. & Pastor, P. (1991) Pollwatching and peacemaking. *Journal of Democracy*. 2 (4), 102–114.

Norris, P. (2014) *Why electoral integrity matters*. Cambridge, Cambridge University Press.

OSCE/ODIHR. (2004) *Ukraine Presidential elections 31 October, 21 November and 26 December 2004: OSCE/ODIHR Election Observation Mission Final Report*.

Regalia, M. (2011) Working for democracy: The effectiveness of election observation. In: Schmeets, H. (ed.) *International election observation and assessment of elections*. The Hague/Heerlen, Statistic Netherlands and Maastricht University Press, pp. 212–236.

Schedler, A. (2002) Elections without democracy: The menu of manipulation. *Journal of Democracy*. 13 (2), 36–50.

Simpser, A. and Donno, D. (2012) Can international election monitoring harm governance? *Journal of Politics*. 74 (2), 501–513.

Smidt, H. (2016) From a perpetrator's perspective: International election observers and post-electoral violence. *Journal of Peace Research*. 53 (2), 226–241.

von Borzyskowski, I. (2019) The risks of election observation: International condemnation and post-election violence. *International Studies Quarterly*. 63 (3), 654–667.

Wilson, A. (2005) *Virtual politics: Faking democracy in the post-Soviet world*. New Haven, CT, Yale University Press.

Žielys, P. & Rudinskaitė, R. (2014) US democracy assistance programs in Ukraine after the Orange Revolution. *Communist and Post-Communist Studies*. 47 (1), 81–91.