

Replication of “Electoral Administration in Fledgling Democracies: Experimental Evidence from Kenya”

Erin Guetzloe

05/14/2020

Abstract

Harris, Kamindo, and Windt (2020a) find that local registration visits by election staff to Kenyan polling sites improve voter registration, while civic education and SMS messages reminding voters to register were less effective at increasing registration. I successfully replicated all of the major findings from Harris, Kamindo, and Windt (2020a). My extension focused on using Bayesian modeling and posterior distributions to reproduce the frequentist models employed in the original paper, which showed that poverty, distance of registration offices from polling stations, and population sparseness were all negatively correlated with higher levels of voter registration. The results of the frequentist and Bayesian models were almost identical, providing corroboration to Harris et al.’s claim that attempts to increase levels of voter registration in Kenya must find ways to better interact with areas that are poorer, that have polling stations distant from registration offices, and that are more isolated.

Introduction

In this experiment, Harris, Kamindo, and Windt (2020a) found that local registration visits to polling stations where citizens were registered by election staff with mobile registration devices was the most effective intervention at increasing levels of voter registration out of three interventions designed. The two other interventions, civic education and SMS reminders

to ask friends and family to register, had near-zero effects on voter registration. However, when combined with local registration visits by election staff, both civic education and SMS reminders increased levels of registration. Importantly, the researchers also discovered that poverty, distance of polling stations from registration offices, and population sparseness are all negatively correlated with the probability of registering to vote.

I was able to successfully replicate all of the results from Harris, Kamindo, and Windt (2020a), who graciously made all their data and replication code accessible within the Journal of Politics dataverse. Both their original work and my replication were completed in R. All of the code I used to complete my extension to the paper is available on my Github.¹

For my extension, I sought to use Bayesian modeling to recreate the regressions Harris, Kamindo, and Windt (2020a) used to demonstrate that as poverty, distance of election constituency offices from polling stations, and population sparseness increase, levels of voter registration decrease. To do this, I created a set of logistic regressions similar to those used by Harris, Kamindo, and Windt (2020a), but instead of using the `lm` function as Harris, Kamindo, and Windt (2020a) did, I employed the `stan_glm` function from the `rstanarm` package. After making the new models, I compared them with the original results and found the differences to be negligible, confirming the claims made by Harris, Kamindo, and Windt (2020a). I then estimated the mean predicted probabilities of at least one voter registering at a polling station based on varying levels of poverty, distance, and population density. Graphing these results as histograms again corroborated Harris et al.'s argument—all of these variables impact levels of voter registration and must be considered when Kenyan policymakers seek to increase registration.

Literature Review

While there are many studies which attempt to examine the impact of civic education on various measures of political participation such as voter turnout or political knowledge, this paper is unique among much of the current literature regarding civic education and

¹<https://github.com/eguetzloe/electoral-administration-kenya>

political participation because it examines the impact of civic education along with two other interventions on voter registration specifically. There are only 2 other major academic papers which explore how civic education affects voter registration. In a paper entitled “The Effects of Civic Education on Political Culture: Evidence from Zambia”, Bratton et al. (2020) look at the impact of canvassing and school-based curriculum approaches to civic education on voter registration. Within the paper “Promoting Democracy in Fragile States: Field Experimental Evidence from Liberia”, Mvukiyehe and Samii (2017) examines how townhall meetings can affect registration. However, neither of these papers uses the same approach to civic education as Harris, Kamindo, and Windt (2020a). Most studies design civic education programs targeting students in primary or secondary school, not adults: by implementing new “curriculum or the introduction of civics exams”, as Harris, Kamindo, and Windt (2020a) write. In contrast, Harris, Kamindo, and Windt (2020a) uses face-to-face canvassing as its form of civic education. Additionally, many papers focus exclusively on civic education without exploring alternate mechanisms for increasing voter registration. Harris, Kamindo, and Windt (2020a) offers new solutions, showing that “the largest benefits to voter registration may lie in changing convenience and cost, not providing civic education.”

Harris et al.’s research also differs from much of the other relevant literature in the type of data and experiment used. Harris, Kamindo, and Windt (2020a) employs administrative data provided directly by the Kenyan government, working in collaboration with the Kenyan electoral commission to design a true experiment. Much of the other research published about civic education relies on self-reported data about political knowledge or participation for measuring outcomes, rather than using actual data about levels of registration. While Mvukiyehe et al.’s paper does utilize an experimental design as well, Bratton et al. (2020) instead used an observational study.

Finally, neither Bratton et al. (2020) nor Mvukiyehe and Samii (2017) use blocking subgroups to explore alternative variables impacting voter registration. Harris et al.’s decision to include blocking based on poverty, distance, and population density stands out in comparison to other papers that investigate civic education and voter registration. Thus, both the methods of civic education employed in Harris et al.’s research and the design of the experiment itself

place it in a unique place within the current literature.

Paper Review

The experiment documented in this paper was the result of a collaboration between a team of independent researchers and the Kenyan government’s electoral commission, the Independent Electoral and Boundaries Commission (IEBC). Both the researchers and the Kenyan government were interested in testing the best mechanisms for increasing levels of voter registration specifically within younger democracies. The researchers developed three interventions which theoretically could increase voter registration: localization, which involved electoral commission staff being sent out with mobile registration equipment to register citizens to vote, canvassing, which used face-to-face meetings of around ten to thirty minutes each to describe the benefits of registration to citizens, and SMS reminders, which sent text messages to citizens already registered to vote asking them to tell their friends and family about the opportunities to register. They selected seven Kenyan counties for this experiment, which were chosen based on their similarity in levels of poverty, distance from registration offices, population sparseness, and levels of governmental support in the previous election cycle. The selected Kenyan counties contained 3,828 polling stations and out of these stations, 1,674 stations were randomly selected to participate in the experiment. Each polling station was randomly assigned to receive one of six possible treatments: localization only, localization with canvassing, localization with SMS reminders, canvassing alone, SMS reminders alone, or no intervention (these polling stations receiving no intervention served as the control group).

The researchers also realized that other variables might contribute to levels of voter registration at these polling sites, so each polling station received a standardized score indicating levels of poverty, distance of registration offices from polling stations, and population density around the polling station. Poverty was measured as the percentage of the population who could register at a polling station who were below the poverty line, average across the county. Distance from polling stations to registration offices was measured as an indica-

tor which took into account kilometer distance as the bird flies, kilometer walking distance according to Google Maps, and kilometer driving distance according to Google Maps. Population density was defined as the population density within a 500 meter radius of a polling station. Data for poverty and population density was drawn from worldpop.org. Using this data in combination with information from the IEBC about which polling stations had at least one individual register there from the continuous voter registration (CVR) period between March 16 to November 14, 2016, the researchers found that as levels of poverty, distances between polling stations and registration offices, and levels of population sparseness increase, levels of voter registration decrease. This evidence reveals something important to understand for future researchers and policymakers seeking to increase voter registration in Kenya- the current system must find ways to better engage voters from more impoverished, remote, and sparse areas.

Replication

After running all of the models from Harris et al.'s replication document, I found my results to be consistent with those presented in the original paper. All of the code required to replicate these results is available within my Appendix. The authors' code can be found at the *The Journal of Politics* Dataverse,² and all of the code I use for this extension is available within my Github repository.³

Extension

While the models run in Harris et al.'s paper are extensive, I decided that I would build on the work of these researchers by applying a Bayesian framework to three of the models created within the original paper. I was especially interested in the authors' findings that poverty, distance of registration offices from polling stations, and population density all impact levels of voter registration. Harris et al. chose to utilize frequentist modeling to

²<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UT25HQ>

³<https://github.com/eguetzloe/electoral-administration-kenya>

investigate this correlation in their original paper. Thus, to explore this relationship further, I created Bayesian models using the `stan_glm` function from the `rstanarm` package. The results of each original regression, explaining voter registration levels by poverty, distance, and sparseness, are compared with the new Bayesian models I created in the table below.

Table 1: Linear and Bayesian Model Comparison

| | Model 1 | New Model 1 | Model 2 | New Model 2 | Model 3 | New Model 3 |
|--------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Intercept | 0.454 (0.021) | 0.454 (0.022) | 0.462 (0.021) | 0.462 (0.021) | 0.469 (0.022) | 0.470 (0.022) |
| Poverty | -0.152 (0.021) | -0.152 (0.021) | | | | |
| Distance | | | -0.262 (0.022) | -0.262 (0.023) | | |
| Sparseness | | | | | -0.306 (0.064) | -0.306 (0.063) |
| Observations | 3823 | 3823 | 3823 | 3823 | 3823 | 3823 |

This table compares the differences in coefficients found from running the frequentist models in the original paper against the Bayesian models created with `stan_glm`. Though there are a few differences between the coefficients produced by new models and the original, they are largely insignificant, confirming Harris et al.’s claim that as poverty, distance of registration offices from polling stations, and population sparseness increase, levels of voter registration decrease. In interpreting the results presented by this table, it is important to remember that the measures of variation listed under each coefficient in the table are not the same. While the frequentist linear models created in Harris, Kamindo, and Windt (2020a) signify the estimated standard errors for each regression coefficient, the Bayesian models that I designed calculate the standard deviation of the median absolute deviation (MAD SD), which takes the median absolute deviation and uses a set scaling factor to make the values interpretable as estimates of standard error. As illustrated within the table, the differences between these values are ultimately negligible, but calculating MAD SD is considered a more “robust measure of dispersion” according to Makowski (n.d.).

After making the new Bayesian models, I wanted to use these models to create predicted probabilities of voter registration at polling stations with varying levels of poverty, distance from registration offices, and population density. For each of these explanatory variables

within the original dataset, I pulled the lowest value, the 25th percentile value, the median value, the 75th percentile value, and the highest value. The highest values represent the highest levels of poverty, distance of registration offices from polling stations, and population sparseness. Next, I placed each of these individual polling stations within “new datasets” which I used to extract posterior draws of the linear predictor transformed into a probabilistic 0-1 scale through the inverse-link function. To do this, I utilized the `posterior_linpred` function, again from the `rstanarm` package. I placed the results of these draws into histograms which can help to visualize the mean predicted probabilities of at least one individual registering to vote at polling stations with differing levels of poverty, distance, and sparseness.

Figure 1: Predicted Probabilities of Registration Based On Poverty

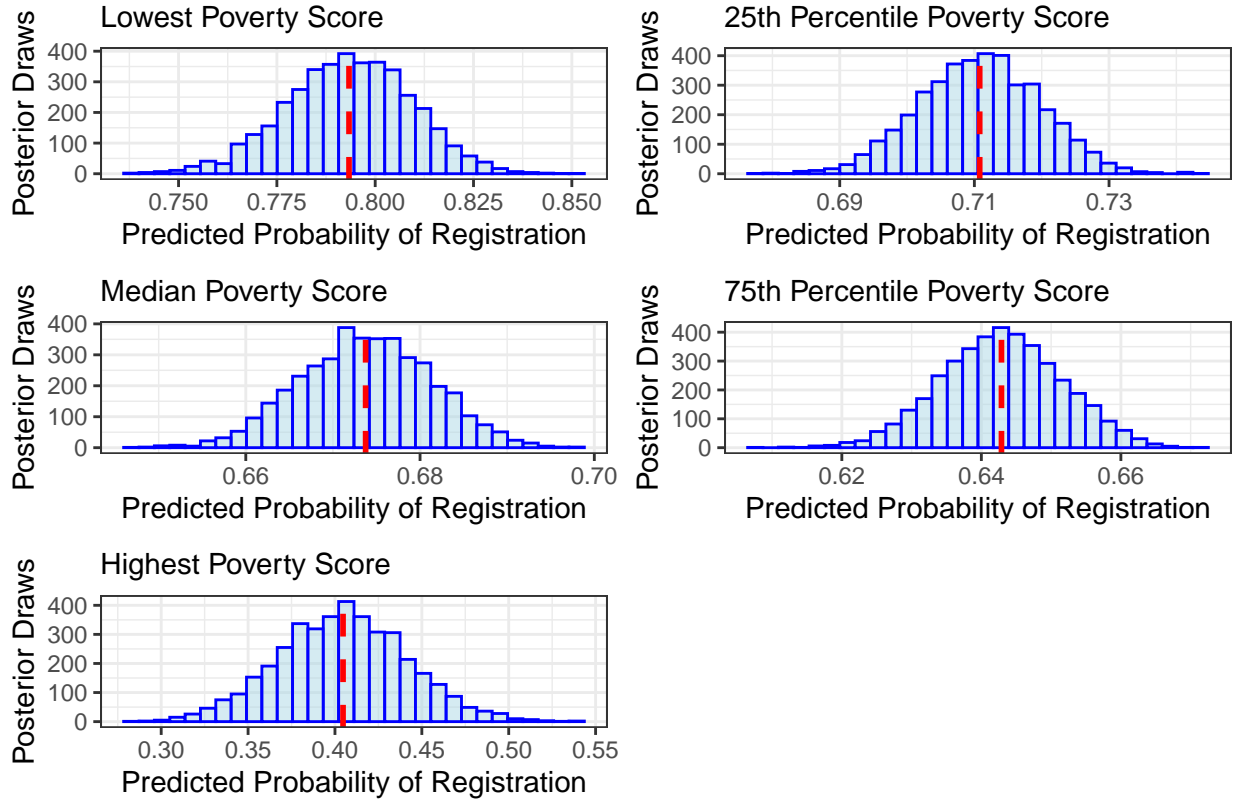
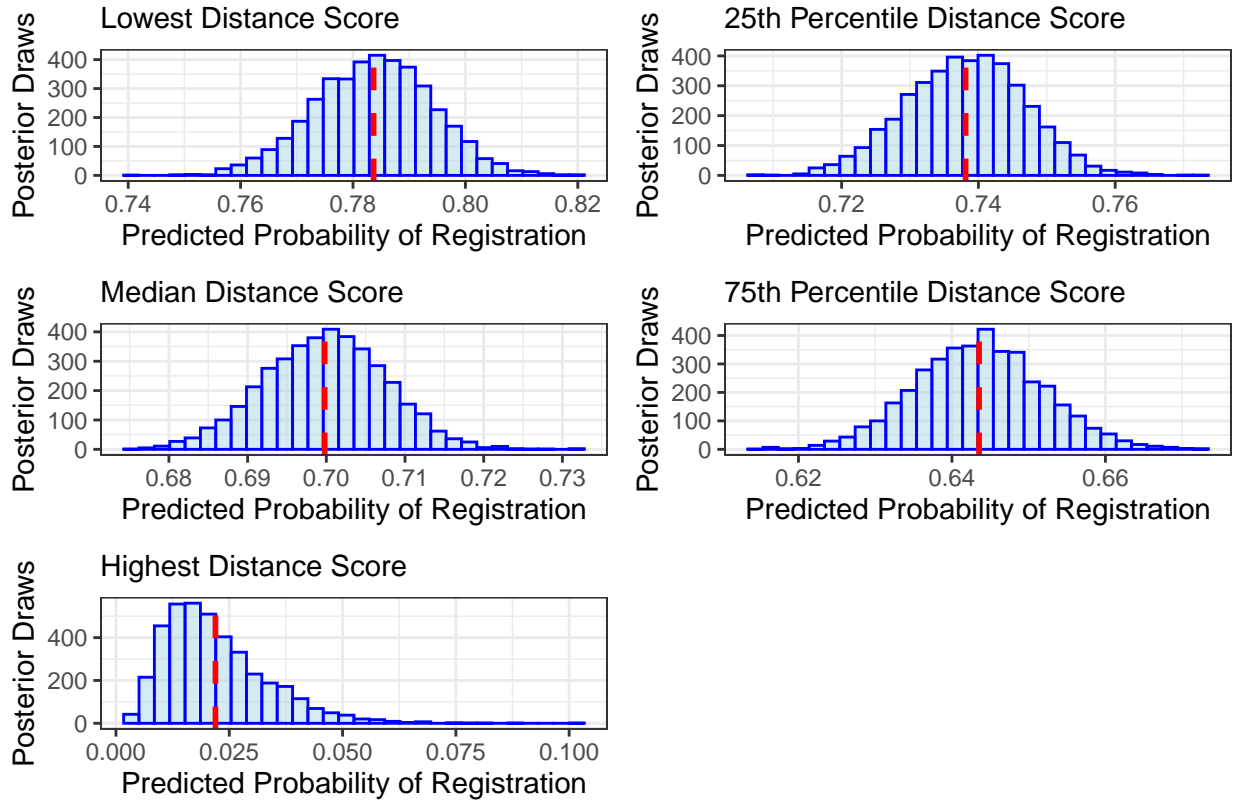


Figure 1, printed above, provides visualizations of polling stations at the lowest poverty level (thus, the wealthiest polling station), the 25th percentile poverty level, the median poverty level, the 75th percentile poverty level, and the highest poverty level (the poorest polling station). As expected from the results of the models created earlier, the histograms show a

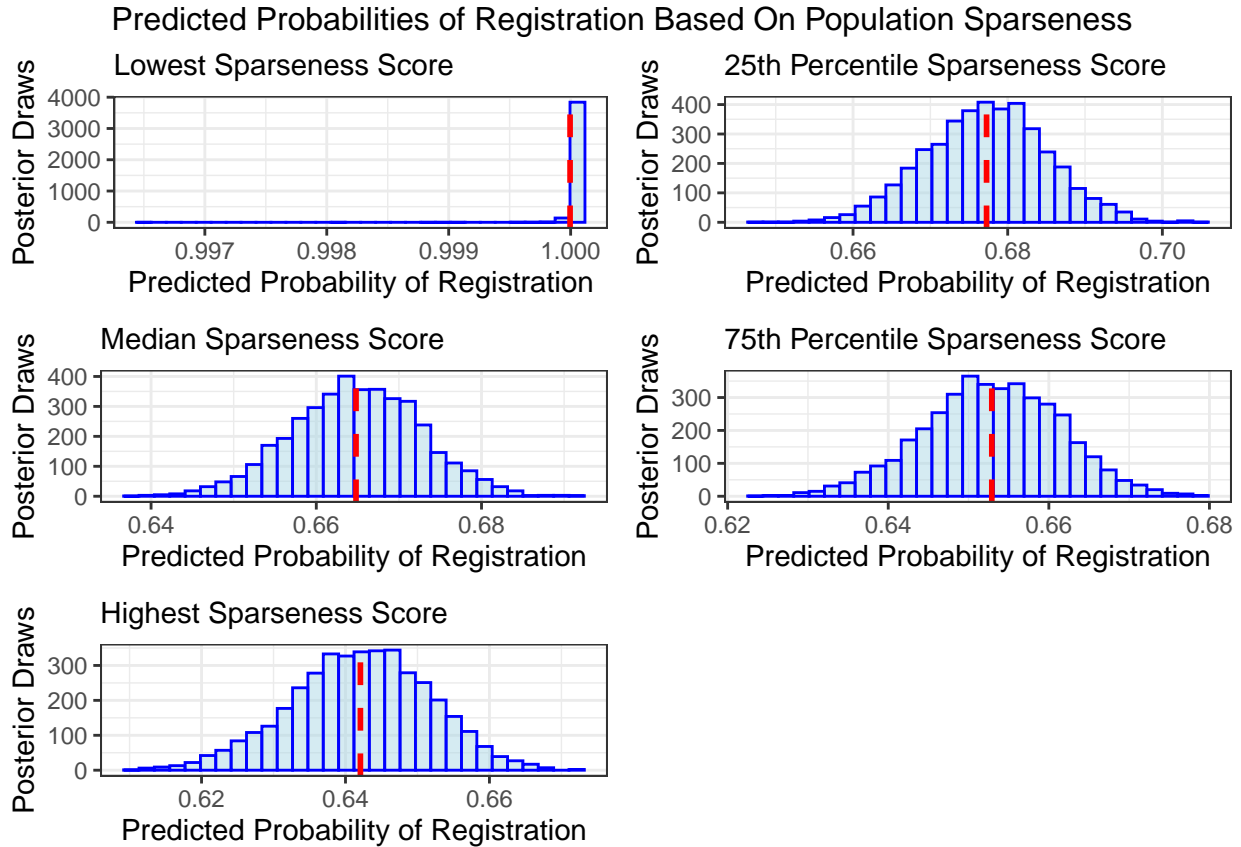
general trend of predicted probability of registration decreasing as poverty scores increase. The red dashed lines indicate the mean predicted probability of at least one individual registering to vote at a certain polling station. For example, the mean predicted probability of at least a single citizen registering at the wealthiest polling station is approximately 0.79, while the mean predicted probability of at least one voter registering at the poorest polling station is about 0.40. This is roughly a 50% decrease in mean predicted probability, indicating that poverty does indeed have a strong impact on voter registration in Kenya.

Figure 2: Predicted Probabilities of Registration Based On Distance



The next variable I visualized in Figure 2 was distance of registration offices from polling stations. In Kenya, voters must first go to a constituency election office to register to vote, while actual voting occurs at polling stations. Thus, Harris, Kamindo, and Windt (2020a) was interested in discovering whether longer distances between the election offices where voters register and the polling stations where voting happens have an impact on levels of registration. Both the results from the authors' original models and my new Bayesian models confirm that longer distances between registration offices and polling stations are negatively

correlated with registration. The mean predicted probability of a voter registering for the polling station with the least distance from its registration office is estimated to be about 0.78, while the mean predicted probability of a single voter registering for the polling station farthest from its registration office is only 0.02. While this predicted probability may seem very low, it seems that this value may be somewhat of an outlier within the dataset since the polling station with the 75th percentile distance score has a mean predicted probability estimated at 0.64. Additionally, the distribution of draws for the highest distance score is heavily skewed to the right, meaning that probabilities as high as 0.10 were estimated within the matrix of simulations. However, even considering the potential outlier value, these histograms confirm that as the distance between election constituency offices and polling stations increase, the predicted probability of at least one voter registering at that polling station decreases.



The final variable I explored was population density. As with the highest distance score in the previous set of histograms, it appears that the polling station with the lowest population

sparseness (thus, the polling station with the most concentrated population) may be an outlier value. Out of 4000 posterior draws, the predicted probabilities for registration at this polling station only extend from approximately 0.99 to 1.00. The mean predicted probability of registration for the polling station with the 25th percentile density score is close to 0.68, while the mean predicted probability of a voter registering at the most isolated polling station is about 0.64. Interestingly, the values produced here indicate that there is not a great difference in predicted probability of registration between a polling station in a more densely populated region and a polling station in the least densely populated region. Though the negative correlation between population sparseness and voter registration indicated in the models is still present, it appears from a closer look at the data that the coefficient for density, which was the largest of all the coefficients shown in Table 1, was being affected one or more outlier polling stations with very high levels of population density. This could mean that population sparseness is not as strongly influential on voter registration as it initially appears.

Conclusion

In “Electoral Administration in Fledgling Democracies: Experimental Evidence from Kenya,” Harris, Kamindo, and Windt (2020a) explores potential mechanisms for increasing levels of voter registration within younger democratic states. The researchers worked with the Kenyan electoral commission to specifically explore three different interventions designed to increase voter registration: first, localization, where electoral staff with mobile registration devices visit polling stations to reduce the time needed to register, second, canvassing, where electoral staff paid visits to citizens’ homes to discuss with them the importance of registering to vote, and third, SMS reminders, where individuals registered to vote receive text messages asking them to remind their friends and family to register as well. Harris, Kamindo, and Windt (2020a) found that localization was the most effective means of increasing registration. They also discovered that poverty, distance of registration offices from polling stations, and population sparseness are all correlated with lower levels of voter registration.

I was successful in my attempt to replicate all of Harris et al.'s results. I used the data made publicly available through Harris, Kamindo, and Windt (2020b). All of the authors' original code and the code I used for this extension was completed in R, and it can be found at my Github repository for this project.⁴

My extension used Bayesian rather than frequentist modeling to investigate the relationship that poverty, distance from registration offices, and population sparseness have on voter registration. The results produced from the Bayesian models were almost identical to the results of Harris et al.'s frequentist models. Additionally, I used a set of histograms to visualize the mean predicted probabilities of at least one individual registering at polling stations with the lowest, 25th percentile, median, 75th percentile, and highest levels of each explanatory variable. These graphs illustrate the same trend central to Harris et al.'s original paper: poverty, distance, and isolation are correlated with lower levels of voter registration, and thus any efforts to increase voter registration must actively seek to better respond to the needs of polling stations in impoverished regions, regions where polling stations are far from registration offices, and regions with isolated populations.

This finding has important implications for any actors attempting to increase voter registration in any nation. While many interventions intended to increase registration focus solely on providing citizens with civic education, it is possible that the focus of such interventions should not be primarily on increasing political knowledge but instead should be on eliminating barriers that discourage people from registering, such as cost, distance, and long waiting periods. This provides a compelling explanation for why localization was the most effective intervention used in this experiment: because localization actively reduced cost, time, and distance by allowing individuals to register on-site at polling stations. Researchers interested in building on this body of work ought to consider designing other interventions that could decrease poverty, distance, and sparseness-related barriers to voting. Doing so could be the key to increasing voter registration in young democracies and could provide a model for how to promote greater political participation across the world.

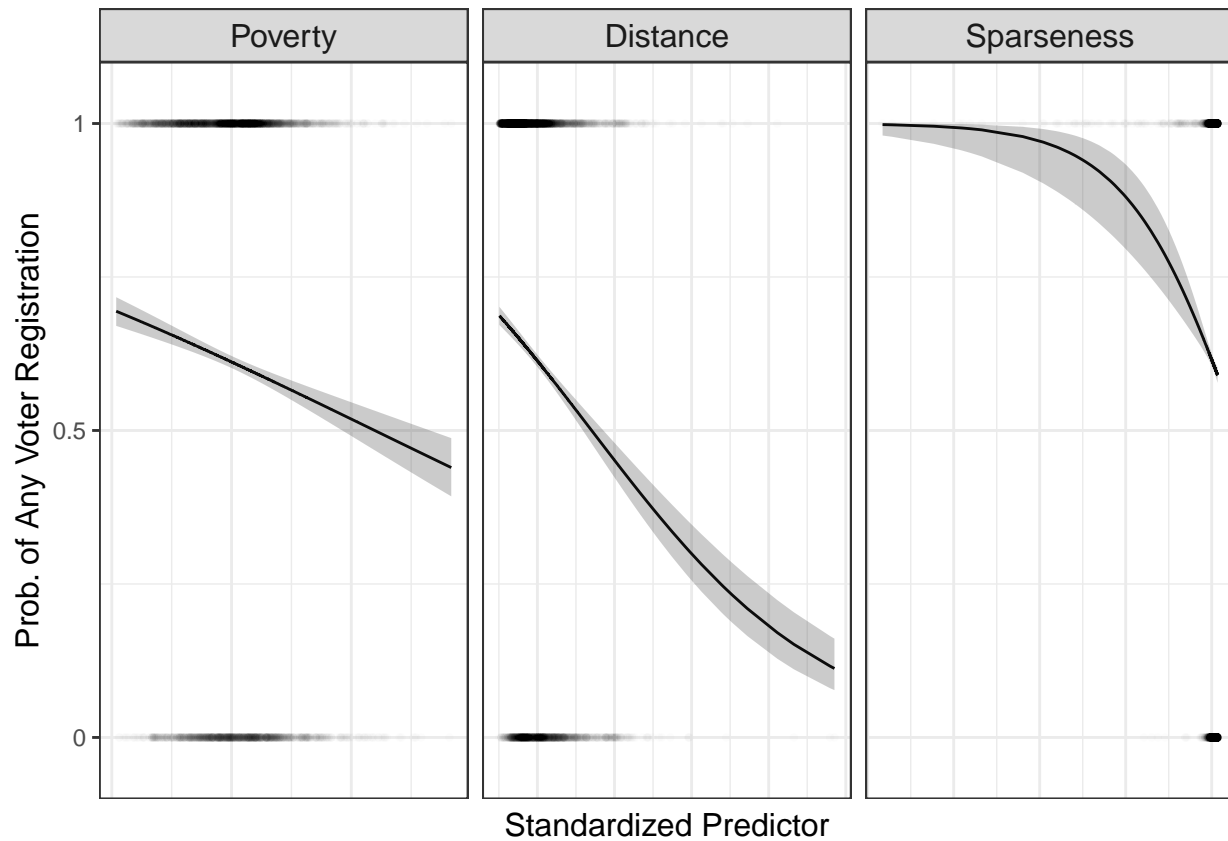
⁴<https://github.com/eguetzloe/electoral-administration-kenya>

References

- Bratton, Michael, Philip Alderfer, Georgia Bowser, and Joseph Temba. 2020. “The Effects of Civic Education on Political Culture: Evidence from Zambia.” <https://www.sciencedirect.com/science/article/abs/pii/S0305750X99000315>.
- Harris, J. Andrew, Catherine Kamindo, and Peter van der Windt. 2020a. “Electoral Administration in Fledgling Democracies: Experimental Evidence from Kenya.” <https://ssrn.com/abstract=3520421>.
- . 2020b. *Replication Data for: Electoral Administration in Fledgling Democracies: Experimental Evidence from Kenya*. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UT25HQ>.
- Makowski, Dominique. n.d. *Analyze Stanreg Objects*. <https://www.rdocumentation.org/packages/psycho/versions/0.4.91/topics/analyze.stanreg>.
- Mvukiyehe, Eric, and Cyrus Samii. 2017. “Promoting Democracy in Fragile States: Field Experimental Evidence from Liberia.” <https://ideas.repec.org/a/eee/wdevel/v95y2017icp254-267.html>.

Appendix

I successfully replicated all the results of Harris, Kamindo, and Windt (2020a). The replication code for the original paper is provided directly by the authors in *The Journal of Politics* Dataverse,⁵ and can also be found in my Github repository for this replication project.⁶ Here is an example of one of the figures I replicated from Harris, Kamindo, and Windt (2020a), which became the basis of my extension in this paper.



⁵<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UT25HQ>

⁶<https://github.com/eguetzloe/electoral-administration-kenya>