Vote Brokers Replication Paper

Cristopher Patvakanian

Abstract

Frye, Reuter and Szakony (2019) examine the voter behavior in Russia and Venezuela and find different types of brokers, appeals, and targets have different effects on voter turnout. I successfully replicated all of their results. As a robustness test, I impute missing values in the dataset and find results in line with that of the original study, but of a slightly different magnitude. These results confirm the authors' original findings and suggest that the missing values in their sampled population do not bias the results.

Contents

1 Introduction

I am replicating "Vote Brokers, Clientelist Appeals, and Voter Turnout: Evidence from Russia and Venezuela" by Timothy Frye, Ora John Reuter and David Szakonyi.[1] The authors use survey-based framework experiments conducted in Russia and Venezuela to analyze voter patterns and turnout in both countries respsectively. Their main finding is that there is higher voter turnout after a voter has been induced by an employer as opposed to a party activist. The results are consistent in both Venezuela and Russia, and are also consistent with existing literature on voter turnout, such as Mares and Young (2016), who also find that employer intimidation significantly influences voter turnout. The data are rich and from face to face interviews in both Russia and Venezuela, with the Venezuelan data coming from a stratified sample whereas the Russian from 20 regions. Thus, it appears that the Venezulean data is a more representative than the Russian. However I believe due to the random sampling and oversampling,the Russian data is still an accurate representation of Russian voters. The models used in the paper include using difference in means between brokers and between inducements, predicted probabilities from individual ordered logit regressions on respondents' likelihood of voting in the survey experiment, and fixed effect linear regressions using interaction variables on willingness to vote.

I was able to sucessfully replicate the paper using the software of RStudio Team (2015). The source of the data and code used for this replication was generously provided online at the authors' publicly accesible Dataverse.[2] The repo with my replication code and data can be found on Github at the link provided in the footnote.[3]

For my extension of this paper, I decided to look at the missing values in this dataset. Given that the treatment assignment was random for all participants, it seemed unlikely that the misssing values the data would likely lead to any to any bias. However, as a robustness check, it is still an important contribution to ensure that the results we see are not due to potential sample bias in either country and that the data truly are missing at random. After imputing data using multivariate imputation by chained equations (MICE), I was able to create mutiple imputations or replacement values, for the gaps in the data. After running similar regressions provided in Table 3, I find results consistent with paper's originally findings, with some regressions' coefficeints becoming slightly higher or lower in magitude.

[1] Frye, Reuter, and Szakonyi 2019
[2] Frye, Reuter, and Szakonyi 2019 Dataverse
[3] Author's Github Replication Repository

2 Literature Review

The findings of Frye, Reuter and Szakony (2019) appear to generally follow what is in the literature. In Aslund (2012), Aslund suggests that in many post soviet countries, voting patterns were heavily influenced by the transition to democracy in instiutions built. In Russia, the case was that there was not enough a big push to transform after communism, and thus the country had to face more difficulties in long term in ensuring fair and free elections due to potential voter interference and corruption on all levels. Similarly, Popova (2010) finds that controlling for different employment statuses and corruption, people are stil likely to vote differently, and more corruption generally induces people to vote more.

Regarding the literature on Venezuela, Jimenez, Hidalgo, and Klimek (2017) too find that there is evidence for significant voter interference in voting in the country through "voter rigging". In addition, Jimenez, Hidalgo, and Klimek (2017) draw paralells in this study to the types of voter manipulation rampant in Russia. Stokes et al. (2013) also find that voter fraud is indeed a large problem in Venezuela, and given the extent of information aviable party activists, who for example are able to access a large portion of individual voter history and tendencies, voter intimidiation through party activists has become a significant issue in the country.

Additionally, other areas of literature focuses on the impact of violence on voting outcomes. In the African context, specifically sub-Suharan African nation of Kenya, Bekoe and Burchard (2017) finds that threats to ones life have been used throughout the country to suppress voter turnout, though on aggregate they do not see an effect. Something which is less prevelant in the literature in most contexts, including both Russia and Venezeula, is the influence of employers on voter turnout in these countries. Some of the latest research, presented in Mares and Young (2016), finds that employer intimidation significantly influences voter turnout, which closely matches the findings of Frye, Reuter and Szakony (2019).

3 Replication

I was succesfully able to replicate all of the figures and the regression table provided in the paper based on the authors' provided data and code. There are minor visual differences, but all of the regression and and table values exactly match those provided in Frye, Reuter and Szakony (2019). In the appendix, I provide one example of successful replication of Table 3 from the original paper. The replications of the other key figures and regression tables can be found in my code.

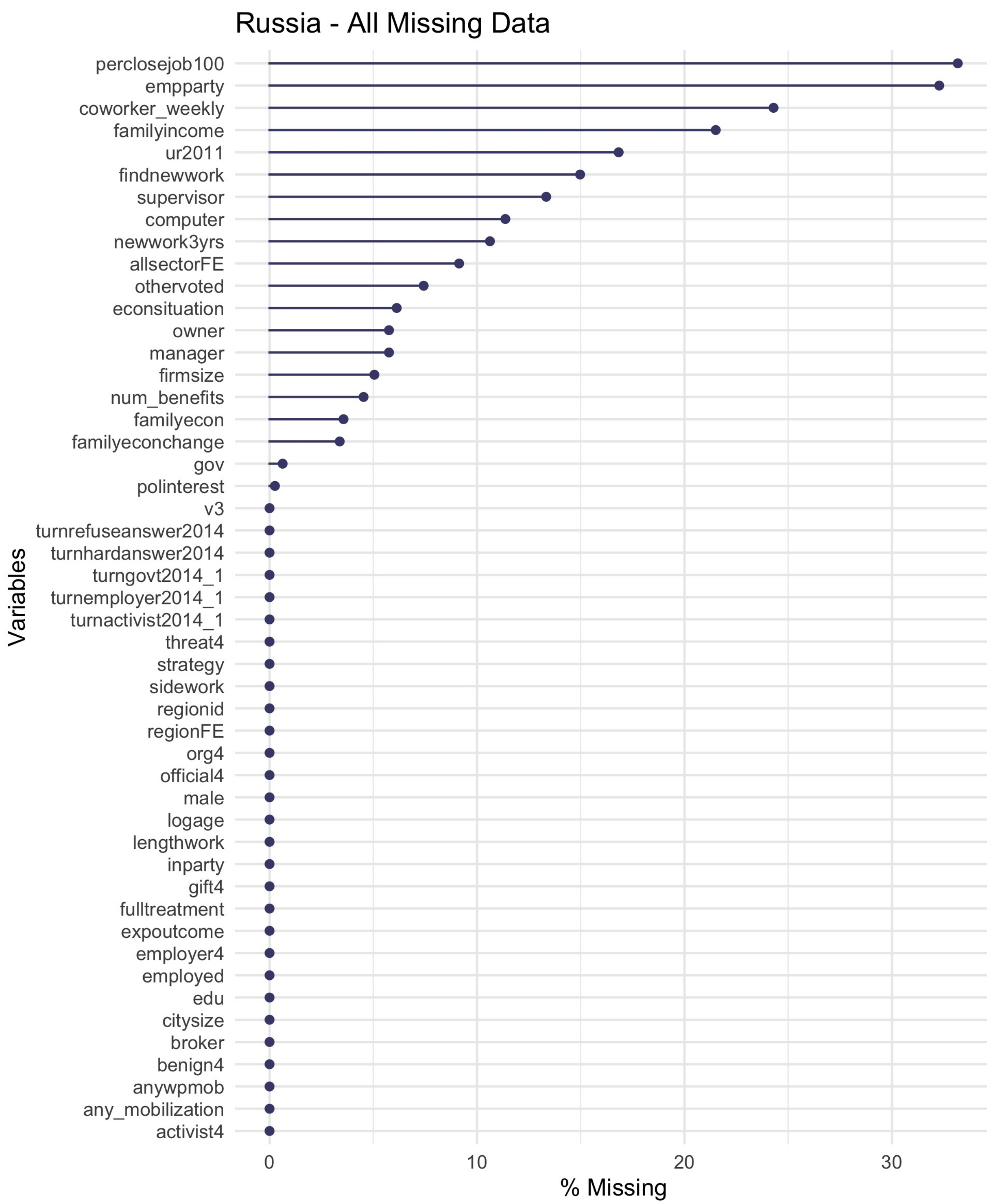
4 Extension

For my extension, I decided to look at missing data in the data sets and use multiple imputition to examine the robustness of the regressions found in Table 3. For my data imputation, I follow the DataScience+ article "Data Management in R Imputing Missing Data with R; MICE package."[4] I choose to use the predictive mean matching method (pmm), which fills in the missing value using a random selection from existing observed "donor values" of other observations in the dataset whose regression-predicted values are closest to the regression-predicted value for the missing observation from the simulated regression model Bruin (2011). Looking at the original data, there do not appear to be too many missing values, at most slighly over 30% for two variables, and only 18 out of the 49 variables in the data have any missing data all. I show the missing data below for the sample as a whole generally, and also for the breakdown per type of broker.

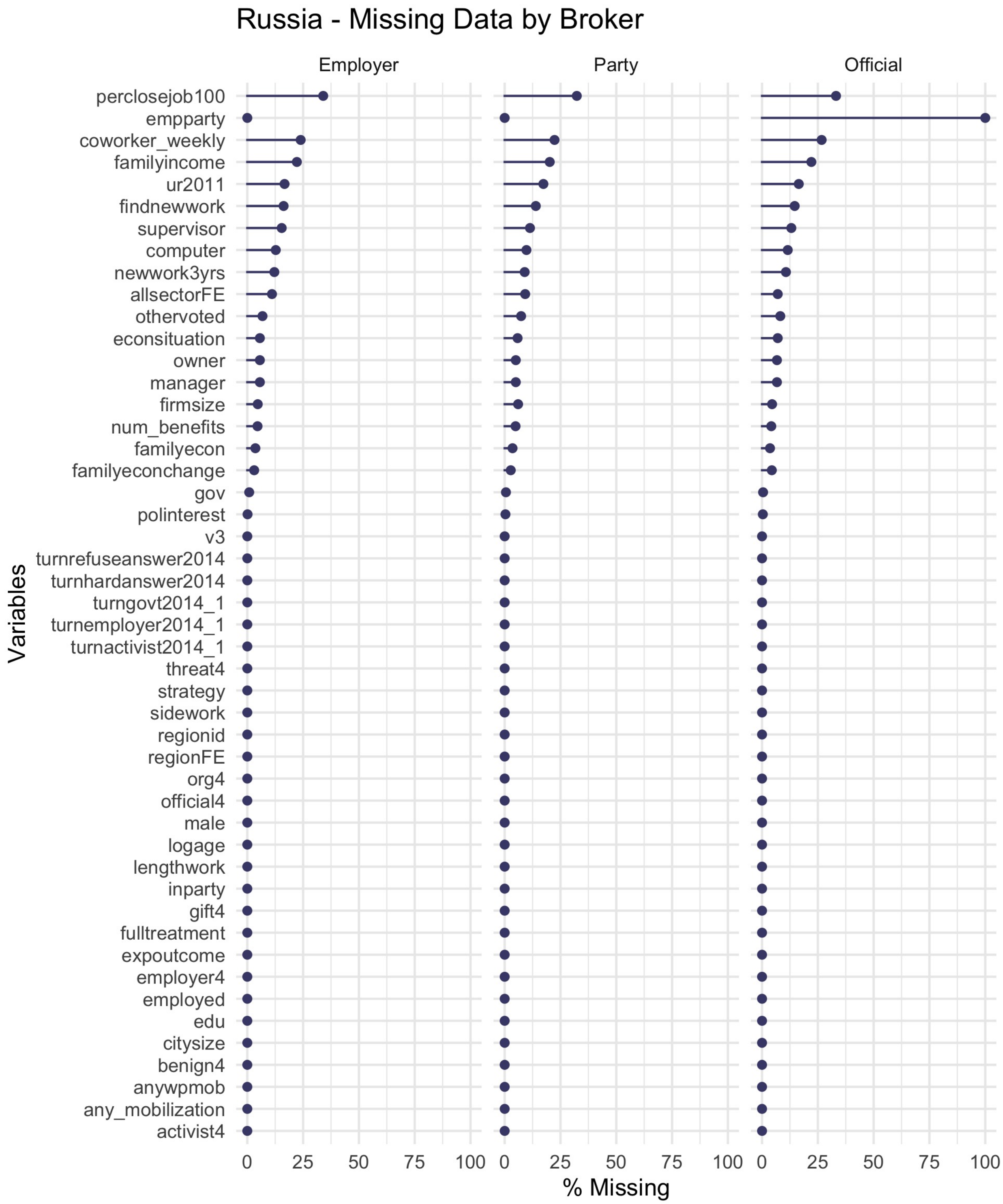
Given that the regression I am using for my extension, Figure 2 (which can be found in the appendix), uses only data from Russia, I decided to impute the missing values for Russia using the mice() function and re-run the regressions.I use the standard 5 iterations to gain 5 data sets of imputed data, and pool the results to get the final coefficeints I would like to use for my model. In Figure 1, I show the density plots for the imputed data as a check to see whether or not the data I have imputed matches the existing values in the data set. It appears that the imputed data generally appear in line with the existing data set, which indicates that the data imputation is valid. Finally in Table 2, which can be found in the appendix, I present the results of re-running the regressions using the newly imputed data.

When I take my results with the imputed data in Table 2 and compare it those of the original study, I find the same significance and sign of the coefficients, however the magnitude of the coefficients appears to differ only very slightly. I think that this just shows the original study is valid and robust, and by being able to not only replicate the data but get very similar results to the original after imputing data, I feel even more confident in the authors' findings.

[4] "Data Management in R Imputing Missing Data with R; MICE package"



Missing data in the provided data for Russia for entire sample.



Missing data in the provided data for Russia broken down by broker type.

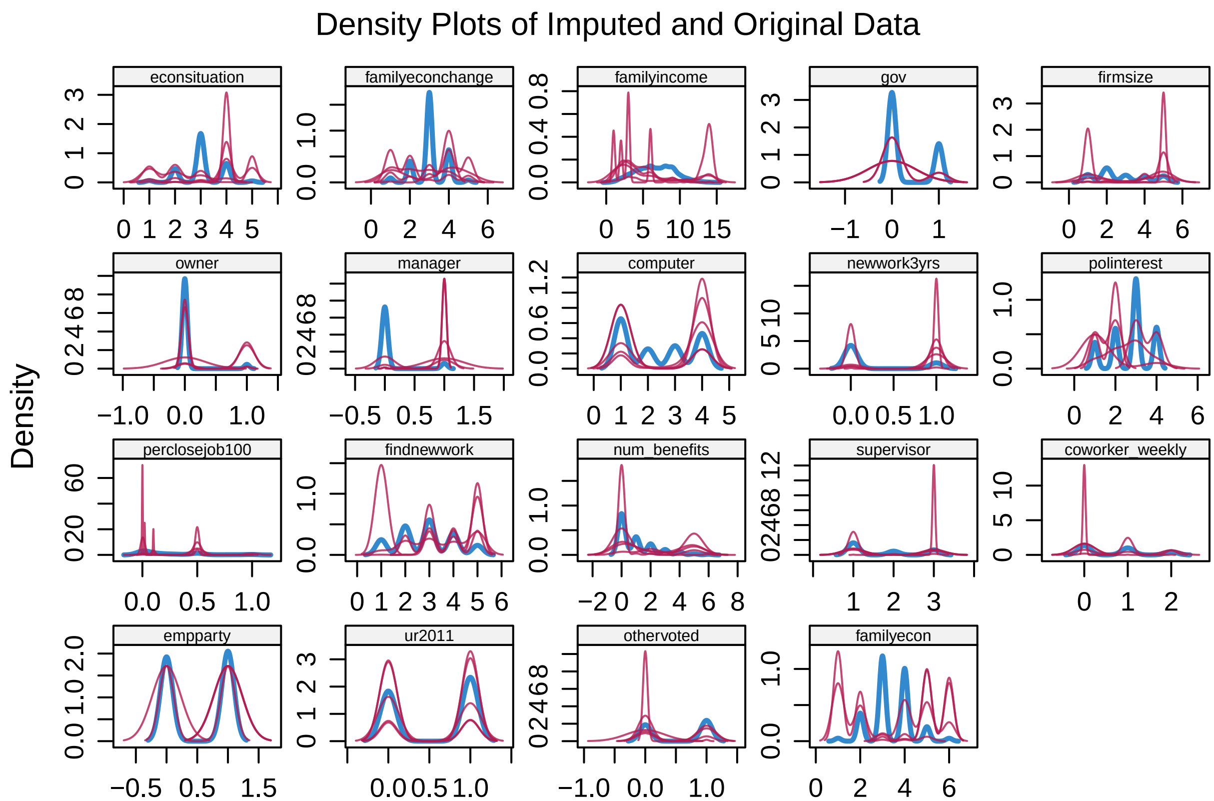


Figure 1: Above are the density plots of the imputed data in crimson overlaid on the density plot for the original sample in blue. The imputed data appears to generally matches the pattern of the original data, and though there are some descrepancies, they do not appear to be extreme. This suggests the data imputation is successful.

5 Conclusion

In this paper, I was succesffuly able to replicate all of the results of Frye, Reuter, and Szakonyi (2019). As a robustness test, I decided to use multivariate imputation by chained equations (MICE) to create mutiple imputations or replacement values for the gaps in the data. After running similar regressions provided in Table 3, I find results consistent with paper's originally findings, with some regressions' coefficeints becoming slightly higher or lower in magitude. These results confirm the authors' original findings and suggest that the missing values in their sampled population do not bias the results. The implications of this replication and extension are that future studies in this area can build on these existing results and have confidence that the findings of Frye, Reuter and Szakony (2019) are in fact robust.

6 References

Aslund, Anders. 2012. *How Capitalism Was Built: The Transformation of Central and Eastern Europe, Russia, the Caucasus, and Central Asia*. 2nd ed. Cambridge University Press. https://doi.org/10.1017/CBO9781139207850.

Bekoe, Dorina A., and Stephanie M. Burchard. 2017. "The Contradictions of Pre-Election Violence: The Effects of Violence on Voter Turnout in Sub-Saharan Africa." *African Studies Review* 60 (2): 73–92. https://doi.org/10.1017/asr.2017.50.

Bruin, J. 2011. "Newtest: Command to Compute New Test @ONLINE." February 2011. https://stats.idre.ucla.edu/stata/ado/analysis/.

Frye, Timothy, Ora John Reuter, and David Szakonyi. 2019. "Vote Brokers, Clientelist Appeals, and Voter Turnout: Evidence from Russia and Venezuela." *World Politics* 71 (4): 710–46. https://doi.org/10.1017/S0043887119000078.

Jimenez, Raúl, Manuel Hidalgo, and Peter Klimek. 2017. "Testing for Voter Rigging in Small Polling Stations." *Science Advances* 3 (6). https://doi.org/10.1126/sciadv.1602363.

Mares, Isabela, and Lauren Young. 2016. "Buying, Expropriating, and Stealing Votes." *Annual Review of Political Science* 19 (1): 267–88. https://doi.org/10.1146/annurev-polisci-060514-120923.

Popova, Olga. 2010. "Corruption, Voting and Employment Status: Evidence from Russian Parliamentary Elections." *SSRN Electronic Journal*.

RStudio Team. 2015. *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, Inc. http://www.rstudio.com/.

Stokes, Susan C., Thad Dunning, Marcelo Nazareno, and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics.* Cambridge Studies in Comparative Politics. Cambridge University Press. https://doi.org/10.1017/CBO9781107324909.

7 Appendix

The results from Frye, Reuter and Szakony (2019) were successfully replicated. As an example, here is Table 3 from page 735. Below I have included my replication, which differs only slightly in terms of table presentation and order of regressions. Those are only minor visual differences which do not take away from the otherwise successful replication of the analysis conducted using the authors' provided code. Further replication of the paper's key graphics and models can be inspected in the RMD file of my code available on the author's Github.[5]

Below I have also included Table 2, which uses the imputed data to replicate the study's original Table 3.

[5] Author's Github Replication Repository

TABLE 3
EXAMINING MECHANISMS: RUSSIA SURVEY EXPERIMENT [a]

	Outcome: Respondent Would Vote						
	Leverage				*Monitoring*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Employer treatment * Chance of job loss	0.360** (0.167)						
Employer treatment * Hard to find a new job		0.156*** (0.038)					
Employer treatment * Receives benefits			0.070** (0.036)				
Employer treatment * Employed in government				0.166** (0.072)			
Employer treatment * Knows supervisor well					0.119** (0.059)		
Employer treatment * Number of years employed						0.012** (0.006)	
Employer treatment * Socializes with coworkers							0.063 (0.076)
Constituent terms	yes	yes	yes	yes	yes	yes	yes
Demographics	yes	yes	yes	yes	yes	yes	yes
Observations	1209	1532	1724	1806	1567	1806	1389

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

[a]The outcome variable is the willingness to turnout outcome (five-point scale) from the survey experiment. The sample includes only respondents who received the employer or political party broker treatment. The employer treatment collapses the data along the inducement treatment arm of the factorial design used in the experiment. The sample is limited to only those who are employed. *Chance of job loss* measures the probability a respondent believes he or she will lose his or her job in the next twelve months. *Hard to find a new job* uses a five-point scale to capture the likelihood that if he or she were to lose his or her job, a respondent could find a similar one; higher values indicate more difficulty. *Receives benefits* captures the number of in-kind benefits (health care, education, transportation subsidies, etc.) respondents received from their employer. Higher values on the three-point scale used in *knows supervisor well* indicate better familiarity with one's boss. *Number of years employed* measures the length of time at one's work. *Socializes with coworkers* captures whether respondents spend time with colleagues outside work. All models include the constituent terms and basic demographic characteristics (gender, age, education, size of settlement, and an indicator for government employment). Models are estimated via OLS and cluster errors at the region level.

Figure 2: Frye, Reuter and Szakony (2019) Table 3

Table 1: Replication of Original Table 3

	Dependent variable:						
	Outcome Respondent Would Vote						
	Leverage					Monitoring	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
empparty:gov				0.166**			
				(0.072)			
empparty:perclosejob100	0.360**						
	(0.167)						
empparty:findnewwork		0.156***					
		(0.038)					
empparty:num_benefits			0.070**				
			(0.036)				
empparty:supervisor					0.119**		
					(0.059)		
empparty:lengthwork						0.012**	
						(0.006)	
empparty:coworker_weekly							0.063
							(0.076)
Observations	1,209	1,532	1,724	1,806	1,567	1,806	1,389

Note: *p<0.1; **p<0.05; ***p<0.01

The outcome variable is the willingness to turnout outcome (five-point scale) from the survey experiment. The sample includes only respondents who received the employer or political party broker treatment. The employer treatment collapses the data along the inducement treatment arm of the factorial design used in the experiment. The sample is limited to only those who are employed. Chance of job loss measures the probability a respondent believes he or she will lose his or her job in the next twelve months. Hard to find a new job uses a five-point scale to capture the likelihood that if he or she were to lose his or her job, a respondent could find a similar one; higher values indicate more difficulty. Receives benefits captures the number of in-kind benefits (health care, education, transportation sub- sidies, etc.) respondents received from their employer. Higher values on the three-point scale used in knows supervisor well indicate better familiarity with one's boss. Number of years employed measures the length of time at one's work. Socializes with coworkers captures whether respondents spend time with colleagues outside work. All models include the constituent terms and basic demographic characteris- tics (gender, age, education, size of settlement, and an indicator for government employment). Models are estimated via ols and cluster errors at the region level.

Table 2: Using Imputed Data to Replicate Table 3

	Dependent variable:						
	Outcome Respondent Would Vote						
	Leverage					Monitoring	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
empparty:gov				0.161^{**}			
				(0.074)			
empparty:perclosejob100	0.433^{***}						
	(0.155)						
empparty:findnewwork		0.092^{***}					
		(0.036)					
empparty:num_benefits			0.057^{*}				
			(0.031)				
empparty:supervisor					0.140^{**}		
					(0.057)		
empparty:lengthwork						0.012^{**}	
						(0.006)	
empparty:coworker_weekly							−0.033
							(0.075)
Observations	1,823	1,823	1,823	1,823	1,823	1,823	1,823

Note: *p<0.1; **p<0.05; ***p<0.01

The outcome variable is the willingness to turnout outcome (five-point scale) from the survey experiment. The sample includes only respondents who received the employer or political party broker treatment. The employer treatment collapses the data along the inducement treatment arm of the factorial design used in the experiment. The sample is limited to only those who are employed. Chance of job loss measures the probability a respondent believes he or she will lose his or her job in the next twelve months. Hard to find a new job uses a five-point scale to capture the likelihood that if he or she were to lose his or her job, a respondent could find a similar one; higher values indicate more difficulty. Receives benefits captures the number of in-kind benefits (health care, education, transportation sub- sidies, etc.) respondents received from their employer. Higher values on the three-point scale used in knows supervisor well indicate better familiarity with one's boss. Number of years employed measures the length of time at one's work. Socializes with coworkers captures whether respondents spend time with colleagues outside work. All models include the constituent terms and basic demographic characteris- tics (gender, age, education, size of settlement, and an indicator for government employment). Models are estimated via ols and cluster errors at the region level.