

# Measuring the Utility of Vote Buying: Reconsidering a Linear Approach to Distance

Chengyu Fu, Frances Cayton, and Molly Hickey

November 30, 2020

## Abstract

Building on Francisco Cantú's 2019 paper "Groceries for Votes: The Electoral Returns of Vote Buying," we reconsider the efficacy of a vote buying scheme in the 2012 Mexican Presidential Elections in which voters were given giftcards in exchange for their support. The author's initial analysis rests on an assumption that the distance between one's precinct and the store where the giftcard is valid is exogenous to political beliefs. We find that this is not the case, as new measures of distance show a correlation between distance and political awareness and beliefs. We conclude that it is not possible to leverage distance as a measure of the perceived utility of a vote buying effort.

## 1 Introduction

In his paper 2019 "Groceries for Votes: The Electoral Returns of Vote Buying," Francisco Cantú argues for the persuasive power and electoral impact of a vote buying scheme in the 2012 Mexican Presidential Election. Cantú's analysis shows a correlation between electoral swings towards the candidate executing the vote-buying scheme (PRI candidate Enrique Peña Nieto), and voters' proximity to a Soriana store, where the gift cards offered as payment could be redeemed. It's worth taking a second look at the measure, proximity, being employed here. Distance and proximity are common independent variables in political science, but they shouldn't be taken as interchangeable.

Proximity to Soriana stores is fundamental to the results of the paper. The author relies on the assumption that proximity to the stores is uncorrelated with voter preference, and thus sufficiently endogenous as to be a useful proxy for measuring the efficacy of vote buying. Cantú backs up this assumption by showing that the coefficients on a broad array of measures of political behavior and previous electoral results are unrelated to proximity to a Soriana store. He also uses placebo tests to illustrate that similar vote swings are absent in previous elections and that proximity to different retailers are not similarly correlated with vote swings.

Important to note is that when Cantú refers to proximity, it is a measure he defines as the inverse distance between the centroid of a precinct and its closest Soriana store. However, Cantú does not explain the rationale behind choosing the proximity stores as opposed to the pure distance. In the following, we will argue that this decision of using proximity, the inverse distance, as opposed to

simply the distance, is integral to Cantú’s results. When we ran the author’s analysis using distance instead of this measure of proximity, we found that his results did not hold.

Besides the differences between measures of proximity and distance, we also expanded our analysis to include non-linear measures of distance. The use of proximity and distance implies a linear relationship between one’s distance from a Soriana store and the utility of the gift card. Of course, there is a difference in utility for someone located 1 km from a store and 50 km, but it’s not clear that the difference in utility is the same between 1 km and 2 km on one hand, and between 50 km and 51 km. In fact, common sense tells us that there is likely diminishing marginal returns to distance. Furthermore, there is likely a distance beyond which the gift card has no utility, and at this point, we could expect to see a drop-off in utility.

The author’s initial argument rested on this implication that distance effectively proxies for a person’s valuation of the giftcard. If we do not have a viable measure of distance in the model, however, this assumption may not hold. In turn, the initial finding that giftcard vote buying schemes are effective may be called into jeopardy. This raised a series of questions: is distance between the centroid of one’s precinct and the nearest Soriana truly endogenous to political beliefs? If so, how do we measure distance? If distance cannot be used as a reliable proxy, how do we measure the success of vote buying attempts that follow the model of the PRI vote buying scheme?

To address these questions this paper proceeds in following manner. First, we re-estimate the model using two new measures of distance: coarsened quantiles and deciles. In relaxing our assumptions that distance is a linear variable, we are able to capture a potential shift utility of a giftcard at different distances. Re-estimating the author’s model for the PRI vote share (Enrique Peña Nieto’s party) using these new measures, we find evidence of model dependence.

Further, we find that the author’s initial assumptions rested on extrapolations of voter behavior to cases of *further* distances to the Soriana store in opposition strongholds with high turnout, when few cases fitting these criteria actually existed. This suggested that the model required further re-assessment. As such, we re-test the exogeneity assumption of distance and voter preference. From these tests we conclude that it is not possible to leverage distance to measure the voter utility of a vote buying effort.

## 2 Replication

### 2.1 Reconsidering Distance

We began our analysis with a re-estimation of the author’s initial models. This is in line with the belief that there is likely to be more difference in the utility of the gift cards at shorter distance than higher distances. Given that the results differ as we change the measures of distance and the model, we hold that the original analysis is model dependent. Our discretized approach allows us to capture this true dispersion of distance between the centroid of a precinct and Soriana stores.

To achieve this, we develop two new measures of distance: coarsened quantiles and deciles. The quantile method simply breaks up the distance metrics into four quantiles, while the deciles sort the distance metric into ten decile bins. While the range of distances included in the initial analysis spanned from under 1 km to 100 km away, we found that these decile and quantile bins were skewed toward shorter distances. In fact, the 75th percentile mark was 4.78km. This provides further evidence that the linearity assumption for the distance measure may not be correct but, instead, needed to better account for varying utilities at different distances from the store.

Notably, once we coarsen distance we run into a problem of collinearity. Here, we find that in each coarsened model the regression breaks down for the interaction term between the variables **HighTurnout** and **PRDstr**. This is substantively interesting, as these two variables are critical to the author’s initial hypothesis that the vote buying scheme increased vote share for the PRI in the opposing PRD strongholds (**PRDstr**) where there was high turnout (**HighTurnout**). For regression to break down here due to the collinearity issue means that we cannot get estimates of these variables.

Tables 1 & 2 show the breakdown of **PRDstr** and **HighTurnout** up to the author’s initial distance cutoff point of 20 km using quantile and decile coarsening. In each, we see that the coarsened distance finds cases of 0 high turnout in PRD strongholds for some distances. Further, joint **HighTurnout** in **PRDstr** cases comprise less than 1% of the entire 11,000-observation dataset. Moreover, of these existing cases, we notice that there are only a handful of PRD strongholds with high voter turnout under 5km away.

Critical to the paper is this assumption is that a person in a high turnout, PRD stronghold who lives closer to a Soriana store would value the giftcard more (and therefore be more willing to cast their vote for a PRI candidate) than someone who lives further away. Yet, only 1-2 strongholds exist under 5km.

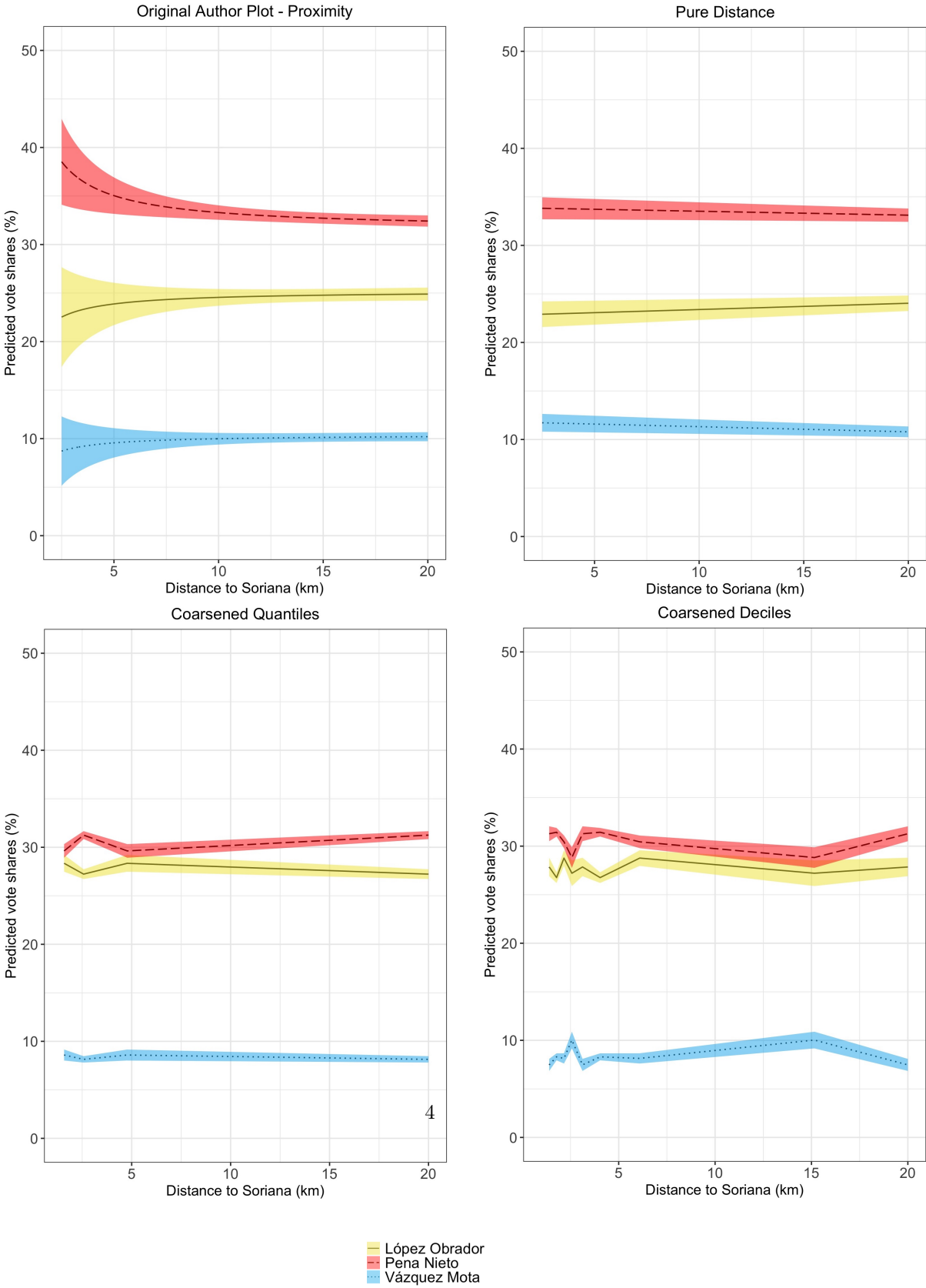
Table 1: **highTURNOUT\*PRDstr** within Quantile Bins

(0.0109,1.59]	(1.59,2.57]	(2.57,4.78]	(4.78,20]
0	0	1	47

Table 2: **highTURNOUT\*PRDstr** within Decile Bins

(.0109, .947]	(0.947,1.4]	(1.4,1.78]	(1.78,2.16]	(2.16,2.57]	(2.57,3.12]	(3.12,4.04]	(4.04,6.09]	(6.09,15.2]	(15.2,20]
0	0	0	0	0	1	0	4	31	12

Figure 1: Vote Share Estimates Shift with Distance Measure



This practical finding has serious substantive implications for the author’s model, as it challenges the assumption that we can use distance to Soriana as a proxy for the efficacy of PRI vote buying in PRD strongholds. Put simply, we are extrapolating a rare phenomenon (under 1% of all cases) that occurs at higher distances from a Soriana store to all distance ranges. We contend that because these strongholds are so rare, and for many coarsened bins do not exist, it is not possible to make this extrapolation.

## 2.2 Relaxed Assumptions Reveal Model Dependence

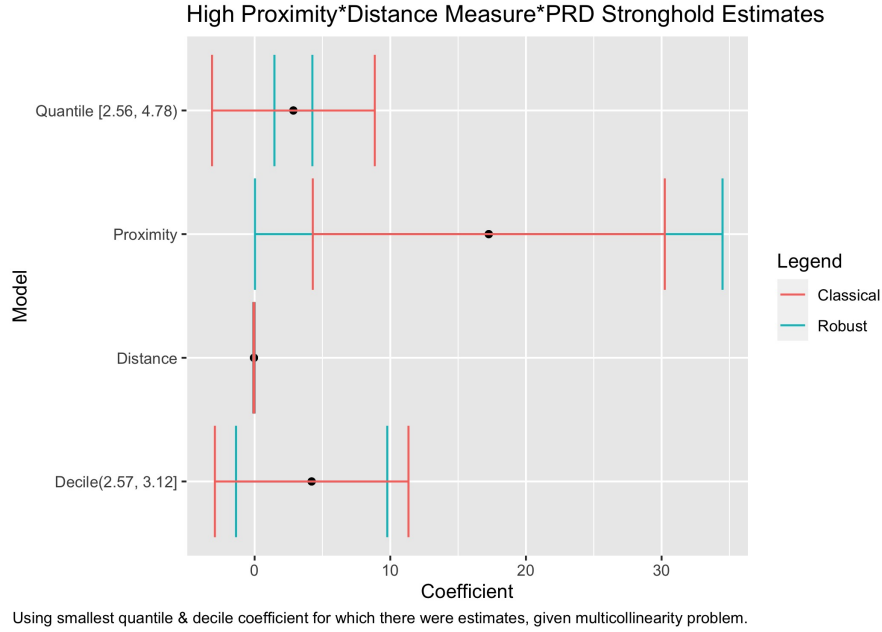
In addition to the collinearity problem, a shift in the measure of distance used changes the results of the model drastically. On page 4 we have four plots, reflecting candidate vote shares using different measurements of distance. These are visualized in the plots within Figure 1. For the plots in which a collinearity problem emerged we removed the `distance*highTURNOUT*PRDstr` interaction term, instead interacting distance solely with `PRDstr` as to provide visualizations.

In Figure 1, we notice not only does the initial PRI (Peña Nieto) vote share drop from the upper 30s to closer to 30-percent, but that this distance-influenced effect on Lopez Obrador vote share goes away entirely or changes in nature through different measures of distance. Notably, in both the coarsened models these confidence intervals of the Lopez Obrador and Pena Nieto vote shares overlap, reflecting the inability to distinguish whether distance to a Soriana is driving a vote difference for each candidate. Still, we note in the coarsened models the confidence intervals shrink, reflecting more precise estimates of our quantities of interest. This reflects the smaller standard errors of the coarsened model’s estimates.

We also note that the Robust Standard Errors in the original models, using proximity, were nearly twice the size of the Classical Standard Errors, providing further evidence of model dependence (King & Roberts, 2015). Here we find that the pure distance and coarsened models eliminate the discrepancy between robust and classical SEs. This is visualized in Figure 2, in which it is evident that, of the estimates, the proximity measure appears to produce both biased and inefficient results in comparison to other distance measures.

The models in Figure 1 & 2 suggest that either a linear distance measure or coarsened decile measure of distance reduces the dissonance between Classical & Robust Standard errors, minimizes bias and the standard errors, and, as such, minimizes model dependence. It is notable that under

Figure 2: Improved Estimates with Distance Measure



either of these models there is no clear distance-driven effect of vote buying. Here, under the traditional distance measure the Peña Nieto voteshare is estimated either consistently above 30% (but lower than the proximity measure) with no short-distance bump. If we turn to the coarsened quantile model, the measure of distance suggests a slight vote-bump in short (under 5km) ranges though, as demonstrated in Table 1, this is an extrapolation as there are no observations with both high turnout and a PRD stronghold within these short ranges.

### 2.3 Exogeneity Questions

Given this extrapolation between long- and short- distance trends, we raise the question as to whether the use of distance to a Soriana store as an independent variable is rather simply a proxy for urban versus rural residence. In the paper, the author says that he controls for urban or rural residence, but when we examined the regression, this involved controlling at the municipality level. Municipalities in the State of Mexico are quite large, often exceeding 100 square miles ([click here for more information](#)).

As a result, it's likely that controlling for municipality does not accurately control for urban versus rural, as a single municipality can contain both urban and rural areas. This means that municipality-level controls cannot account for highly-local trends, such as whether one has a Soriana

store within walking distance or 20km away. We investigate this further below, in Figure 3, which examines how distance to a Soriana store also correlates with beliefs, so it cannot be considered as an exogenous variable to measure the impact of vote buying.

It's important to take a step back and look at the validity of the claims being made in the paper. Activists claimed that Peña Nieto's campaign distributed cards worth between 100 and 1,000 pesos (roughly 8 USD to 80 USD) to 1.8 million voters in Mexico state in the days prior to the election (Nichter Palmer-Rubin, 2015). In 2019, the average Mexican made 17,595 USD, or around 8.45 USD an hour (OECD, 2017). It's possible to imagine that perhaps a gift card worth, on the lower end, an hour's wage, might not be valuable enough to swing someone's vote, but on the upper end, the gift cards could be worth more than an average day's wage, which would likely be meaningful to most. However, it's impossible to be certain about to what degree recipients would value the cards, and they likely had a mixed effect among recipients.

Beyond the value of the cards among recipients, there also remains the possibility that recipients might defect and vote for another candidate. In the Mexico 2012 Panel Study, 87 percent of respondents reported believing that their vote choice remains secret (Greene, 2016). If 87 percent of the recipients believed that their votes were secret, even if the value of the gift card was contingent on Peña Nieto's victory, it's reasonable to imagine that some might defect and vote for their favored candidate. Given these contingencies, there are multiple places where the vote-buying effort could have broken down, making it difficult to make sense of it in this analysis.

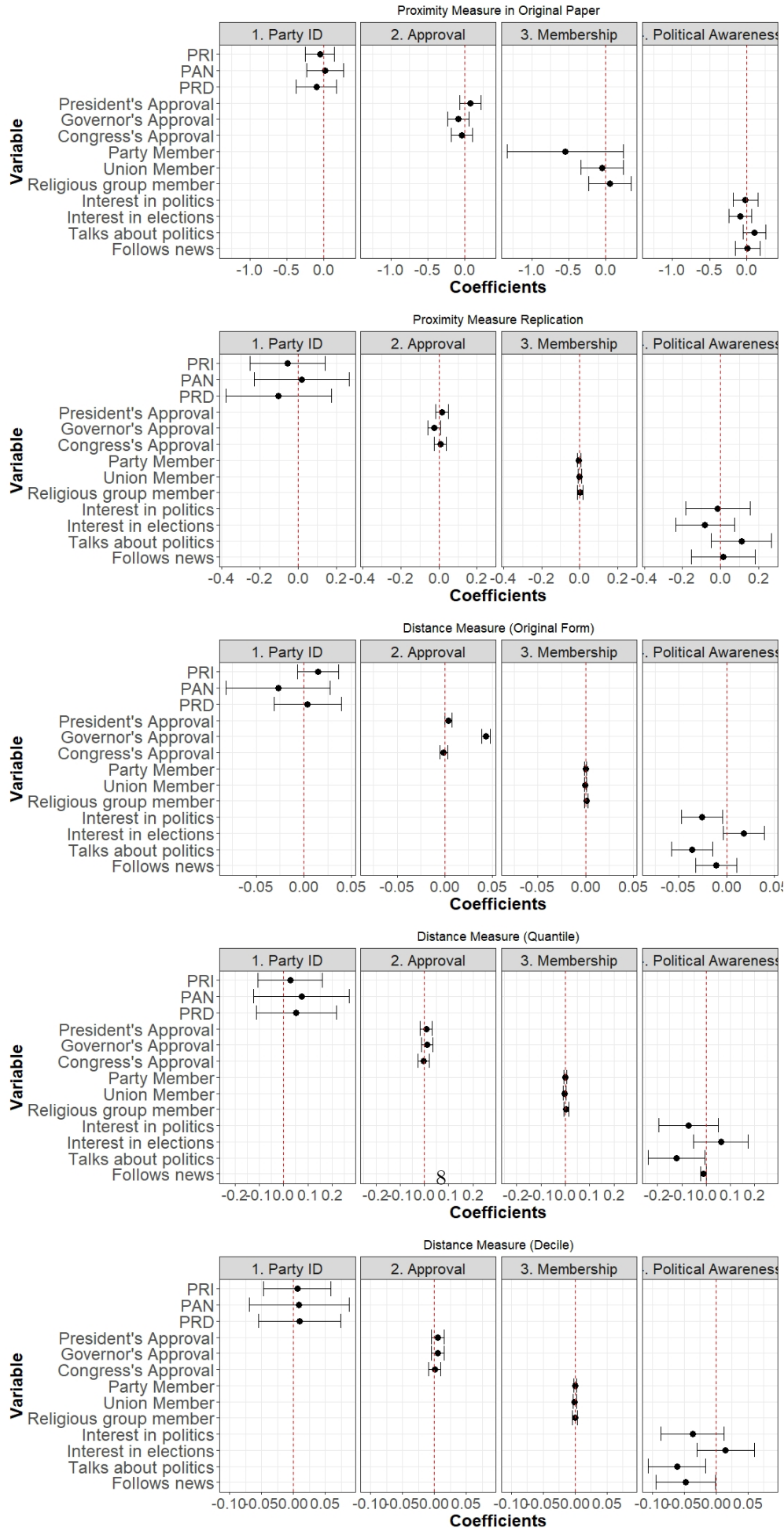
With the understanding that the current estimation of distance induces model dependence, we turn to the central assumption: that distance is not exogenous to voter preference. Figure 3 (below) is an extension and replication of Figure 1 in the author's original paper, which tests for this potential endogeneity.

The first row is the original figure, the second is our replicated figure (with no changes in codes or data). Both two rows use proximity as measurement of gift card values. Notably, here, our replication of the initial figure without changing any of the author's initial assumptions, code, or data finds far smaller confidence intervals for the correlation of distance and party membership, and far wider confidence intervals for the correlation between distance and political awareness.

For rows 3 to 5 of this figure, the authors re-estimate the model with distance, rather than proximity, as measurement of gift card values. Rows 3 to 5 use pure distance, distance quantiles, and distance deciles for measurement separately.

We can see that, in row 3, when replicating Figure 1 with "distance", the assumption that proximity is exogenous may not hold. For the pure distance model, distance is positively correlated with approval for the president and the governor. It is also negatively correlated with political awareness. Though the results for political approval may not be robust for distance quantile and decile models, we still find that distance is negatively correlated with political awareness, e.g., people's discussion about politics and people's following political news. Therefore, the distance to Soriana may not be used as a variable exogenous to political attitudes and political behavior that reflects purely people's expected value of gift cards.

Figure 3: Correlation between Distance and Political Beliefs, by Distance Measure



### 3 Conclusion

From these tests, we argue that one must take a careful approach to distance when using its as a proxy for one's political beliefs. As our final tests demonstrate, distance may not be truly exogenous to political beliefs. If we take either the less biased and more efficient pure distance or coarsened quantile models, we note that distance could be correlated with politician approval (row 3) and negatively correlated with political awareness (rows 3 and 4). We conclude that it is not possible to leverage distance to a 'point of redemption' as a measure the voter-level utility of a vote buying effort.

### 4 Sources cited

- Cantú, Francisco. (2019). "Groceries for Votes: The Electoral Returns of Vote Buying." *The Journal of Politics*, 81(3), 790-804.
- Greene, Kenneth. *Mexico Panel Study*, 2012, Inter-University Consortium for Political and Social Research [Distributor], 11 Mar. 2016, [www.icpsr.umich.edu/web/RCMD/studies/35024](http://www.icpsr.umich.edu/web/RCMD/studies/35024).
- King, Gary, and Roberts, Margaret E. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis* 23, no. 2 (2015): 159-79.
- "Mexico - OECD Data." *The OECD*, Organisation for Economic Co-Operation and Development, 2019, [data.oecd.org/mexico.htm](http://data.oecd.org/mexico.htm).
- Nichter, Simeon, and Brian Palmer-Rubin. "Clientelism, Declared Support, and Mexico's 2012 Campaign." *Mexico's Evolving Democracy*, Johns Hopkins University Press, 2015, pp. 200-226.