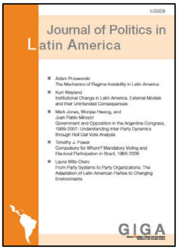


Journal of Politics in Latin America

Bahamonde, Héctor (2018),
Aiming Right at You: Group versus Individual Clientelistic Targeting in Brazil, in:
Journal of Politics in Latin America, 10, 2, 41–76.

URN: http://nbn-resolving.org/urn:nbn:de:gbv:18-4-11211

ISSN: 1868-4890 (online), ISSN: 1866-802X (print)

The online version of this article can be found at: <www.jpla.org>

Published by
GIGA German Institute of Global and Area Studies, Institute of Latin American Studies
and Hamburg University Press.

Journal of Politics in Latin America 2/2018: 41–76 GIGA

Aiming Right at You: Group versus Individual Clientelistic Targeting in Brazil

Héctor Bahamonde

Abstract: Do parties target individuals or groups? Although this question is fundamental to understanding clientelism, the literature does not offer an answer. This paper argues that, depending on certain conditions, brokers target individuals when they are identifiable, and groups when brokers need to rely on the spillover effects of clientelism. Both identifiability and spillovers depend on individual poverty, group poverty, and political competition. Though the theory I outline focuses on targeting, I also argue that structural factors, such as the density of the poor, should be considered in the vote-buying literature. Structural factors are one of the few observables upon which brokers can base their decision regarding investing in clientelism. Using survey and census data from Brazil, the paper exploits variations in personal incomes within contexts of differing levels of municipal poverty. I find that political parties engage in segmented or ad-hoc strategies, targeting individuals when identifiability is high, and groups when there are economies of scale. Importantly, non-poor individuals can also be offered clientelism.

Manuscript received 18 June 2017; accepted 29 June 2018

Keywords: Brazil, clientelism, vote-buying

Héctor Bahamonde is an assistant professor at O'Higgins University (Rancagua, Chile). In 2017 he received his PhD in Political Science from Rutgers University (New Brunswick, NJ). Then, he served one year as a post-doctoral fellow at CIPR, Tulane University (New Orleans, LA). His current research focuses on clientelism, the political economy of taxation, state formation, and Latin American politics. Personal website: <www.hectorbahamonde.com>
E-mail: <hector.bahamonde@uoh.cl>

There is no agreement on when, how, and why parties choose to aim clientelist practices at individuals or groups.[1] The distributive politics and vote-buying literatures have traditionally pursued one of two approaches. The former has mostly focused on group targeting, usually districts or provinces (Dixit and Londregan 1996; Khemani 2015; and Calvo and Murillo 2004), showing that incumbent parties deliver public-sector jobs or construction projects contingent on the support of groups of people. The latter has typically focused on individuals and their characteristics, such as their socio-economic or electoral profiles. Substantively, however, it is not clear when or why clientelist brokers use either strategy.

In fact, the decision to investigate group-based and/or individual-based targeting seems to be attributable to distinct research designs and agendas, rather than theory. For example, ethnographers generally focus on individuals, while others have traditionally focused on groups (Scott 1972; Auyero 2000; Szwarcberg 2013; Weitz-Shapiro 2012; and González-Ocantos et al. 2012).[2]

What is most concerning, however, is that it is relatively assumed or implied that individual and group clientelist targeting strategies are interchangeable, when they are clearly not. Individuals pertaining to groups and individuals by themselves have different incentives to defect to the incumbent. For instance, individuals belonging to larger groups have greater incentives to defect (Stokes 2005), while individuals who are personally targeted have fewer incentives to defect (Auyero 2000). Anticipating this, brokers adjust their strategies accordingly. In the first instance, brokers deal with low-informational environments that increase principal-agent problems. In the second instance, brokers – who know their clients better – are able to leverage this knowledge, reducing the probability of defection. However, these differences have not been sys-

---

1 I am grateful to Robert Kaufman, Daniel Kelemen, Richard Lau, Paul Poast, Geoffrey Wallace, Douglas Jones, Ezequiel González-Ocantos, Juan Pablo Luna, Jorge Bravo, Eric Davis, Adam Cohon, Edwin Camp, Luciana Oliveira Ramos, Giancarlo Visconti, William Young, Johannes Karreth, and the reviewers and editor of *JPLA*. I also thank participants of the Latin American Studies Association 2014 conference, the Southern Political Science Association 2015 meeting, the Western Political Science Association 2015 meeting, and the 2014 Graduate Conference at the Political Science Department, Rutgers University. Any errors that remain, of course, are my responsibility. This work was partially funded by the Center for Latin American Studies at Rutgers University. I am grateful to the School of Arts and Sciences and the Department of Political Science for their travel grants.

2 I wish to thank Ezequiel González-Ocantos for this suggestion.

tematized in the literature. In the present paper, I propose a framework that explains when it is more efficient to target groups or individuals.

Particularly, by focusing on brokers, the paper advances an argument about the decision process regarding whom to target. The crux of the argument is that this decision is a function of three factors: individuals' discount factors explained by income levels, the incentives of clientelist brokers to rely on spillover effects caused by the nesting structure of individuals (that is, whether individuals are nested in poor or non-poor contexts), and brokers' incentives to engage in clientelism explained by higher electoral pressures and political competition.

Overall, I share Carlin and Moseley's (2015: 14) opinion that "[e]xisting research looks almost exclusively at individuals' socio-economic and, specially, electoral profiles [and] [y]et our knowledge of who parties target remains incomplete." The present paper seeks to contribute to this issue by incorporating both structural and individual factors that foster clientelism in the same theory. Analytically, the structure of the argument (and the empirics) allows for disentangling the effects of "being poor" and "living in a poor area." Another important implication of the argument is that I am able to suggest why parties that adopt clientelism as a strategy, target their resources to both poor and non-poor individuals, an empirical regularity that, to the best of my knowledge, has been unexplored so far.

Perhaps the area in which there is the most agreement among scholars is on the relationship between poverty and vote-buying (Calvo and Murillo 2004; Weitz-Shapiro 2012; Kitschelt 2000; and Kitschelt and Altamirano 2015). For example, Brusco, Nazareno, and Stokes (2004), Stokes et al. (2013), and Nazareno, Brusco, and Stokes (2008) explained that since the poor derive more utility from immediate transfers than the uncertain returns associated with future policy packages, clientelist political parties only target the poor. In fact, Weitz-Shapiro explained that "[a]lmost *universally*, scholars of clientelism treat and analyze [this] practice as an exchange between politicians and their poor clients" (Weitz-Shapiro 2014: 12; my emphasis).

However, this canonical predictor has recently been contested (Hicken 2007: 55). Szwarcberg (2013: 32) "challenges the assumption [that brokers] with access to material benefits will always distribute goods to low-income voters in exchange for electoral support," while González-Ocantos et al. (2012) and Holland and Palmer-Rubin (2015) found that income (measured at the individual level) had little or no effect on vote-buying. In fact, Figure 1 shows that non-poor individuals in Brazil did receive clientelist offerings. *Why would brokers target non-poor individuals?*

And relatedly, *why does contemporary scholarly work report null findings for poverty, traditionally the most important predictor of vote-buying?* I present an argument where individual income alone is not relevant (similarly, see Weitz-Shapiro 2012: 568). What matters is how noticeable individuals are. Wealthier individuals living in poor contexts and poor individuals living in non-poor contexts are more identifiable, increasing their respective probabilities of being targeted. I also contend in this article that, in low-information environments, brokers use these kinds of observables to reduce the probability of defection of their clientele.

Another often-considered contextual factor in the literature is the size of the community in which clientelism takes place. Large-sized communities impose severe principal-agent problems. Stokes (2005: 323) explained that the "community structure" mediates the incentives to defect. Large communities make voters more anonymous, increasing their probability of defection. In fact, Rueda (2017: 164) found that in Colombia vote buying is more effective in contexts of small polling places.

Figure 1. Individual Wealth and Vote-Buying in Brazil

Note: Following the advice of Córdova (2008) and Córdova and Seligson (2009, 2010), different socio-economic variables in The Latin American Public Opinion Project (2010) dataset were used to construct a relative wealth index (RWI). With this information, in addition to the frequency of clientelism question (*clien1*), the figure shows that clientelist brokers target individuals at all levels of income.

Several scholars have then argued that brokers prefer smaller groups because individuals nested in small communities should defect less (Brusco, Nazareno, and Stokes 2004; Kitschelt and Wilkinson 2006: 10; and Magaloni 2008: 67. See also Bratton 2008, for Nigeria, and Gingerich and Medina 2013: 456, for Brazil). Yet, even when brokers might prefer to target small communities (with fewer voters relative to large communities), it is not clear how political parties gain enough electoral returns, especially considering that clientelism is expensive.

Vote-buying is an expensive strategy (Zarazaga 2014: 35), and more so when clients are individually targeted.[3] Stokes (2005: 317) argued that brokers develop skills that allow them to infer whether individual clients in small-sized communities voted for their party by looking at them in the eyes. Gay (1993, 1998) documented similar findings for the Brazilian case. This strategy requires brokers to sustain close relationships over time with their clients in a personal and individualized way. Knowing the clients' needs, delivering them benefits, monitoring their political behavior (and punishing them in case of defection), all in an individualized fashion, makes this strategy an extremely expensive choice – and it becomes even more expensive as more individuals are added to the broker's portfolio.

The cost of individual targeting increases linearly with the size of the targeted population (Hicken 2007: 56). This intuition is important because the brokers' production-possibility frontier cannot be shifted upwards either. Since the number of brokers is a depletable resource, at some point party machines run out of brokers, implying that monitoring capacities are bounded. In fact, Auyero (2000: 74) explained that the capacity brokers have to deliver benefits is "finite," and "only for a restricted number of people." However, and despite this constraint, brokers still have incentives to secure a large number of votes. Yet, the literature explains that clientelism should decrease in large communities. However, it is hard to conceive that brokers will stop being clientelist just because the size of the population is large. A priori, it seems a missed opportunity for brokers to let go of a large number of votes. In fact, survey data for the Brazilian case indicate that inhabitants of large, medium, and small municipalities are targeted in virtually the same proportion (Speck and Abramo 2001: 2). This article explains that when

3 Dixit and Londregan (1996: 1147) explained that brokers track "constituents' likes and dislikes, *compulsively* participating in a spectrum of events [such as] baptisms and bar mitzvahs, weddings and funerals [and even, holding] *daily* meetings with constituencies [even] *after* nine o'clock [hearing] what anyone wished to tell [them]" (My emphasis).

brokers need to secure large amounts of electoral support, especially when political competition is high, they turn to group-targeting strategies, relying on the spillover effects of clientelism. In these contexts, clientelism mobilizes electoral support from "actual" and "potential" beneficiaries, minimizing the costs of clientelist targeting while maximizing electoral benefits, a mechanism that I explain later on in the paper.

Civic associations might help solve some of the challenges large-sized groups present to brokers. As low-information environments prevent brokers from really observing individual electoral behavior (Zarazaga 2014: 35), they usually resort to alternative methods that allow them to make safer inferences. For example, Schaffer and Baker (2015: 1094) argued that clientelism is "socially multiplied" as party machines target individuals "who are opinion-leading epicenters" in informal situations, or "partisan networks" (Calvo and Murillo 2013), in what has been called "organization buying" (Stokes et al. 2013: 250–251).[4] If parties buy "turnout" (Nichter 2008), then they will most probably target associations too, as "citizens immersed in clientelist networks [...] have a higher probability of voting than the rest" (Carreras and Castaneda-Angarita 2014: 7). I acknowledge the positive relationship between group membership and clientelism. However, what has not been explored yet is whether clientelism is explained by association membership itself, or by the fact that poor individuals usually address their problems as a *group*, since otherwise it would be too costly to solve them individually. If this is the case, group membership should be spuriously related to clientelism. While I find that group membership does have a positive effect on clientelism, I find that structural contexts that foster group-targeting have even more explanatory power.[5]

Moving forward, Weitz-Shapiro's (2012) important paper found that in several Argentine municipalities, higher levels of political competition and low socioeconomic levels fostered higher levels of clientelism. In her paper, losses are conceptualized in terms of "moral costs." Evidence for these types of costs has been presented in the literature very recently. For example, Carlin and Moseley (2015) argued that citizens

4 Holland and Palmer-Rubin (2015: 16) explained that when "parties lack their own brokerage networks [they seek] to capitalize on organizational networks instead." Similarly, Rueda (2015: 13) argued that parties tend to target very specific civic associations of "seniors and associations of single mothers, organizing trips to recreational centers outside the city where all their expenses are covered." Paradoxically, the stronger the civic society (that is, the more organized it is), the more clientelism there is.

5 These results are presented in Figure A4.

endowed with more democratic values feel more "moral repugnance" to clientelism. Vicente (2014) showed that vote-buying practices have an "immoral/illegal connotation," while González-Ocantos et al. (2012) found that individuals wanting to avoid social stigma usually do not give truthful answers when asked directly about clientelism. Building on this literature, I contend that when political competition is high, clientelism will be higher in contexts where poor individuals live in poor economic contexts.

When Do Parties Target Individuals and When Groups?

Table 1 presents four ideal types in four quadrants; cases where individuals are highly identifiable; that is, non-poor individuals living in poor areas (Q1), and poor individuals living in non-poor areas (Q4). Identifiability in these cases reduces the cost of defection, permitting clientelist brokers to closely target individuals. While individual targeting is more expensive, it is also safer (compared to group targeting). The table also shows cases where individuals are hard to identify; that is, poor individuals living in poor areas (Q2), and non-poor individuals living in non-poor areas (Q3). In these cases, voters are more anonymous, making direct individual-based targeting and monitoring more costly. Since brokers still have incentives to seek electoral support, they engage in group targeting by relying on the spillover effects of clientelism. In these cases, the effects of vote-buying disseminates by mobilizing targeted voters and latent untargeted (but potential) clients. This form of targeting is cheaper but more uncertain.

Table 1. Strategy Set: Group versus Individual Targeting

	Non-Poor Individuals	**Poor Individuals**
High Competition	Poor Areas, identifiable, individual targeting.	Poor Areas, spillover effects, group targeting, cheap vote-buying.
Low Competition	Non-Poor Areas, group targeting, expensive vote-buying, lack of checks and balances, embezzlement.	Non-Poor Areas, identifiable, individual targeting.

Source: Author's compilation.

Individual Targeting

This is the safest bet a broker can make, but also the most expensive one, as it requires brokers to have sustained closed relationships with their clients. For instance, Zarazaga (2014: 26) stated that "brokers have detailed information about their neighborhood and clients' needs." Keeping track of every single client (and their respective needs) is an expensive strategy. After all, as Auyero (2000: 73) put it, brokers are "problem solvers." Importantly, the kind of care given ranges from material needs to symbolic and immaterial necessities, making clientelism a relationship based on "trust, solidarity, reciprocity, caring, and hope." Such broker-client symbiosis is both material and personal-intensive, making it very costly. As an investment, however, it pays off electorally. The same detailed information brokers have about their clients' needs is then used to infer coercively (or know directly) the electoral behavior of their respective clientele, administering punishments or rewards accordingly (Stokes 2005: 317).

The transaction costs of clientelism are reduced by targeting identifiable clients. In 2009, Luna et al. (2011) made extensive participant observations in several campaigns, accompanying a number of candidates for several months in their campaigns for the legislative election in Santiago de Chile. With one incumbent, we spent considerable time on the ground, traveling in her district. On several times, as we drove throughout the district in her personal car, the candidate was able to recall who the head of household was (including his/her name), what her district office had contributed to solve their needs, and whether the household members were on good terms with her.[6] Importantly, the economic diversity of the district provided a number of useful observables. In non-poor areas, poor houses with an unpainted wall, a rusty front yard fence, a two-story house with a bodega market on the first, a household with a broken window, or a junk diesel truck aground in the front yard, among others, provided distinctive points of reference. Identifiability, as an observable, made these receivers less anonymous, raising their cost of defection and making them more prone to cooperate. Table 1 portraits individuals living in these heterogenous contexts in Q4.

Households in Q4, being more noticeable, stand out in their respective contexts, making it easier for brokers to notice whether they need construction materials, whether there are wakes to which they could contribute flowers or birthday parties to which they could bring cakes. In

6 The actual gender of the candidate might have been changed for confidentiality purposes.

addition, it makes their possible defection more obvious and memorable for the brokers. In summary, higher levels of visibility supply brokers with good-quality information about their clients.[7] In addition, when political contestation is low, the demand for votes is less astringent, shaping brokers' incentives to target in a more accurate, less massive fashion, identifiable and particularized individuals, not groups.

The capacity brokers have to identify potential clients not only comes from third-party sources, as the "organization buying" proponents explain (Holland and Palmer-Rubin 2015; Rueda 2015 and Stokes et al. 2013). In a similar account, others have pointed out that brokers are also "reliable neighbors" (Zarazaga 2014: 38); that is, members of the same community of targeted individuals. Acknowledging this approach, the argument presented in this article contends that brokers have incentives to expand their immediate local networks by colonizing visible targets outside of their own proximate neighborhood. By conceptualizing brokers as active political entrepreneurs who seek new supporters outside of their immediate context, the proposed framework complements other accounts, as presented in Szwarcberg (2013: 32) or Zarazaga (2016: 681), where brokers are neighborhood party agents. Clientelist entrepreneurship can be performed directly or indirectly. For instance, Auyero (2000: 65–66) described the situation of Cholo, a member of the inner circle of one of the brokers in Buenos Aires, Argentina, who visited "other poor neighborhoods of the area adjacent to" the place where the broker (and himself) lived, to spread news about some government plan, the governor, and the Peronist party, but importantly, also reporting to the broker any unattended material needs he had noticed. This illustrates how, via different channels, brokers expand their client portfolio outside of their immediate community.

An important implication is that individual poverty does not play a role by itself. Non-poor individuals living in poor areas (Q1) are also noticeable, and consequently, possible targets as well. Political competition shifts the demand for votes upwards. As elections become more contested, brokers need to secure even higher levels of electoral support. Since newly elected representatives are more likely to bring new people to their machines, brokers are also interested in seeing their candidates elected. Consequently, brokers will have even more incentives to engage in clientelism when political competition is high. In these cases, political competition is high enough to even mobilize non-poor individuals in a

7 Importantly, poor households do not need to be close to each other, but visible enough.

clientelist way. Since these votes are more expensive to purchase (given decreasing marginal utility from income, see Stokes 2005: 321), this strategy is less preferred. However, costly clientelism is worth the investment given the risk of losing the election.

Group Targeting

This is the least accurate targeting strategy, but also the cheapest one available to brokers. It leverages the spillover effects provided by larger concentrations of individuals who share the same socio-economic backgrounds. This strategy is less accurate because it mobilizes electoral support from "actual" clients (individuals who have actually been targeted), and "potential" clients (individuals who have not received benefits yet). It is preferred when poor individuals are nested in poor areas (Q2), or vice-versa (Q3). In these cases, individuals are masked by their environments, which means that identifiability is hard to achieve. As explained before, identifiability facilitates individual targeting, an important factor in reducing the probability of defection of targeted clients. When individuals are hard to identify, however, individual targeting becomes prohibitively expensive. Yet, brokers who still need to secure electoral support do not opt out of clientelism and instead turn to group targeting.

Auyero (2000: 65) described the case of Alfonsina in Argentina. Alfonsina was part of the broker's inner circle and got a job as a cleaning lady in a public school. As the broker explained to her, before getting the job, Alfonsina had to be *patient* because as a member of "the circle," she was in the pool of potential beneficiaries; it was only a "matter of time" until she could get the job. The idea of expectations and hope are important. Auyero explained that the

> *hope* of a job serves as important glue within the inner circle. Although not everyone is employed at the municipality, the fact that someone gets [a] job has an important *demonstration* effect. (Auyero 2000: 65; my emphasis)

Building on this intuition, two ideal types are suggested: actual and potential beneficiaries. The former receive particularistic benefits "today" and vote for the broker's candidate "tomorrow," while the latter do not receive benefits "today" (in the expectation of receiving them in the future) but still vote for the broker's candidate "tomorrow."

Group targeting is cost-effective because it mobilizes two types of voters at the cost of investing in just one (i.e. the "actual"). Actual beneficiaries want to remain actual beneficiaries since they want to keep receiving benefits; thus, they keep supporting the broker's candidate. In turn, potential beneficiaries want to become actual beneficiaries, but are uncertain when that might happen; as a result, they also support the broker's candidate. In this sense, from the broker's perspective, this strategy reduces the sunk costs by half, multiplying the gross benefits by two. In other words, the broker's reputation of a "problem solver" disseminates twice as fast relative to individual targeting. It is in this sense that this is a massive (but less precise) form of clientelist targeting.

Given that potential clients support the broker's candidate in the absence of current inducements, brokers need to effectively calibrate the timing when potential beneficiaries become actual beneficiaries. In other words, brokers need to infer the discount factors of their potential clients, making it expensive for them to defect. Reputation, as a form of capital, is fundamental for brokers since "voters prefer to support [brokers] with a reputation for delivering because they are a more reliable source of future rewards" (Zarazaga 2014: 24). However, potential clients are also interested in investing in their reputation. From their perspective, they know that the flow of resources is dependent on the brokers' electoral success. Also, they do not know whether new brokers might have access to fewer resources or distribute them to other people. For them, the cost of switching brokers (or defecting) is very high since it also involves building relationships of confidence with another broker from scratch, which is costly. Hence, the incentives are for the broker to deliver benefits before it is too late, while the incentives for the potential client are to support the broker's candidate.

Since it does not matter what type an individual is, both actual and potential beneficiaries keep voting for the broker's candidate. While cost-effective, group targeting is less accurate since brokers hope to mobilize potential beneficiaries only indirectly; that is, by targeting actual beneficiaries. This makes this strategy a fragile one. However, besides the reputation costs described above, low-income voters have additional incentives to support the broker's candidate. This is described in Q2. Given that the poor are risk-averse, potential beneficiaries are better-off waiting (and voting for the broker's candidate) than defecting. In the same vein, but on a slightly different subject, Magaloni (2008: 20) posited that the Mexican PRI lasted as long as it did not because of electoral fraud but because voters supported the "known devil." Economic underdevelopment played a fundamental role in this equilibrium as well. Finally, higher levels of electoral contestation force brokers to engage in this less accurate, but massive form of clientelist targeting, leveraging (1) the incentive structure of potential clients to support the candidate even in the absence of current inducements, and (2) the higher levels of risk aversion poor individuals have.

Importantly, vote-buying is also targeted to non-poor individuals nested in non-poor groups (Q3). Vote-buying has decreasing returns to scale in non-poor individuals. That is, wealthier individuals derive fewer advantages from a bag of rice relative to poorer individuals (Kitschelt 2000). Anticipating this, brokers will not offer the same benefits to wealthy individuals, but will customize the type of offerings. This distinction is important, since most of the literature assumes that clientelist practices decrease when individual incomes rise. However, that approach does not explain the counterintuitive empirical regularity depicted in Figure 1; that is, non-poor individuals get targeted too. *Why are non-poor individuals targeted?* This article seeks to contribute to the literature by explaining that brokers make their offers more attractive to non-poor individuals by offering goods that are relatively more expensive. This is more likely when districts are wealthier.

While buying votes from non-poor individuals costs more, brokers in non-poor areas have more resources to spend. Along the same lines, Hicken (2007: 55) questioned the implicit assumption that the broker's vote buying funds remain fixed; stating that "a candidate's capacity to buy votes increases commensurate with increases in average incomes." In other words, higher levels of economic development not only raise personal incomes, but also shifts the broker's vote-buying capacities upwards. Similar evidence has been found in the Philippines (Schaffer 2004). The link between higher incomes and vote buying is particularly relevant for Brazil, since its electoral laws allows political parties to get *unlimited* funds (Abramo and Speck 2001: 14), enabling brokers greater capacities to buy more expensive votes.

Besides having more resources to spend, brokers in politically uncontested districts have fewer political constraints, facilitating the spending of expensive clientelism. In Q3 it is suggested that lower levels of political contestation allow brokers to spend on more expensive means of clientelism. Uncompetitive districts lack proper *de facto* mechanisms of checks and balances, giving local incumbents more "room to move," allowing them to divert local resources into more expensive means of targeting. I call this "embezzlement clientelism." Given these relatively more expensive costs, however, I expect this form of clientelism to be less frequent. In a dynamic similar to Q2, potential clients also support the broker's candidate, hoping to become actual beneficiaries. However – and unlike poor clients in Q2 – non-poor clients in Q3 (both actual and potential) have smaller discount factors. That is, non-poor individuals – given their relatively higher incomes – have more "patience." This is especially important for brokers. In practice, potential clients' timing constrains are more elastic, putting less pressure on brokers to deliver benefits in the short run.

Case Selection, Research Design, and Data Analyses

I. Data

This section empirically tests the theoretical proposition stated in Table 1 – that is, the combined effects of individual income, of being nested in poor/non-poor communities, and being exposed to different levels of political competition – on receiving clientelist benefits. Brazil is a good case because its poverty structure is such that it is possible to find low-income individuals nested in non-poor areas (and vice versa). This case is also interesting from an institutional perspective. The Brazilian electoral system incentivizes clientelism. Several factors such as multimember districts with open lists, and the institution of the *candidato nato*,[8] "clearly [makes] Brazil one of the most personalistic systems of democratic governance" (Kitschelt and Altamirano 2015: 257), which might foster higher levels of clientelism. In fact, Gingerich (2014: 290) found that vote-buying drastically changed electoral results, concluding that "[v]ote brokerage can still pay electoral dividends in contemporary Brazil."

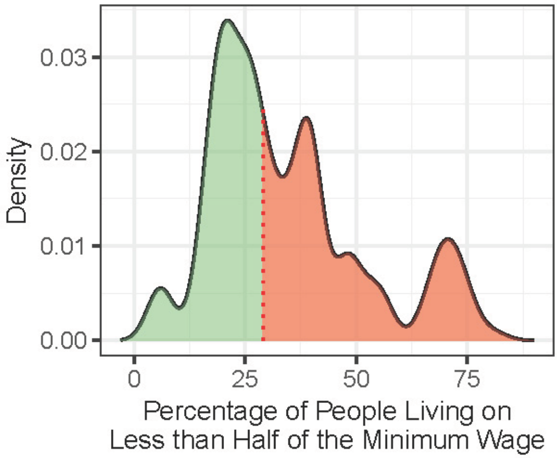
To test this hypothesis, I use survey data from 2010 from The Latin American Public Opinion Project (LAPOP) (2010).[9] Though the LAPOP survey provides a question for income, people who are somewhat better off than their neighbors but live in poor areas may not "feel" poor. If this is the case, it could confound the results. Additionally, when answering the questioner, individuals might not want to reveal their true incomes (either because they are too low or too high). Following the advice of Córdova (2008) and Córdova and Seligson (2009, 2010), a

8 "[R]ule that removed parties' control over the nominations process and let an electoral legislator decide to run on any party ticket." See Kitschelt and Altamirano (2015: 257).

9 "I thank the Latin American Public Opinion Project (LAPOP) and its major supporters (the United States Agency for International Development, the United Nations Development Program, the Inter-American Development Bank, and Vanderbilt University) for making the data available." The sample consists of five strata representing the five main geographical regions of Brazil. Each stratum was further sub-stratified by urban and rural areas.

relative wealth index (RWI) was constructed (see also Santos and Villatoro 2018). Using principal component analyses, the index measures wealth based on actual assets weighted by how common these assets are. Different indices were constructed for urban and rural contexts. Figure 1 plots the distribution of the index.

Figure 2. Distribution of the Density of the Poor



Note: Employing Brazilian census data from the IBGE (2010), the figure shows the percentage of individuals who live on less than half of the minimum wage in a given municipality. While individual income is measured using the relative wealth index (in Figure 1), the variable plotted here is used to measure economic development at the group level. Due to statistical reasons explained in this paper, the variable had to be dichotomized at its median (29 percent). However, in separate statistical analyses shown in Table A3 (weighted model), the variable is used without dichotomizing it, showing the same results.

II. Main Variables of Interest

To measure economic development at the group level, I constructed a variable that I call "the density of the poor" following a strategy similar to that of Weitz-Shapiro (2012). The variable, which is plotted in Figure 2, measures the degree of poverty at the municipal level. Using information from the 2010 Brazilian census,[10] a semi-continuous variable was constructed to measure the percentage of individuals who live on less than half of the minimum wage in a given municipality. Given that the

10 Official data comes from the Bureau of Statistics of Brazil IBGE.

municipality of residence for each individual in the LAPOP survey is recorded, I was able to merge the census percentage with the LAPOP dataset. It is important to stress that the unit of analysis is the individual, and that this variable captures the economic context in which each individual lives. Just like other scholars in the past have tested the effect of being nested in rural areas,[11] this paper focuses on another class of contextual variable. Although the density of the poor group was originally a semi-continuous variable (that is, a percentage), it had to be dichotomized at the median (29 percent) to be able to construct a matched sample, which I justify and explain below. Figure 2 shows the continuous distribution dichotomized at the median (dotted line).

Finally, to measure political competition, I again follow Weitz-Shapiro (2012). Using official electoral data from the 2008 municipal elections,[12] I constructed a variable that measures the percentage of seats that are not controlled by the mayor's party in a given municipal council.

III. Matched Design

There is a built-in lack of relationship between "being poor" and "living in a poor municipality," confirming that Brazil is in fact a good case to test this theory. Figure A1 in the Appendix shows that the unmatched/raw dataset already has embedded low levels of correlation between these two variables ($r = -0.44$).[13]

I was able to break this relationship down further using matching methods. Matching is a two-stage process. In the first stage, the analyst "preprocesses" the data, seeking to break any systematic relationship between, in this case, the density of the poor and the relative wealth index RWI (Ho et al. 2011). Matching does so by deleting observations for which similar observations cannot be found.[14] The idea is to obtain a good covariate balance, as in Figure A3 (in the Appendix), to then estimate any appropriated statistical model.[15] From a statistical standpoint,

11 See, for example, Brusco, Nazareno, and Stokes (2004) and Stokes (2005). Both studies used the log of population, which is a proxy for urban/rural.

12 Data from the Tribunal Superior Eleitoral.

13 The figure shows that, for both the matched and raw datasets, "being poor" and "living in a poor municipality" are not confounded, as it is possible to find poor individuals living in non-poor areas, and vice versa.

14 The final procedure matched 761 individuals living in the low-density poverty condition with 676 individuals living in the high-density poverty condition.

15 The idea is that the propensity of being exposed to the "high" density of the poor condition (or "propensity score") has a similar distribution in both "treat-

preprocessed datasets are less model-dependent (Ho et al. 2007),[16] and prevent analysts from making extreme counterfactuals.[17] The preprocessed data used in the matching approach has 54 municipalities, while the raw data used in the generalized propensity score (GPS) approach (which I explain below) also has 54. Figure 3 lists the municipalities and shows which ones are considered "high" or "low" in terms of the density of the poor after the dichotomization process. The figure also shows that there is considerable variance in income/RWI in both high- and low-poverty density conditions (bubbles).[18]

One could argue that dichotomizing the density of the poor variable at the median is an arbitrary decision. While there have been theoretical advances regarding general treatment effects regimes for continuous or semi-continuous response doses (Imai and Dyk 2004; and Hirano and Imbens 2004), algorithms with the ability to match on continuous treatment variables are not common. In order to obtain covariate balance in a non-parametric way (as matching does) but without dichotomizing the density of the poor, I also use the original (that is, continuous) density of the poor variable to construct a generalized propensity score GPS (Imbens 2004; Guardabascio and Ventura 2014; and Imai and Ratkovic 2014).[19] The score is used to weight each observation in the model.[20]

---

ed" and "control" groups. It is important to note that, despite the language, I do not claim any causal relationship in this paper.

16 Table A2 and Table A1 in the Appendix provide summary statistics for both the matched and raw datasets. Tables were generated using the *stargazer* R package. See Hlavac (2015).

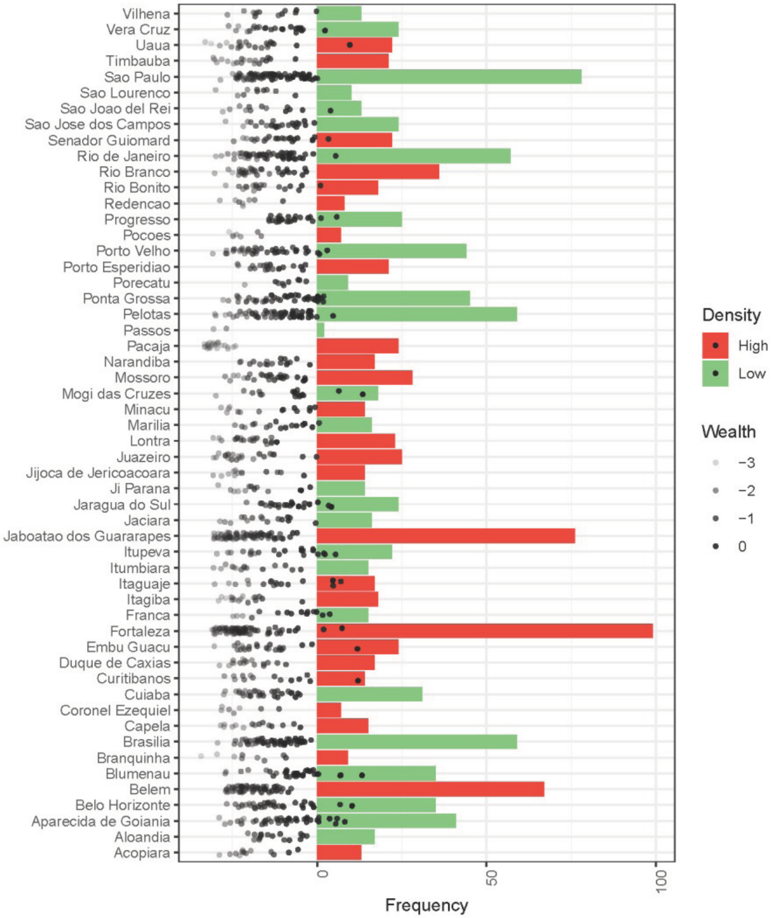
17 King and Zeng (2005). The matching routine used was the *full* matching routine (Hansen 2004; Rosenbaum 2010), via the *MatchIt* R package (Ho et al. 2011).

18 Figure A2 in the Appendix shows the frequency of individuals by municipality in both raw and matched datasets.

19 To generate the weighting vector, I used the *CBPS* R package (Fong et al. 2018).

20 Besides matching on and weighting by the RWI index, I also included the following variables to match on/weighting by: municipal opposition, municipal population and individual involvement in civic associations.

Figure 3. Distribution of Observations by Municipality, Wealth Index and Density of the Poor



Note: The figure shows the municipalities in the analyses (matched set). For every municipality, the figure shows (1) the number of inhabitants (Y-axis), and (2) whether the municipality is considered having a high or low density of the poor. High-density municipalities have more than half of their inhabitants living on less than half of the minimum wage. The figure also shows (3) individual wealth indexes (bubbles).

IV. Model Specification

The dependent variable is clientelism. To measure it, I use the question that asks if a candidate or someone from a political party offered the respondent something, like a favor, food, or any other benefit or thing in return for her/his vote or support. Subjects could answer that this had happened often, sometimes, or never. Carreras and Irepoglu (2013) and Holland and Palmer-Rubin (2015) used the same dataset and outcome variable. As they explained, the question did not ask whether respondents took the offer, hence it should not be an important source of social desirability bias (González-Ocantos et al. 2012). For statistical and substantive reasons, I dichotomized this variable, combining the alternatives often (n = 91) and sometimes (n = 150), leaving never (n = 1,196) unchanged.[21]

The following control variables were considered in the statistical analyses. Perception of corruption was included to hold constant the effect of respondents who declared clientelist activity when in reality they were referring to corruption scandals.[22] Brokers usually target civic associations. Following Holland and Palmer-Rubin (2015: 28), an additive index to measure civic participation (Political Involvement) was created.[23] Some studies have also found group size to be important (Stokes et al. 2013). A variable to measure population size at the municipal level was constructed using Brazilian census data.

Following the convention in statistical studies of clientelism, an urban/rural dummy was also included. Some have argued that parties target their own supporters (Dixit and Londregan 1996, and Cox and Mccubbins 1986), moderate opposers (Stokes 2005), or unmobilized supporters (Nichter 2008). To keep these effects constant, a variable to capture party identification (Political Id.) was included. Higher levels of democratic support should be negatively associated with clientelism. To control for that, a variable measuring democratic support was included. González-Ocantos, Kiewiet de Jonge, and Nickerson (2014) found that schooling plays a negative role on clientelism; hence, I control for education too.

21 These numbers come from the matched dataset.
22 I thank Cesar Zucco for this suggestion.
23 This variable was constructed by adding the frequency of attendance at religious meetings, community improvement meetings, and political party meetings (variables *cp6*, *cp8* and *cp13*, respectively).

V. Functional Form

Observations are clustered on a number of important factors such as levels of municipal political competition, municipal poverty, and municipal population size. In order to account for these clustering effects, I use a "generalized estimating equations" approach. GEE were introduced by Liang and Zeger (1986) to fit clustered, repeated (that is, correlated), and panel data. This method is especially efficient when the data are binary (Hanley et al. 2003). GEE models are similar to random effects models (Gardiner, Luo, and Roman, 2009), in that they allow observations to be nested in hierarchical structures. This method requires analysts to parameterize the working correlation matrix. While Hedeker and Gibbons (2006: 139) stated that "the GEE is robust to misspecification of the correlation structure,"[24] Hardin and Hilbe (2013: 166) pointed out that "[i]f the observations are clustered (not collected over time), then [...] the exchangeable correlation structure" is the most appropriate working correlation matrix. Given that the data do not follow a panel but rather a clustered structure, the "exchangeable" correlation matrix was specified in all models.

While this method is very flexible, GEE estimates remain uninterpretable in practice (Carlin et al. 2001), making regression tables useless from a substantive standpoint. In this case, the problem is even more severe due to the interactive nature of the hypothesis being tested in this paper, which is a parameter for the multiplicative term between the variables wealth index, political competition, and density of the poor.[25] Methodologists agree about "not interpret[ing] the coefficients on the constitutive terms," as they lack substantive meaning (Brambor, Clark, and Golder 2006: 77). These problems become more complex when it comes to generalized models, as a number of challenges arise.[26] Given that cross-partial derivatives are not advisable either, simulation methods

---

24 Carlin et al. (2001: 402) argued that "[r]elatively minor differences in estimates may arise depending on how the estimating equations are weighted, in particular within the generalized estimating equation (GEE) framework." Westgate and Burchett (2017) and Gardiner, Luo, and Roman (2009, 227) made the same point.

25 Brambor, Clark, and Golder (2006: 74) offer the same advice.

26 As Ai and Norton (2003) explained, "(1) the interaction effect could be non-zero, even when the estimation says it is zero, (2) the statistical significance of the interaction effect cannot be tested with a simple t-test on the coefficient of the interaction term, (3) the interaction effect is conditional on the independent variables, [...] and (4) the interaction effect may have different signs for different values of covariates."

are required (Zelner 2009). In particular, I follow the simulation approach introduced in King, Tomz, and Wittenberg (2000). This procedure samples via simulation from the point estimates, generating a new and larger distribution. In more detail, taking the single estimated parameters (that is, the regression coefficients), I constructed a distribution of estimated values for each coefficient. Relying on the central limit theorem, with enough sampling draws, the new simulated distribution is a transformation that approximates with a great degree of precision the (uninterpretable) coefficients. Subsequently, means and uncertainty measures can be constructed for each of these distributions. From a substantive standpoint, simulation methods also allow for the sampling of new distributions at different values of the independent variables. This will be important in simulating the expected value of clientelism for different "profiles," such as non-poor individuals nested in high-poor dense municipalities in contexts of high political competition, among other profiles.

Since it is "impossible to evaluate conditional hypotheses using only the information provided in traditional results tables" (Brambor, Clark, and Golder 2006: 76), I have focused instead on the substantive results from the simulation methods. However, I still present the raw results in Table A3 in the Appendix.[27] Analogous to Table 1, in Figure 4 I simulate the predicted probabilities of being targeted using both the matched and weighted/GPS models. The horizontal panel depicts simulations for the "upper" ("non-poor," 75 percent) and "lower" ("poor," 25 percent) quartiles of the wealth index. In turn, the vertical panel shows the simulated values for the maximum (100 percent) and minimum (43 percent) values of the municipal opposition index. Each quadrant shows simulations for individuals nested in poor municipalities (high density of the poor), and non-poor municipalities (low density of the poor). Each profile shows two simulated probability distributions (with 95 percent confidence intervals): one for the matched sample, and one for the weighted/GPS model.[28] The idea is to show that the decision of dichotomizing the density of the poor variable at its median gives substantively

27 Table generated via the *texreg* R package. The first column shows the estimates for the matched dataset, while the second column shows the results for the GPS-weighted model. Virtually all coefficients have the same size and sign.

28 In the case of the weighted/GPS model, which does not use the dichotomized variable, I use the continuous version of the size of the poor variable, where "low density" represents the lower quartile while "high density" represents the upper quartile.

exact results than using the continuous version of that variable via the GPS analysis.

VI. Results

All quadrants in Figure 4, regardless of the approach used,[29] suggest that brokers engage in individual targeting when individuals are identifiable, and in group targeting when brokers need to rely on the spillover effects of clientelism.

In Q1, clientelism is more likely (with a 26 percent probability) in situations where non-poor individuals are nested in poor groups (i.e. where the density of the poor is "high")[30] and living in electorally contested municipalities. As I have argued, these types of individuals are still targeted because they are more identifiable. For instance, a similar individual (same quadrant) who is nested in a non-poor group ("low" density of the poor), and consequently harder to identify, has a much lower probability of being targeted (7 percent). Similarly, individuals in Q4, such as poor individuals nested in non-poor areas ("low" density of the poor), and living in lowly contested municipalities, are more likely to be targeted (13 percent) relative to harder-to-identify individuals who live in poor areas (11 percent). In Q1, higher levels of electoral competition put heavier pressure on brokers to mobilize more expensive ways of clientelism. These pressures decay when incumbents face lower levels of electoral contestation (Q4).

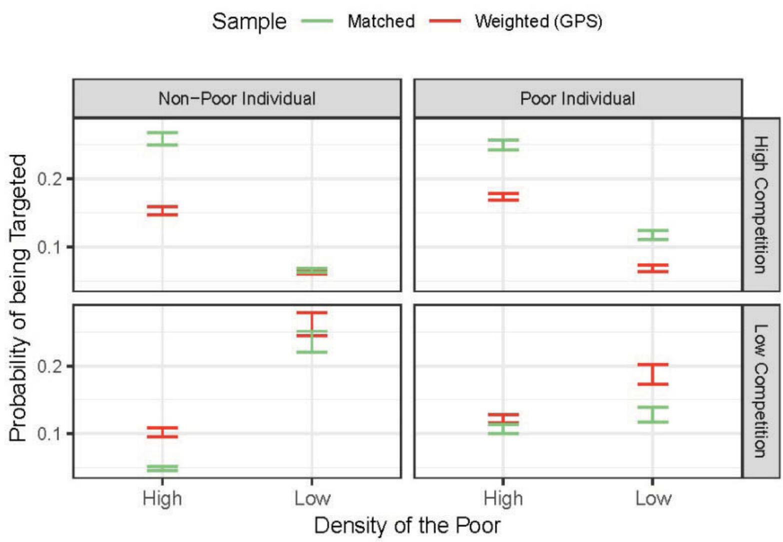
Figure 4 shows in Q2 that clientelism is more likely (25 percent) in situations where poor individuals are nested in poor groups ("high" density of the poor). As I have argued here, brokers will have incentives to engage in group targeting, taking advantage of the spillover effects of clientelism, leveraging the electoral support of potential clients by mobilizing actual clients. This is especially the case when the incumbent is seriously contested. Individuals that are similar (same quadrant), but nested in a non-poor group ("low" density of the poor), have a much lower probability of being targeted (12 percent). Individuals in Q3, who are non-poor individuals nested in non-poor areas ("low" density of the poor), and those living in lowly contested municipalities, are more likely to be targeted (24 percent) than similar individuals nested in non-poor areas (5 percent). Areas with higher levels of economic development also allow brokers to have more resources to distribute in what it was called

29 Although there are statistical differences, the differences across datasets are proportional.

30 Matched sample.

"embezzlement clientelism." Lowly contested municipalities give brokers and political incumbents more room to allocate and distribute more expensive goods. However, and as theoretically expected, given that the net costs of this form of clientelism are higher, this is the least likely form of clientelism (reflected in the lower probabilities).

Figure 4. Simulated Expected Values of Clientelism



Note: After fitting the models shown in Table A3, this figure shows the predicted probabilities of being targeted under different scenarios, with 95 percent confidence intervals. Substantively, the figure emulates the theoretical predictions shown in Table 1. Clientelism is higher when non-poor individuals are nested in poor groups ("high" density of the poor) in highly contested municipalities (Q1), when non-poor individuals are nested in non-poor groups ("low" density of the poor) in scarcely contested municipalities (Q3), when poor individuals are nested in poor areas in highly contested municipalities (Q2), and when poor individuals are nested in non-poor areas in scarcely contested municipalities (Q4). For every quadrant, estimates from both the matched and weighted datasets are shown. The idea is to show that the decision to dichotomize the density of the poor variable at its median (as shown in Figure 2) gives substantively exact results than using the continuous version of that variable via the GPS analysis.

Discussion

This paper has argued that when poor individuals live in poor areas, brokers engage in group targeting relying on the spillover effects of clientelism. This strategy mobilizes targeted and untargeted clients by disseminating the broker's reputation of delivering benefits among potential beneficiaries. In a similar way, non-poor individuals clustered in non-poor areas are also targeted. In these cases, higher levels of economic development not only raise personal incomes, but also shift the broker's vote-buying capacities upwards. Lower levels of political contestation allow these more expensive forms of clientelism. However, in heterogeneous areas, brokers adapt their strategies and execute clientelism in a different way, relying on how identifiable individuals are. Identifiability raises the cost of defection by making their households more memorable, making receivers more likely to cooperate.

Incentives to offer or take clientelist offerings are not guided solely by structural or individual factors. This paper has suggested that both are necessary to understand clientelism better. Clearly, pressures to partake in clientelism, an expensive and uncertain strategy, rise as political competition raises (from 18 percent to 25 percent).[31] However, the outcomes of this strategy differ largely depending on whether brokers face homogeneous or heterogeneous groups of individuals. Each one provides a different cost/benefit structure for both clients and brokers. Finally, I hope that the literature considers that groups and individuals provide different incentives to both brokers and clients, and hence, this distinction should be incorporated to better understand clientelism.

References

Abramo, Claudio, and Bruno Speck (2001), *Report on Brazil*, paper presented to the Global Forum II on Fighting Corruption.

Ai, Chunrong, and Edward Norton (2003), Interaction Terms in Logit and Probit Models, in: *Economics Letters*, 80, 1, 123–129.

Auyero, Javier (2000), The Logic of Clientelism in Argentina: An Ethnographic Account, in: *Latin American Research Review*, 35, 3, 55–81.

Bratton, Michael (2008), Vote Buying and Violence in Nigerian Election Campaigns, in: *Electoral Studies*, 27, 4, 621–632.

Brusco, Valeria, Marcelo Nazareno, and Susan Stokes (2004), Vote Buying in Argentina, in: *Latin American Research Review*, 39, 2, 66–88.

31 Grand mean considering the most likely scenarios only.

Calvo, Ernesto, and María Victoria Murillo (2013), When Parties Meet Voters: Assessing Political Linkages Through Partisan Networks and Distributive Expectations in Argentina and Chile, in: *Comparative Political Studies*, 46, 7, 851–882.

Calvo, Ernesto, and María Victoria Murillo (2004), Who Delivers? Partisan Clients in the Argentine Electoral Market, in: *American Journal of Political Science*, 48, 4, 742–757.

Carlin, John, Rory Wolfe, C. Hendricks Brown, and Andrew Gelman (2001), A Case Study on the Choice, Interpretation and Checking of Multilevel Models for Longitudinal Binary Outcomes, in: *Biostatistics*, 2, 4, 397–416.

Carlin, Ryan, and Mason Moseley (2015), Good Democrats, Bad Targets: Democratic Values and Clientelistic Vote Buying, in: *The Journal of Politics*, 77, 1, 14–26.

Carreras, Miguel, and Néstor Castaneda-Angarita (2014), Who Votes in Latin America? A Test of Three Theoretical Perspectives, in: *Comparative Political Studies*, 47, 8, 1079–1104.

Carreras, Miguel, and Yasemin Irepoglu (2013), Trust in Elections, Vote Buying, and Turnout in Latin America, in: *Electoral Studies*, 32, 4, 609–619.

Córdova, Abby (2008), Methodological Note: Measuring Relative Wealth Using Household Asset Indicators, in: *AmericasBarometer Insights*, 6, 1–9.

Córdova, Abby, and Mitchell Seligson (2010), Economic Shocks and Democratic Vulnerabilities in Latin America and the Caribbean, in: *Latin American Politics and Society*, 52, 02, 1–35.

Córdova, Abby, and Mitchell Seligson (2009), Economic Crisis and Democracy in Latin America, in: *PS: Political Science & Politics*, 42, 04, 673–678.

Cox, Gary, and Mathew Mccubbins (1986), Electoral Politics and Redistributive Game, in: *The Journal of Politics*, 48, 2, 370–389.

Dixit, Avinash, and John Londregan (1996), The Determinants of Success of Special Interests in Redistributive Politics, in: *The Journal of Politics*, 58, 4, 1132–1155.

Fong, Christian, Marc Ratkovic, Kosuke Imai, and Xiaolin Yang (2018), *CBPS:* Covariate *Balancing Propensity Score (CBPS) Estimation*, R Package version 2.15.0.

Gardiner, Joseph, Zhehui Luo, and Lee Anne Roman (2009), Fixed Effects, Random Effects and GEE: What Are the Differences?, in: *Statistics in Medicine*, 28, 2, 221–239.

Gay, Robert (1998), Rethinking Clientelism: Demands, Discourses and Practices in Contemporary Brazil, in: *Luso-Brazilian Review*, 36, December, 7–24.

Gay, Robert (1993), *Popular Organization and Democracy in Rio De Janeiro: A Tale of Two Favelas*, Temple University Press.

Gingerich, Daniel (2014), Brokered Politics in Brazil: An Empirical Analysis, in: *Quarterly Journal of Political Science*, 9, 3, 269–300.

Gingerich, Daniel, and Luis Fernando Medina (2013), The Endurance and Eclipse of the Controlled Vote: A Formal Model of Vote Brokerage Under the Secret Ballot, in: *Economics & Politics*, 25, 3, 453–480.

González-Ocantos, Ezequiel, Chad Kiewiet de Jonge, and David Nickerson (2014), The Conditionality of Vote-Buying Norms: Experimental Evidence from Latin America, in: *American Journal of Political Science*, 58, 1, 197–211.

González-Ocantos, Ezequiel, Chad de Jonge, Carlos Meléndez, Javier Osorio, and David Nickerson (2012), Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua, in: *American Journal of Political Science*, 56, 1, 202–217.

Guardabascio, Barbara, and Marco Ventura (2014), Estimating the Dose-Response Function through a Generalized Linear Model Approach, in: *The Stata Journal*, 14, 1, 141–158.

Hanley, James, Abdissa Negassa, Michael Edwardes, and Janet Forrester (2003), Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation, in: *American Journal of Epidemiology*, 157, 4, 364–375.

Hansen, Ben (2004), Full Matching in an Observational Study of Coaching for the SAT, in: *Journal* of *the American Statistical Association*, 99, 467, 609–618.

Hardin, James, and Joseph Hilbe (2013²), *Generalized Estimating Equations*, Boca Raton, FL: CRC Press.

Hedeker, Donald, and Robert Gibbons (2006), *Longitudinal Data Analysis*, Wiley Series in Probability and Statistics, Hoboken, NJ, USA: John Wiley & Sons, Inc.

Hicken, Allen (2007), How Do Rules and Institutions Encourage Vote Buying? (Chapter 4), in: Frederic Schaffer (ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*, Boulder, Colorado: Lynne Rienner Pub., 47–60.

Hirano, Keisuke, and Guido Imbens (2004), The Propensity Score with Continuous Treatments, in: Andrew Gelman and Xiao-Li Meng (eds), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley, 73–84.

Hlavac, Marek (2015), *stargazer: Well-Formatted Regression and Summary Statistics Tables*, R package version 5.2.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart (2011), MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, in: *Journal of Statistical Software*, 42, 8, 1–28.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart (2007), Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference, in: *Political Analysis*, 15, 3, 199–236.

Holland, A. C., and Brian Palmer-Rubin (2015), Beyond the Machine: Clientelist Brokers and Interest Organizations in Latin America, in: *Comparative Political Studies*, 48, 9, 1186–1223.

Imai, Kosuke, and David van Dyk (2004), Causal Inference With General Treatment Regimes: Generalizing the Propensity Score, in: *Journal of the American Statistical Association*, 99, 467, 854–866.

Imai, Kosuke, and Marc Ratkovic (2014), Covariate Balancing Propensity Score, in: *Journal of the Royal Statistical Society*, 76, 1, 243–263.

Imbens, Guido (2004), Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review, in: *The Review of Economics and Statistics*, 86, 1, 4–29.

Khemani, Stuti (2015), Buying Votes Versus Supplying Public Services: Political Incentives to Under-Invest in Pro-Poor Policies, in: *Journal of Development Economics*, 117, 84–93.

King, Gary, Michael Tomz, and Jason Wittenberg (2000), Making the Most of Statistical Analyses: Improving Interpretation and Presentation, in: *American Journal of Political Science*, 44, 2, 341–355.

King, Gary, and Langche Zeng (2005), The Dangers of Extreme Counterfactuals, in: *Political Analysis*, 14, 131–159.

Kitschelt, Herbert (2000), Linkages between Citizens and Politicians in Democratic Polities, in: *Comparative Political Studies*, 33, 6-7, 845–879.

Kitschelt, Herbert, and Melina Altamirano (2015), Clientelism in Latin America Effort and Effectiveness, in: Ryan Carlin and Matthew Singer (eds), *The Latin American Voter: Pursuing Representation and Accountability in Challenging Contexts*, Ann Arbor: University of Michigan Press, Chapter 10.

Kitschelt, Herbert, and Steven Wilkinson (eds) (2006), *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition*, Cambridge: Cambridge University Press.

Liang, Kung-Yee, and Scott Zeger (1986), Longitudinal Data Analysis Using Generalized Linear Models, in: *Biometrika*, 73, 1, 13–22.

Luna, Juan Pablo, Pilar Giannini, Héctor Bahamonde, Rodolfo López, Martín Ordóñez, and Gonzalo Recart (2011), El Secreto de mi Éxito: Parte II. Los Caminos a Vaparaíso en 2009, in: *Revista de Ciencia Política*, 31, 2, 285–310.

Magaloni, Beatriz (2008), *Voting for Autocracy: Hegemonic Party Survival and its Demise in Mexico*, Cambridge: Cambridge University Press.

Nazareno, Marcelo, Valeria Brusco, and Susan Stokes (2008), *Why Do Clientelist Parties Target the Poor?*, online: <http://paperroom.ipsa.org/papers/paper_2443.pdf> (11 July 2018).

Nichter, Simeon (2014), Conceptualizing Vote Buying, in: *Electoral Studies*, 35, 315–327.

Nichter, Simeon (2008), Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot, in: *American Political Science Review*, 102, 01, 19–31.

Rosenbaum, Paul (2010²), *Design of Observational Studies*, Springer Series in Statistics, Springer.

Rueda, Miguel (2015), Buying Votes with Imperfect Local Knowledge and a Secret Ballot, in: *Journal of Theoretical Politics*, 27, 3, 428–456.

Rueda, Miguel (2017), Small Aggregates, Big Manipulation: Vote Buying Enforcement and Collective Monitoring, in: *American Journal of Political Science*, 61, 1, 163–177.

Santos, Maria Emma, and Pablo Villatoro (2018), A Multidimensional Poverty Index for Latin America, in: *Review of Income and Wealth*, 64, 1, 52–82.

Schaffer, Joby, and Andy Baker (2015), Clientelism as Persuasion-Buying, in: *Comparative Political Studies*, 48, 9, 1093–1126.

Schaffer, Frederic (2004), *Vote Buying in East Asia*, Unpublished Manuscript.

Scott, James (1972), Patron-Client Politics and Political Change in Southeast Asia, in: *The American Political Science Review*, 66, 1, 91–113.

Speck, Bruno, and Claudio Abramo (2001), Transparencia Brasil/IBOPE Survey - Summary Report, in: *Transparencia Brasil*, 1–6.

Stokes, Susan (2005), Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina, in: *American Political Science Review*, 99, 3, 315–325.

Stokes, Susan, Thad Dunning, Marcelo Nazareno, and Valeria Brusco (2013), *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*, Cambridge: Cambridge University Press.

Szwarcberg, Mariela (2013), The Microfundations of Political Clientelism. Lessons from the Argentine Case, in: *Latin American Research Review*, 48, 2, 32–54.

The Latin American Public Opinion Project (LAPOP) (2010), *The AmericasBarometer*.

Vicente, Pedro (2014), Is Vote Buying Effective? Evidence from a Field Experiment in West Africa, in: *The Economic Journal*, 124, 574, F356–F387.

Weitz-Shapiro, Rebecca (2014), *Curbing Clientelism in Argentina: Politics, Poverty, and Social Policy*, Cambridge: Cambridge University Press.

Weitz-Shapiro, Rebecca (2012), What Wins Votes: Why Some Politicians Opt Out of Clientelism, in: *American Journal of Political Science*, 56, 3, 568–583.

Westgate, Philip, and Woodrow Burchett (2017), A Comparison of Correlation Structure Selection Penalties for Generalized Estimating Equations, in: *The American Statistician*, 71, 4, 344–353.

Zarazaga, Rodrigo (2016), Party Machines and Voter-Customized Rewards Strategies, in: *Journal of Theoretical Politics*, 28, 4, 678–701.

Zarazaga, Rodrigo (2014), Brokers Beyond Clientelism: A New Perspective Through the Argentine Case, in: *Latin American Politics and Society*, 56, 03, 23–45.

Zelner, Bennet (2009), Research Notes and Commentaries: Using Simulation to Interpret Results from Logit, Probit, and other Nonlinear Models, in: *Strategic Management Journal*, 30, 12, 1335–1348.

Appendix

Table A1. Summary Statistics: Raw Sample

	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Clientelism	1,483	0.171	0.376	0	0	0	1
Wealth Index	1,483	-1.543	0.846	-3.05	-2.261	-0.843	0.899
Municipal Opposition	1,483	81.761	11.821	43	75	89	100
Density of the Poor	1,483	2.435	1.12	1	1	3	4
Municipal Population	1,483	5.393	2.841	1	3	8	10
Urban	1,483	0.86	0.347	0	1	1	1
Political Involvement Index	1,483	1.792	1.619	0	0	3	9
Support for Democracy	1,483	5.426	1.682	1	4	7	7
Party Id.	1,483	5.939	1.15	1	6	6	12
Perception of Corrup-tion	1,483	2.027	1.003	0	1	3	3
Years of Education	1,483	9.398	3.857	1	6	12	18

Table A2. Summary Statistics: Matched Sample

	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Clientelism	1,437	0.168	0.374	0	0	0	1
Wealth Index	1,437	-1.557	0.811	-3.05	-2.261	-0.866	0.899
Municipal Opposition	1,437	81.912	11.749	43	75	89	100
High Density of the Poor	1,437	0.47	0.499	0	0	1	1
Municipal Population	1,437	5.384	2.792	1	3	8	10
Urban	1,437	0.86	0.347	0	1	1	1
Political Involvement Index	1,437	1.784	1.613	0	0	3	9
Support for Democracy	1,437	5.417	1.684	1	4	7	7
Party Id.	1,437	5.934	1.16	1	6	6	12
Perception of Corruption	1,437	2.029	1	0	1	3	3
Years of Education	1,437	9.359	3.843	1	6	12	18

Table A3. Generalized Estimating Logistic Equations: Clientelism

	Matched	Weighted
(Intercept)	1.404	2.958
	(1.968)	(2.691)
Wealth Index	1.374	1.320
	(0.990)	(1.209)
Municipal Opposition	-0.040	-0.061
	(0.025)	(0.032)
High Poor Density	-6.550**	
	(2.399)	
Municipal Population	-0.115*	-0.101
	(0.048)	(0.053)
Urban	-0.091	-0.077
	(0.401)	(0.416)
Political Involvement	0.046	0.047
	(0.055)	(0.055)
Support for Democracy	-0.056	-0.051
	(0.046)	(0.048)
Party Id.	-0.082	-0.087
	(0.053)	(0.052)
Perception of Corruption	0.240**	0.267**
	(0.088)	(0.089)
Years of Education	0.051*	0.054**
	(0.021)	(0.020)
Wealth Index * Municipal Opposition	-0.018	-0.013
	(0.013)	(0.015)
Wealth Index * High Poor Density	-2.509	
	(1.319)	
Municipal Opposition * High Poor Density	0.085**	
	(0.030)	
Wealth Index * Municipal Opposition * High Poor Density	0.029	
	(0.016)	
Density of the Poor		-1.992*
		(0.921)
Wealth Index * Density of the Poor		-0.555
		(0.372)
Municipal Opposition * Density of the Poor		0.024*
		(0.011)
Wealth Index * Municipal Opposition * Density of the Poor		0.005
		(0.004)
Num. obs.	1,437	1,483
Num. clust.	54	54

Note: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Clustered standard errors at the municipality level. First column shows the estimates using the matched dataset. Second column shows the estimates of the weighted model (the generalized propensity score was omitted in the table). Both models are logit GEE.

Figure A1. Distribution of Pre- and Post-Matching Observations by Wealth Index and Density of the Poor

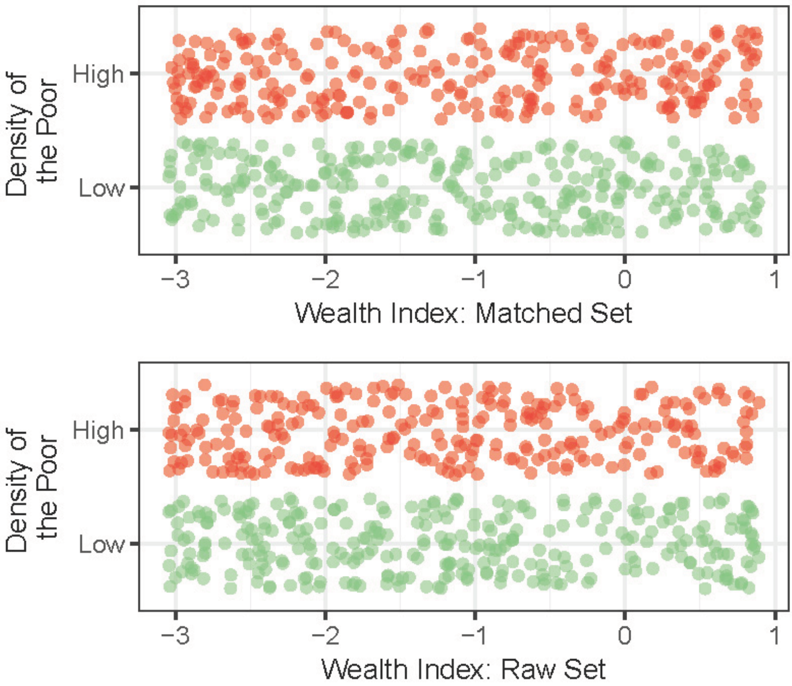


Figure A2. Frequency of Individuals by Municipality, Pre- and Post-Matching Deletion

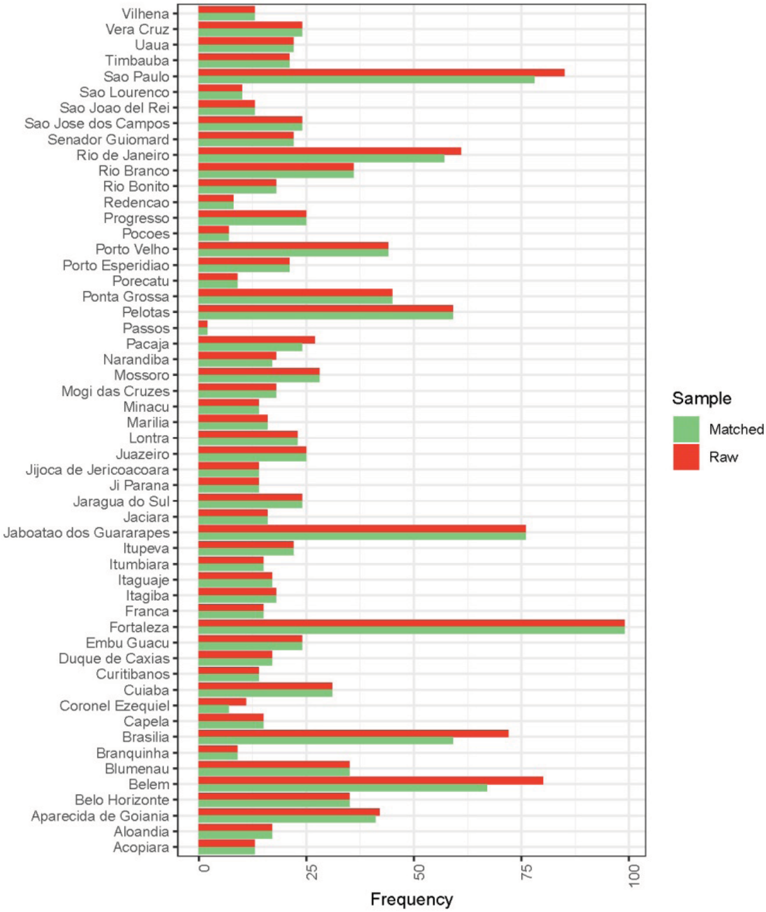


Figure A3. Pre- and Post-Matching Balance: Distribution of Propensity Scores

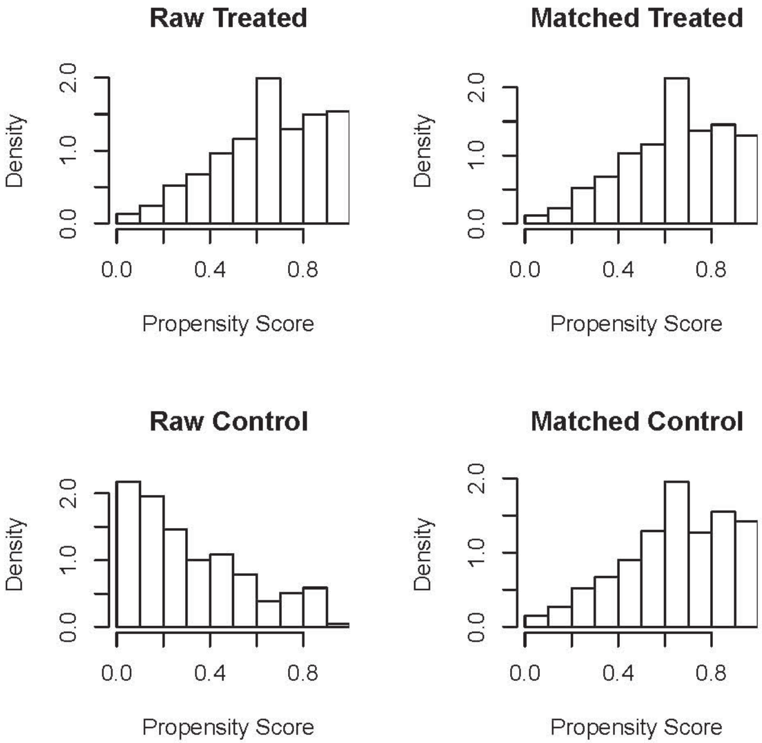
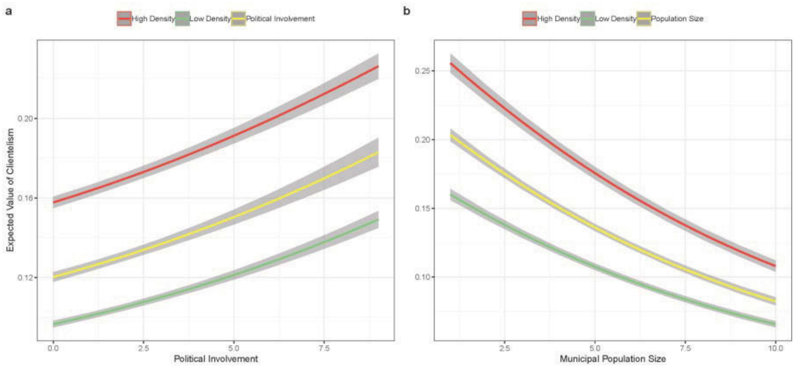


Figure A4. Simulated Expected Probability of Being Targeted: Political Involvement and Population Size



Note: Using the estimations in Table A3, the figure shows the probability of being targeted at different values of political involvement (**a**) and population size at the municipal level (**b**). The figure suggests that being nested in high-poor density areas contributes substantially more to explaining clientelism.

Figure A4 shows a plot divided in two panels. Panel **a** shows the simulated expected probabilities (with 95 percent confidence intervals) of being targeted at different levels of political involvement. As the blue lines suggests, individuals who participate in civic associations have higher probabilities of being targeted. This is in line with findings in previous research (Schaffer and Baker 2015; Carreras and Castaneda-Angarita 2014: 7; Calvo and Murillo 2013; Holland and Palmer-Rubin 2015: 16; and Rueda 2015). However, once I decompose these effects, being nested in high-poor density areas contributes substantially more to the model. These differences are statistically significant. Panel **b** shows the probability (with 95 percent confidence intervals) of being targeted at different increments of the size of the population. In line with the literature, I also see that this relationship is negative (Stokes 2005: 323; Kitschelt and Wilkinson 2006: 10; Magaloni 2008: 67; Rueda 2017; Bratton 2008; and Gingerich and Medina 2013: 456). However, the effect of being nested in high-poor density municipalities outperforms the effect of population size, suggesting spillover effects.

Apuntando Justo a Ti/Ustedes: Blancos Clientelares Grupales e Individuales en Brasil

Resumen: ¿Los partidos apuntan a grupos o individuos? Aunque esta pregunta es fundamental para entender el clientelismo, la literatura no ofrece una respuesta clara. Este trabajo argumenta que, dependiendo de ciertas condiciones, los compradores de votos apuntan a individuos cuando pueden identificar a sus blancos, y a grupos cuando necesitan utilizar los efectos indirectos que provee la lógica del clientelismo. Tanto la identificación individual como los efectos indirectos del clientelismo grupal, dependen de los niveles de pobreza individual, pobreza grupal, y los niveles de competencia partidista. Aunque la teoría de este trabajo se concentra en los blancos clientelares (grupales e individuales), también argumenta que factores estructurales, como la densidad de pobreza, deberían ser considerados en la literatura acerca de la venta de votos. Estos factores estructurales son de los pocos observables sobre los cuales los compradores de votos basan su decisión acerca de si invertir en clientelismo o no. Usando datos de opinión pública y censos de Brasil, el trabajo examina las variaciones en rentas individuales dentro de diferentes contextos de pobreza a nivel municipal. Los resultados sugieren que los partidos políticos emplean estrategias segmentadas o ad-hoc, apuntando a individuos cuando son altamente identificables, y a grupos cuando se presentan situaciones de economías de escala. Además, individuos que no están en situación de pobreza también pueden recibir ofertas clientelares.

Palabras clave: Brasil, clientelismo, venta de votos

Received: 14 May 2020 | Revised: 10 August 2020 | Accepted: 23 August 2020

DOI: 10.1111/rsp3.12337

ORIGINAL ARTICLE

Employment effects of COVID-19 across Chilean regions: An application of the translog cost function

Félix Modrego[1] | Andrea Canales[1,2] | Héctor Bahamonde[1]

[1]Institute of Social Sciences, Universidad de O'Higgins, Rancagua, Chile

[2]Instituto de Economía, Pontificia Universidad Católica de Chile, Santiago, Chile

Correspondence
Félix Modrego, Institute of Social Sciences, Universidad de O'Higgins, Libertador Bernardo O'Higgins 611, Rancagua 2820000, Chile.
Email: felix.modrego@uoh.cl

Funding information
Chilean National Agency of Research and Development (ANID); Institute for Research in Market Imperfections and Public Policy, Grant/Award Number: MIPP, ICM IS130002; FONDECYT Postdoctoral, Grant/Award Number: 3200650; FONDECYT Initiation in Research Fund 2019, Grant/Award Number: 11190112

Abstract

Estimating an aggregated translog cost function for the period 2013–2018, and using alternative scenarios of product loss based on expert projections, this article provides a preliminary forecast of the regional employment effects of COVID-19 across Chilean regions. The total estimated loss in the average scenario was around 705,000 jobs (577,000 in the optimistic and 870,000 in the pessimistic scenarios). Relative impacts were spatially heterogeneous, ranging from 1.5% (Antofagasta Region) to 13.6% (Los Lagos Region) of total regional jobs in the average scenario. Estimated impacts may inform regionally-targeted social protection and economic stimulus policies at a time in which the virus has not fully spread and total regional employment impacts have not been realized. In any region and scenario, estimated losses were sizeable and call for rapid and spread implementation of job and production protection initiatives recently passed as well as others still being discussed in congress.

KEYWORDS

Chile, COVID-19, forecasting, regional employment, translog cost function

JEL CLASSIFICATION

I15; J23; R11

1 | INTRODUCTION

COVID-19 is considered the greatest threat to world health since the Spanish Flu of 1918,[1] and the economic impacts are expected to be comparable in magnitude only to those of the 1929 Great Depression (IMF, 2020). Strict policies have been implemented by most governments to contain the spread of the disease during what has been called the "Great Lockdown" (IMF, 2020). As a result, around half of the world population is currently under confinement.[2] Restrictions to production, consumption and trade start striking harshly the global and national economies. For instance, the US Federal reserve reported a 4.8% reduction in the GDP in the first quarter of 2020. China's GDP, shrank by 6.8%, an outcome that has not been seen in 40 years. In addition, Italy and France fell into a technical recession in the first months of 2020. International trade is expected to decrease between 13 to 32% in 2020 (WTO, 2020). Consequently, the global employment impacts of COVID-19 are expected to encompass around 195 million jobs lost (UN News, 2020).

Several studies have reported that economic recessions are usually characterized by high rates of job destruction (Darby, Haltiwanger, & Plant, 1985, 1986; Davis & Haltiwanger, 1990, 1992). However, the COVID-19 crisis is showing some peculiarities in the way that economic impacts unfold, with the risk of turning a transient shock into one with relatively permanent consequences. Guerrieri, Lorenzoni, Straub, and Werning (2020) observed that in the case of the COVID-19 crisis, the demand may overreact to the supply shock, since the shock can be amplified by mechanisms such as firms' closures and job destruction, thus aggravating the recession. In this context, standard fiscal stimuli may be less effective, which warrants the implementation of large-scale social protection policies. Because these market adjustment mechanisms may operate differently (or at least at a different pace) across regions, it is important to gather estimates of regional job losses to inform regional support strategies. However, the economic effects of the pandemic are currently ongoing and still very few estimations of its sub-national economic effects exist. This paper contributes to filling this gap in knowledge by estimating the employment effects of COVID-19 in Chile, an upper middle-income economy in Latin America with large regional inequalities.

From a regional point of view, Chile is an interesting case to estimate the employment effects of COVID-19. Chile is currently organized administratively in 16 regions, 56 provinces and 346 municipalities (*comunas*). Chilean regions are arranged one on top of each other from north to south (Figure 1). In this paper, we consider the 15 regions that existed up to the end of 2018. [3] The population of Chile is highly concentrated in the metropolitan region of Santiago, Chile's primate city (with around 40% of the national population) and the economic and political core of the country. However, Chile's regional labour markets are extremely heterogeneous. Although most jobs across all Chilean regions constitute retail and unsophisticated services, there are significant variations in the relative importance of the different industries, which are directly related to Chile's physical geography. In the sparsely populated regions of the north, for instance, a large percentage of employees work in mining and mining-related services. In the more densely populated regions in the centre and south of the country, in contrast, agriculture and the agri-food industry account for a relatively large share of the labour force (Olfert et al., 2014). The metropolitan region of Santiago is a largely urban economy, oriented to services and with a much more diverse industrial structure. We approached such heterogeneity by following a regional approach to the estimation of unemployment effects of COVID-19.

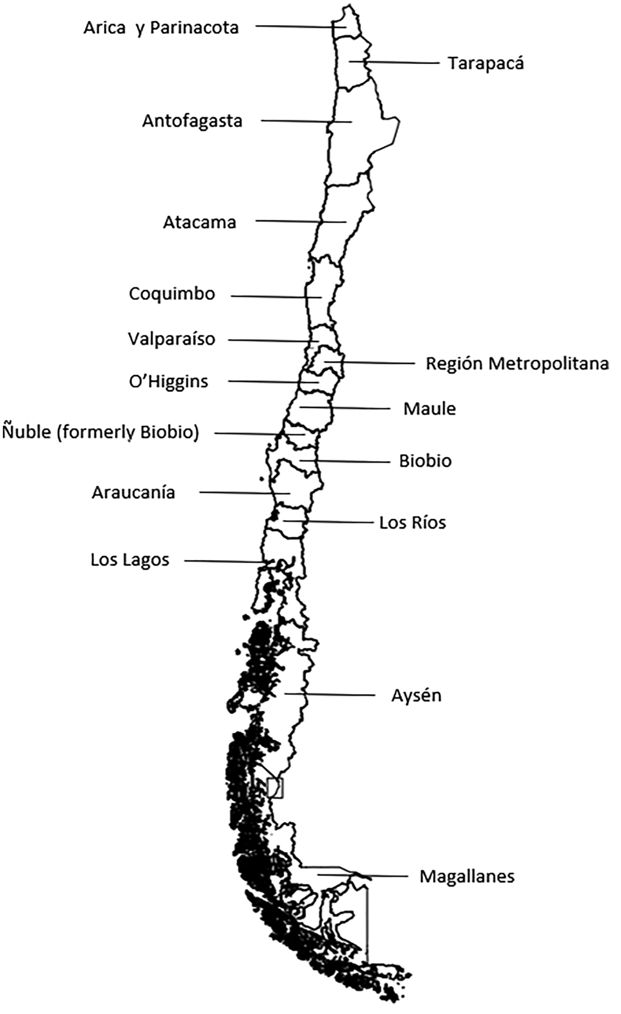
This estimation is based on fitting an aggregated cost function using data from the Chilean regions for the period of 2013–2018. We used a flexible functional form (translog) fitted using regional data from several official sources. Despite data limitations, the estimation results are largely consistent with theory. Estimated (region-specific)

[1]See https://www.telegraph.co.uk/news/2020/03/06/coronavirus-poses-serious-threat-public-health-since-spanish/

[2]See: https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou

[3]The region of Ñuble was officially established in September 2018. In all calculations, Ñuble was treated as part of the former region of Biobio from which it originates.

FIGURE 1 Chilean regions



elasticities were used to forecast the impact of Covid-19 on employment across Chilean regions due to product losses, using available expert forecasts to define three scenarios of national product loss. Results indicate a total loss of around 705,000 jobs in the average scenario. In the optimistic scenario, the estimated loss is around 577,000, and 870,000 is estimated in the pessimistic scenario. The approach followed here is based on dual production theory and is similar to the method used by Anríquez and López (2007) to study the employment and wage effects of agricultural expansion in Chile during the 1990s. This approach is grounded on well-established results in microeconomics, and is not particularly data-intensive. However, despite its strengths and applicability, we are unaware of its use for studying the regional employment impacts of COVID-19 (or other public health crises). Thus, this method may provide an econometric alternative to early estimations based on assessments of occupations at risk (Lund, Ellingrud, Hancock, Manyika, & Dua, 2020; Muro, Maxim, & Whiton, 2020).

As expected, the bulk (32%) of predicted job losses was estimated in the metropolitan region of Santiago. However, this is not the region with the largest estimated impacts in relative terms. The results indicated that the relative losses ranged from around 1.5% of regional jobs in the region of Antofagasta (a mining region in the north) to around 13.6% in the region of Los Lagos (a region in the south, mainly oriented to agriculture, aquaculture and the downstream agri-food industry). In any scenario and region, estimated impacts were considerable; therefore, the implementation of determined and widespread policies for the protection of employment and production is well warranted.

The scope of this paper is admittedly limited since it does not provide explanations of the mechanisms by which COVID-19 impacts regional labour markets nor does it compare our forecasts against others obtained using alternative methods. However, this paper contributes to the body of literature on the economic impacts of COVID-19 in several ways. First, it proposes a theory-grounded, regression-based method to account for the employment effects of the disease, a method which is feasible to implement with regional data usually available. Second, it provides statistical inferences at the subnational level, which we think are useful for Chilean regional scientists and national and regional decision makers concerned with the labour impacts of the pandemic. Third, in assessing our estimations, we observed that in the first months of the pandemic in Chile, massive job losses coincided with many workers (many of them informal) becoming inactive. This phenomenon is itself intellectually interesting and politically relevant. While skilled workers are likely to cope better with the pandemic - for instance, by offering their services from home - unskilled workers lack what has been a traditional unemployment-buffering mechanism. This is an important phenomenon to monitor in the different regions of the country. To the best of our knowledge, we are unaware of previous work noting this adjustment in Chilean labour markets in the present COVID-19 context.

The following section presents the method used for estimating the employment effects of COVID-19 in Chilean regions. Section 3 presents the results, and the final section concludes this work.

2 | METHOD

The method for estimating the employment impacts of COVID-19 across Chilean regions is based on dual production theory (Chambers, 1988; Diewert, 1974). We fit a system of equations comprised of an aggregated cost function and the corresponding factor demand equations, using aggregated data from Chilean regions for the period of 2013–2018. Estimated output elasticities were used to calculate job losses in alternative scenarios, which were defined according to expert forecasts of product loss for the Chilean economy due to COVID-19. Product losses were allocated to regions using the initial sectorial impacts in the metropolitan region of Santiago and the regional shares in each sector.

The econometric model uses a translog specification, a flexible functional form which is a second-order approximation of an arbitrary cost function. The properties of the translog cost function have been previously described in the literature (e.g., Berndt, 1991; Christensen, Jorgenson, & Lau, 1971). The use of a flexible functional form has the advantage of, first, acknowledging the substitution relationships among production factors while imposing minimum restrictions to the underlying production technology. Second, it allows for imposing parametric restrictions which are consistent with standard production theory. While the econometric approach proposed here has several advantages, it also has some limitations. The translog cost function is based on a local approximation (i.e. within the neighbourhood of a point) and does not guarantee good "regional" (that is, in a wider region of the factor price space) approximations nor global satisfaction of the regularity conditions (Pollak, Sickles, & Wales, 1984). Moreover, at the cost of flexibility, many parameters must be estimated, which may reduce statistical power when performing inference in small samples like ours (Finch & Finch, 2017). However, we believe this method is functional for the problem at hand, particularly considering the still limited information available on the regional labour impacts of the pandemic.

The starting point is the following translog specification with technological change proposed by Diewert and Wales (1987):

$$
\begin{aligned}
lnC(p,y,t) = \alpha_0 + \sum_i \alpha_i lnp_i + \alpha_y lny + \alpha_t (t-t^*) + \left(\frac{1}{2}\right)\sum_i \sum_j \alpha_{ij} lnp_i lnp_j + \sum_i \alpha_{iy} lnp_i lny \\
+ \sum_i \alpha_{it}(t-t^*) lnp_i + \alpha_{yt}(t-t^*) lny + \frac{1}{2} a_{tt} (t-t^*)^2,
\end{aligned} \tag{1}
$$

where C is the production cost, i,j index factors, p is the factor's price, y is the output, t is the period and t^* is a reference year, such that $t-t^*$ is a measure of technological change. Homogeneity of degree one in factor prices is imposed through the following parametric restrictions in Equation 1 (Ryan & Wales, 2000):

$$\sum_i \alpha_i = 1; \sum_i \alpha_{iy} = 0; \sum_i \alpha_{it} = 0; \sum_j \alpha_{ij} = 0 \text{ for the } i = 1,\ldots,n \text{ factors}. \tag{2}$$

By Shepard's lemma, one arrives at the conditioned factor demands, as cost shares (s), differentiating 1 with respect to factor prices. For each factor i:

$$s_i(p,y,t) = \alpha_i + \sum_j \alpha_{ij} lnp_j + \alpha_{iy} lny + \alpha_{it}(t-t^*), \quad \forall i. \tag{3}$$

The theoretical restriction of symmetry of the second-order derivatives of the cost function (Young's theorem) is imposed by the following restrictions to the cross-price coefficients in the n factor share equations 3:

$$\alpha_{ij} = \alpha_{ji}, \ \forall i,j. \tag{4}$$

Although Equation 1 contains all the parameters in 3, the system 1–3 (with parametric restrictions 2 and 4) is jointly estimated, in order to increase the statistical efficiency of the estimates (Hussain & Bernard, 2018). Considering the available data, our estimation includes three factors: labour, machinery, and construction. However, of the three share equations in 3, we only estimated the labour and machinery equations, to avoid a singularity problem. System 1–3 is estimated using Zellner's (1962) seemingly unrelated regressions in its iterative version (IT-SUR), which ensures that the arbitrary decision of which equation to exclude carries no consequences in the estimation results (Baum & Linz, 2009).

A check of global concavity of the cost function in each sample point was performed using the approach proposed by Diewert and Wales (1987), and implemented in the Stata software by Baum and Linz (2009). Diewert and Wales (1987) showed that the cost function will be (quasi)concave if the matrix M:

$$M = H - S^k + SS', \tag{5}$$

is negative (semi)definite. This will depend on all the eigenvalues (λ**)** being (non-negative) positive. In 5, H is the Hessian matrix, S is the shares matrix and S^k is a diagonal matrix of shares, all of order $k \times k$, with k being the number of factors in the cost function.

Estimated coefficients of system 1–3 were used to calculate changes in total costs associated with changes in output. Total labour costs (average wages times the number of workers) (C_l) equals:

$$C_l = C * s_l. \tag{6}$$

Taking logarithms in 6 and differentiating both sides with respect to the output, after some manipulations, the change in total labour cost (dCl) in region r as a function of estimated coefficients is:

$$\mathrm{d}C_{lr} = C_{lr} * dlny_r\left(\sigma_{Cyr} + \sigma_{sly}/s_{lr}\right), \tag{7}$$

where $\sigma_{Cyr} \equiv \left(\frac{dlnC}{dlny}\right)_r$ is the output elasticity of total costs obtained from the estimation of 1, and $\sigma_{sly} \equiv \frac{dsl}{dlny} \equiv a_{ly}$ is the output semi-elasticity of the labour share in total costs. It worth noting that since it is dependent on the values of the covariates, the output elasticity of total cost (σ_{Cyr}) in Equation 9 is specific for each observation, while the output

TABLE 1 The data

Variable in the model	Empirical variable	Source
y (output)	Regional GDP (thousands of millions Ch.$)	Central Bank of Chile
p_l (price of labour)	Average income of employed people in the region as an index (base 100 = metropolitan region in 2013)	National Institute of Statistics, national employment survey (ENE), (Nov.–Jan. moving quarter).
p_k (price of machinery)	Capital goods imports price index (national) (base 100 = 2013)	Central Bank of Chile
p_c (price of constructions)	Deflator of fixed capital consumption in the construction sector (national) (base 100 = 2013)	Central Bank of Chile
C_l (labour costs)	$Cl = p_l$ x number of employed people in the region (thousands of millions Ch.$)	Employed people: National Institute of Statistics, national employment survey (ENE), (Nov-Jan. moving quarter).
C_k (machinery costs)	Fixed capital consumption in machinery and equipment (thousands of millions Ch.$). Allocated to regions using the share of each region in total capital.	Fixed capital consumption in machinery and equipment: Central Bank of Chile. Regional share in total capital estimated by Cerda (2018).
C_c (construction costs)	Fixed capital consumption in construction (housing plus rest of construction) (thousands of millions Ch.$). Allocated to regions using the share of each region in total construction workers each year.	Fixed capital consumption in construction (housing plus rest of construction): Central Bank of Chile. Regional share o in total construction workers each year.: National Institute of Statistics, national employment survey (ENE), (Nov.–Jan. moving quarter).
C (total cost)	$C_{l\ +} C_k + C_c$	
s_i	Share of factor i in total cost: $s_i = C_i/C$	

elasticity of the labour share (σ_{sly}) is constant across the entire sample space. From Equation 7, it is clear that the change in labour triggered by a change in output includes the effect of output changes on total regional cost (σ_{Cyr}) and factor substitution effects due to changes in the production scale (σ_{sly}).

Changes in total labour costs due to changes in output are then converted into changes in jobs (dl) using the average incomes of workers in each region (p_{lr}):

$$dl_r = \frac{dC_{lr}}{p_{lr}}. \tag{8}$$

The system 1–3 is fitted using annual data from the 15 Chilean regions existing up to 2018 and for the period 2013–2018. Regional GDP as chained volume (based on the 2013 Chilean input–output matrix) in thousands of millions Ch.$ (variable y) is reported by the national accounts system managed by the Chilean Central Bank. Fixed capital consumption in machinery (in thousands of millions Ch.$) is used as a proxy of machinery costs (C_k) and was retrieved from the same source. The fixed capital consumption in machinery was allocated to regions using regional shares in total capital calculated by Cerda (2018). We used the price index of capital goods imports (base 2013 = 100) as a proxy of the price of machinery (p_k). This price index is also reported by the Chilean Central Bank. This variable does not have regional variation, which can introduce some measurement errors. However, as argued by Anríquez and López (2007), the mobility of capital goods across regions is subject to few frictions due to a well-integrated capital market in Chile; thus, the price of capital should not have major regional differences. Fixed capital consumption in construction (housing

TABLE 2 Expert forecasts of product growth for the Chilean economy in 2020

Source	Forecast end of 2019 (% of GDP) (1)	Forecast in June 2020 (% of GDP) (2)	Impact of COVID-19 (% of GDP) (1)–(2)
World Bank	2.5	−4.3	−6.8
OECD	2.4	−5.6	−8
OECD 2/1	2.4	−7.1	−9.5
ECLAC	1	−5.3	−6.3
Average			−7.7

Note: /1 OECD refers to the forecast with only one virus outbreak and OECD2 with a second.

plus the rest of construction, in thousands of millions of Ch.$) was used as the construction costs variable (C_c),and was also taken from the national accounts system. The capital consumption in construction was allocated to regions using the share of each region in total construction workers each year, with the share calculated using the National Employment Survey (ENE, mobile quarter November–January) run by the Chilean National Institute of Statistics (INE). For the price of construction (p_c), we used the deflator of fixed capital consumption in the construction sector in the national accounts system (base 100 = 2013), also unavailable for regions. This is another source of measurement error, which is possibly more important due to, for instance, regional differences in land prices. These are differences that we cannot control with the data at hand. The average income of workers in each region each year was obtained from the INE's Supplementary Survey of Incomes (ESI), an extra module of the ENE survey added each year in the mobile quarter October–December. The workers' average income was multiplied by the regional number of workers to calculate the regional labour costs (C_l) each year. The workers' mean income was turned into an index (base 100 = the metropolitan region of Santiago in 2013) to use in the regressions as the labour price variable (p_l). Total regional costs were obtained adding labour, machinery and construction costs; subsequently, the share of each factor (s_{lr}, s_{kr}, s_{cr}) was computed. A summary of the variables used in the econometric estimation can be found in Table 1.[4]

We defined three scenarios of product loss following the most recent expert forecasts for the Chilean economy (World Bank, OECD and United Nations' ECLAC). The product impact of COVID-19 was defined as the difference between the national product change forecasted before the outbreak of the virus (December 2019–January 2020) and after the outbreak (first days of June 2020). The *average* scenario was defined as an impact equal to the average of the four forecasts and amounts to 7.7%. The *pessimistic* scenario was defined according to the largest impact, the OECD forecast with a second COVID-19 outbreak, which amounts to a 9.5% reduction. The *optimistic* scenario corresponds to the mildest product impact, by UN's ECLAC, which indicates a shrinkage of around 6.3%. The scenarios and the simulation parameters are summarized in Table 2.

Using the expert forecasts, the forecasted product impact in each scenario was allocated to regions using regional weights (ω_r, w_r), such that Equation 8 now reads as:

$$dl_r = C_{lr}/p_{lr} * dy * \omega_r/(y * w_r)\left(\sigma_{Cyr} + \sigma_{sly}/s_{lr}\right), \tag{9}$$

where y is the national product in 2019 and dy is the national product loss (in levels) for 2020 in each scenario. ω_r is a regional parameter used for allocating the forecasted national product loss in 2020 among the (former) fifteen regions, and w_r is the regional share in total product (in 2018, last year available).[5]

The ω_r parameter was calculated as:

[4]Data complied by the Chilean Central Bank was retrieved from its statistics website: https://www.bcentral.cl/areas/estadisticas. Data reported by INE was obtained from INE.stat (https://stat.ine.cl/) and INE databank (http://bancodatosene.ine.cl/) websites.

[5]Total here means the sum of the 15 regions, which is not equal to the national product, as the national GDP includes items which are not suitable for regional allocation.

TABLE 3 System (1)–(3) estimation results

Variable	C	sl	sk
lnpl	0.940*** (0.098)	–0.065*** (0.017)	0.015 (0.020)
lnpl*lnpl	–0.065*** (0.017)		
lnpk	0.060 (0.098)	0.015 (0.020)	–0.000*** (0.000)
lnpk*lnpk	–0.000*** (0.000)		
lnpc	0.000*** (0.000)	–0.052** (0.024)	–0.000 (0.000)
lnpc*lnpc	0.056*** (0.020)	–0.006 (0.026)	0.000*** (0.000)
lny	0.489** (0.202)	–0.001 (0.005)	0.005 (0.006)
lny*lny	0.020* (0.011)		
lnpl*lnpk	0.015 (0.020)		
lnpl*lnpc	–0.006 (0.026)		
lnpl*lny	–0.001 (0.005)		
lnpk*lnpc	–0.000 (0.000)		
lnpk*lny	0.005 (0.006)		
lnpc*lny	–0.005 (0.005)		
t–t*	0.110 (0.108)	0.000** (0.000)	0.000*** (0.000)
(t–t*)2	–0.013** (0.006)		
lnpl*(t–t*)	0.000*** (0.000)		
lnpk*(t–t*)	0.000*** (0.000)		
lnpc*(t–t*)	0.000*** (0.000)		
lny*(t–t*)	0.009 (0.010)		
Constant	–2.889*** (0.985)	0.940*** (0.098)	0.060 (0.098)
Observations	88	88	88
chi2 p-value	0.0000	0.0000	0.4725
Breusch–Pagan test of errors independence (p-value)	0.0000		

Note: significant at *10%, **5%, ***1%. Standard errors in parenthesis.

$$\omega_r = \sum\nolimits_s w_s^{RM} * w_{sr}. \quad (10)$$

w_s^{RM} is the contribution of each economic sector (s = 1, …,21) to total employment change observed in the metropolitan region of Santiago between February–April 2020 and February–April 2019 (same quarter to avoid seasonality problems), and w_{sr} is the share of region r in total workers in the s sector. Thus, ω_r considers the likely sectoral differences in the employment effects of the pandemic and the importance of each sector in each regional economy.[6]

The use of employment changes in the metropolitan region to calculate the ω_r parameter is grounded in the following reasons. In the period between the first draft of this paper and the revised version, the National Institute of Statistics released the results of the National Employment Survey (NES) for the moving quarter February–April 2020, which allows for assessing the initial employment effects of the virus outbreak. Currently (the first days of June 2020), the Chilean National Government has opted for a strategy of "dynamic lockdowns," in which specific areas (municipalities or specific areas within a municipality) are locked down and released based on the epidemiologic situation, an assessment which is largely based on the spatial concentration of detected COVID-19 cases. By the end of April 2020, the metropolitan region of Santiago accounted for around 80% of total cases (twice its share in the national population) and was the only Chilean region where a sizeable population went into a prolonged (more than

[6]This way of regionally allocating the impacts of COVID-19 was motivated by the comments of an anonymous reviewer to whom we are very grateful.

TABLE 4 Output elasticities of total costs in 2018 for Chilean regions

Region	σ_{Cyr}	Standard error	P-value	95% Confidence interval	
Arica y Parinacota	0.838	0.057	0.000	0.726	0.950
Tarapacá	0.884	0.046	0.000	0.794	0.973
Antofagasta	0.942	0.050	0.000	0.845	1.039
Atacama	0.882	0.046	0.000	0.792	0.972
Coquimbo	0.892	0.045	0.000	0.804	0.980
Valparaíso	0.934	0.048	0.000	0.840	1.028
Metropolitana	0.991	0.068	0.000	0.858	1.124
O'Higgins	0.911	0.045	0.000	0.823	0.998
Maule	0.899	0.044	0.000	0.812	0.986
Biobio	0.930	0.047	0.000	0.838	1.023
Araucanía	0.889	0.045	0.000	0.801	0.978
Los Ríos	0.861	0.050	0.000	0.762	0.959
Los Lagos	0.896	0.045	0.000	0.809	0.984
Aysén	0.828	0.061	0.000	0.708	0.948
Magallanes	0.854	0.052	0.000	0.752	0.956

a few weeks) lockdown. For instance, a large section of the municipality of Santiago, where the city's main business and public administration district is located, entered into lockdown on 26 March 2020, along with six other municipalities, amounting to around 1.3 million people in lockdown. By the first days of June 2020, the municipality of Santiago had not been released yet; on the contrary, the strategy of dynamic lockdowns had shifted to locking down the entire greater Santiago City (34 municipalities) and some surrounding municipalities. The Santiago metropolitan region is the core and largest region, constituting around 40% of the national population, and has the highest sectoral diversity, with all economic sectors well represented. Thus, the results of the February–April round of the NES survey for the metropolitan region of Santiago provide what is arguably the best snapshot currently available of the sectorial employment effects in Chile in the likely event of the spread of the disease within the entire country and the subsequent implementation of total regional lockdowns.

To estimate the employment effects of COVID-19 in each region, total labour costs (C_{lr}), average regional workers' income (p_{lr}), the output elasticity of total costs (σ_{Cyr}) and the share of labour in total costs (s_{lr}) were all evaluated for each region in 2018, which was the last year with all the data available. Confidence intervals for the impacts calculated with equation 9 were obtained using the delta method.

3 | RESULTS

Table 3 summarizes the estimation results of system 1–3 for a filtered sample of 88 observations.[7] The results are largely consistent with the theory. The total marginal effects indicate that costs are non-decreasing in output and factor prices. Factors shares decrease as their prices increase, and point estimates suggest substitution between

[7]The system 1–3 was initially estimated using the full sample of 90 observations (15 regions × 6 years). Estimated elasticities were used to make retrodictions of employment changes in the 2008–2013 period, using Equations 7 and 8. An analysis of the standardized residuals (i.e., the standardized difference between observed and retrodicted employment changes) revealed two clear outliers where the rule-of-thumb criterion of a standardized residual greater than two (in absolute terms) was exceeded. Such outliers were excluded in the final estimations. We are grateful to an anonymous reviewer for motivating this validation exercise.

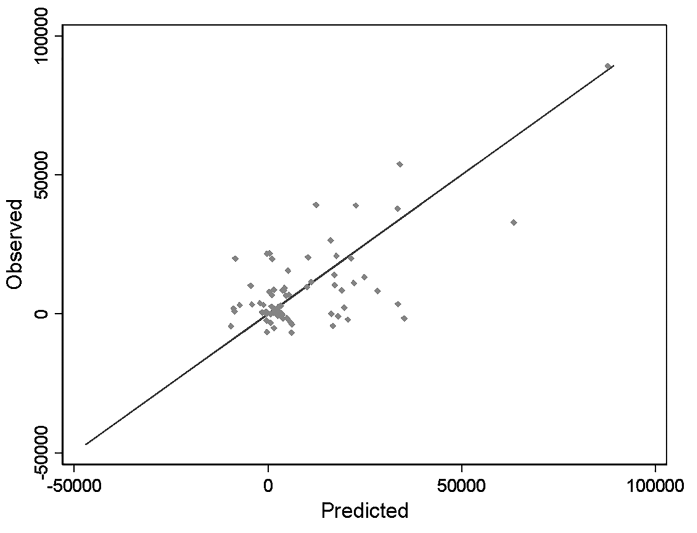


FIGURE 2 Model validation – predicted vs observed regional employment changes for the 2013–2018 period

machinery and constructions. In contrast, there is no clear substitution relationship between labour and the other production factors, although this is probably related to the small sample size and the lack of regional variation of capital and construction prices. Concavity was verified in approximately 76% of the sample. The Breusch–Pagan test confirmed the lack of independence of the errors across equations, which is expected since all derive from the same underlying technology (implicit in the cost minimization problem). Altogether, the results provide confidence in that in spite of considerable data limitations, the estimated translog cost function provides a reasonable representation of the aggregated regional production technology.

Output elasticities of total costs (σ_{Cyr}) in 2018 and their standard errors are reported for each region in Table 4. They range from 0.828 (Aysén, south end of the country) to 0.991 (metropolitan region of Santiago). At a 95% confidence, in 11 out of the 15 regions, total costs responded less than proportionally to changes in output. In particular, the northernmost region (Arica y Parinacota) and the two southernmost (Aysén and Magallanes) are estimated as the less responsive to output changes. The estimated output elasticity of the labour share (constant across the sample space) indicates, however, that the share of labour in total costs is largely invariant to the scale of production (σ_{sly} = −0.0005, not significant). This indicates that the impact of COVID-19 on employment would be mostly due to the reduction in output and would not be substantially buffered (nor amplified) by a substitution of factors triggered by the large product shrinkage. Again, we cannot exclude the possibility that the lack of significance of this second elasticity is due to the lack of regional variation in factor prices and the small sample size. As a simple validation exercise, Figure 2 displays observed against predicted employment changes for the period of 2013–2018. Predictions were obtained using the estimated elasticities in Equations 7 and 8. A 45-degree line was added to ease visualization. A simple regression indicated a R^2 of 0.57 and a slope coefficient of 0.74.[8] Overall, the model seems to capture well the fundamental underlying economic relationships, and the proposed method has been demonstrated to be useful for estimating regional employment effects of COVID-19 in Chile.

Table 5 summarizes the regional parameters used in the calculation of labour impacts with Equation 9, showing the large share (40%) of the Santiago metropolitan region in the national impact (ω_r). The calculated share is, nevertheless, lower than its share in the total product (46%). The region of Antofagasta, the largest mining region in the country, has the lowest calculated share in the national impact (2.4%) relative to its share in the total product (11%). Conversely, Biobio, a region with the third largest urban agglomeration in the country (Concepción), has a high calculated share (12%) in the national impact given its share in the total product (around

[8]The standardized residuals were all below the rule of thumb value of two in absolute terms.

TABLE 5 Regional parameters for the employment impact calculations

Region	Regional GDP 2018 (thousands of millions Ch. $)	Regional weight (ω_r)	Regional share in total product 2018 (w_r)	Total labour costs 2018 (thousands of million Ch.$)	Mean workers annual wage 2018 (thousands Ch.$)	Share of labour in total costs 2018 (%)
Arica y Parinacota	1,111.9	0.009	0.008	436.8	5,819	62.3
Tarapacá	3,433.1	0.016	0.025	1,031.6	6,138	62.8
Antofagasta	14,787.8	0.024	0.106	2,422.4	8,310	59.3
Atacama	3,323.5	0.013	0.024	942.4	6,592	63.4
Coquimbo	4,251.4	0.043	0.030	2,093.7	5,542	67.4
Valparaíso	12,135.3	0.098	0.087	5,584.0	6,589	67.4
Metropolitana	65,031.3	0.404	0.465	27,255.7	8,036	72.5
O'Higgins	6,733.1	0.052	0.048	2,659.4	5,862	70.8
Maule	4,999.9	0.067	0.036	2,692.9	5,287	66.0
Biobio	11,018.5	0.116	0.079	5,361.9	5,550	68.3
Araucanía	3,951.6	0.053	0.028	2,512.0	5,333	64.2
Los Ríos	1,947.4	0.027	0.014	1,122.5	5,868	66.3
Los Lagos	4,706.1	0.060	0.034	2,686.9	6,047	68.9
Aysén	856.1	0.008	0.006	468.3	7,494	67.7
Magallanes	1,654.8	0.010	0.012	890.7	10,134	77.2

Notes: /1 dy = – 11,961.1 (th. MM Ch. $) in the average, (th. MM Ch. $) -14,757.2 in the pessimistic and –9.786.4 (th. MM Ch. $) in the optimistic scenario. σ_{sly}= – 0.0005 for all regions. σ_{Cyr} in Table 4. /2 "Total" costs are the sum of labour, machinery and constructions costs.

TABLE 6 Estimated impacts of COVID-19 on regional employment

	Average scenario			Optimistic scenario			Pessimistic scenario		
Region	Estimate	95% confidence interval		Estimate	95% confidence interval		Estimate	95% confidence interval	
Arica y Parinacota	−5,468	−6,208	−4,727	−4,474	−5,079	−3,868	−6,746	−7,659	−5,832
Tarapacá	−7,535	−8,313	−6,758	−6,165	−6,802	−5,529	−9,297	−10,257	−8,337
Antofagasta	−4,753	−5,255	−4,251	−3,889	−4,299	−3,478	−5,864	−6,483	−5,245
Atacama	−5,303	−5,854	−4,752	−4,339	−4,789	−3,888	−6,542	−7,222	−5,863
Coquimbo	−36,561	−40,213	−32,910	−29,914	−32,901	−26,926	−45,108	−49,613	−40,603
Valparaíso	−68,895	−75,910	−61,879	−56,368	−62,109	−50,628	−85,000	−93,656	−76,344
Metropolitana	−224,874	−255,276	−194,471	−183,987	−208,862	−159,113	−277,441	−314,951	−239,932
O'Higgins	−34,320	−37,663	−30,976	−28,080	−30,816	−25,344	−42,343	−46,468	−38,217
Maule	−65,848	−72,327	−59,369	−53,876	−59,176	−48,575	−81,241	−89,234	−73,248
Biobio	−102,136	−112,385	−91,887	−83,566	−91,951	−75,180	−126,012	−138,656	−113,367
Araucanía	−59,980	−66,027	−53,934	−49,075	−54,022	−44,128	−74,002	−81,462	−66,542
Los Ríos	−24,500	−27,340	−21,659	−20,045	−22,369	−17,721	−30,227	−33,731	−26,723
Los Lagos	−54,964	−60,405	−49,523	−44,970	−49,422	−40,519	−67,812	−74,525	−61,100
Aysén	−5,279	−6,052	−4,506	−4,319	−4,952	−3,687	−6,513	−7,467	−5,559
Magallanes	−5,013	−5,623	−4,403	−4,102	−4,601	−3,603	−6,185	−6,937	−5,433
Total	−705,428			−577,168			−870,333		

8%). This is due to the specific regional employment structure, which is characterized by a relatively large weight of agriculture and manufacturing, two sectors showing large initial employment impacts. According to the February–April ENE survey, retail, lodging and restaurants and manufacturing are the sectors that have been most affected initially by the COVID-19 shock, while others sectors such as health, education or mining are far less. While high-risk sectors such as lodging and restaurants correspond to the sectors at the greatest risk in the US (Muro et al., 2020), others like mining (described as a high-risk sector in the US) are not among the industries with the highest initial job losses according to the ENE survey. This finding suggests some particularities of the sectorial employment effects of COVID-19 in Chile.

Table 6 summarizes the estimated employment losses in each region and their 95% confidence intervals and Figure 3 maps these losses as a percentage of total regional jobs. Overall, total job losses were more or less proportional to the expert forecasts of product shrinkage, since regional output elasticities were close to unity and the estimated output elasticity of the labour share in total costs was negligible. Total job losses amounted to around 705,000 in the average scenario. In the pessimistic scenario, job losses amount to nearly 870,000 and to 577,000 in the optimistic scenario. Examining the regional variation of impacts, almost a third of the national variation (nearly 225,000 jobs) was estimated in the metropolitan region of Santiago, although this figure was considerably less than its share in national employment (44%). The least impacted region in relative terms would be the region of Antofagasta, a region where mining accounts to around 54% of the regional GDP. In the average scenario, job

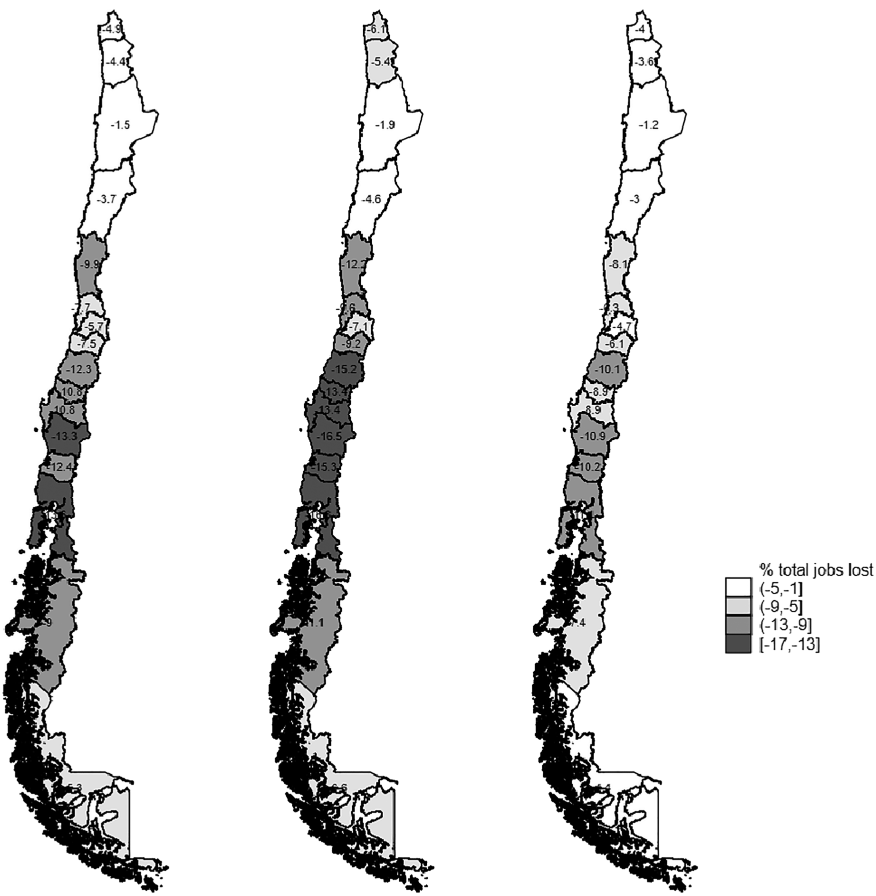


FIGURE 3 Regional employment impacts as a percentage of total jobs. Left: Average scenario. Centre: Pessimistic scenario. Right: Optimistic scenario

losses would be of around 4.800, which means around 0.7% of the national job loss. This is proportionally less than its share in the national employment (3.6%), and only a 1.5% of total regional jobs. While the expenditures in production factors have been relatively sensitive to output variations (Table 4) in Antofagasta compared to other regions, the regional employment structure is highly specialized in mining, a sector that has not been one of the most affected by the disease based on the evidence so far. Antofagasta's economy is highly sensitive to the copper price cycle (Atienza & Modrego, 2019), and the slow recovery in copper-buying countries (mainly China) will likely help mitigatethe employment effects of the pandemic. In contrast, the region of Los Lagos is the most affected region in relative terms (13.6% of total regional employment in the average scenario, Figure 3). Los Lagos is a region located in the south of the country, oriented to the production of primary natural resources and its downstream agri-food industry. Agriculture, livestock, forestry, fishing and aquaculture alone account for around 12% of the regional product and manufacturing adds another 26%. Although the region's factors expenditure had an approximately average sensitivity to output changes (Table 4), its specialization in agriculture and (agri-food) manufacturing (two of the sectors most initially affected by COVID-19) has resulted in the large forecasted employment impacts in relative terms. More generally, Figure 3 illustrates how mining regions in the north of the country (Tarapacá, Antofagasta and Atacama) are predicted to have lower relative job losses compared to agriculture and agri-food-oriented regions in the centre and centre–south (Maule, Araucanía, Los Ríos and Los Lagos).

Overall, in all regions and scenarios, the estimated regional employment effects of COVID-19 are sizeable and call for determined actions to the protect employment and production within the entire country. Since the pandemic is currently ongoing, and forecasts of economic impacts for the Chilean economy may still change, these estimations should be considered to be a preliminary approximation.

4 | CONCLUSIONS

This paper presents preliminary forecasts of employment losses in Chile at the regional level resulting from the COVID-19 pandemic. Following a methodological approach based on the estimation of an aggregated cost function for the period of 2013–2018 and on dual production theory, the employment losses were calculated for different scenarios of national product loss. The estimations suggest sizeable impacts. In the average scenario, the estimated impact was around 705,000 jobs country-wide, which amounts to 870,000 and 577,000 in the pessimistic and optimistic scenarios, respectively. In regional terms, impacts are heterogeneous across regions, with agriculture and agri-food oriented regions in the centre–south more affected than mining regions in the north.

These estimates are preliminary, and should be taken with caution. A first limitation of this study is based on the data, particularly the lack of regional data on consumption and prices of several production factors. Producing better estimates would be possible with more complete regional data, particularly data on consumption and prices of production factors. A second limitation is that we did not explicitly model the spatial spread of the disease nor the output impacts of COVID-19 in each region. Instead, we relied on expert forecasts for the national economy, which are regionally allocated in a more or less *ad-hoc* way, and implicitly assuming a uniform spread of the virus throughout the country. Finally, statistical relationships were estimated using data for a period of moderate growth of the national economy. However, estimated elasticities were used to forecast job losses in a situation of product loss not seeing during the period for which such elasticities were estimated. It is unclear whether employment adjustments in the COVID-19 context will be directly proportional to those seen in the 2013-2018 period. Indeed, the results of the ENE survey for the moving quarter February–April 2020 suggest that adjustment mechanisms that have not been previously observed, at least in the last decades, are currently taking place in the Chilean labour market. Traditionally, self-employment and informality have been powerful buffer mechanisms containing unemployment outbreaks during previous crises. During the COVID-19 crisis, an unusual rise in the number of inactive people (around 720 thousand people) has accompanied an increase in unemployment of only 0.9 percentage points between January–March and February–April 2020. At the same time, there was a large reduction in informal jobs (around 415,000). Both these observations suggest that many jobs lost are due to informal and self-employed workers exiting the labour market, a pattern not observed at such a large scale during the 2013–2018 period. This phenomenon is important itself and deserves further research and monitoring by national and regional authorities.

Based on these considerations, total job losses at the end of 2020 may exceed the present forecasts. Indeed, between the last two moving quarters, a period capturing only the initial employment effects of COVID-19, the National Employment Survey (ENE) reported a fall of around 706 thousand in the number of employed individuals already, very close to what we forecasted for the average scenario. The inactive people and the unemployed people behind these figures may amplify the recession in the sense of Guerrieri et al. (2020). If that is the case, the expert forecasts of product losses we are using may also fall short, as the downwards corrections in most expert projections between April and June suggest.

Keeping these caveats in mind, the extent of estimated impacts is large (around 700 to 870,000 jobs potentially lost), calling for resoluted actions to protect employment and production in the entire country. As a response to the virus outbreak, sizeable support packages have been implemented in Chile and others are currently being discussed. Actions already implemented include modifications to the national employment insurance to be used in situations of temporary job suspensions (thus avoiding firings), insurance supported by supplementary public funds, state-guaranteed loans with near-zero real interest rates for firms and additional cash and in-kind transfers to vulnerable households. Regarding the reformed unemployment insurance, the last ENE survey indicated an increase of around 365,000 “absent employees” (people with a job that was nevertheless not performed on-site last week), a change of a magnitude which is partly explained by the reformed employment insurance.

Regional authorities should be prepared for a rapid and effective implementation of these support initiatives, maintaining close monitoring of their communities and collaborating in targeting and implementing national support initiatives. Employers can also contribute to this effort by making a correct use of the employment protection insurance, limiting its use to cases when it is truly needed, and rapidly processing the requests to activate the insurance. Financial institutions can also play an important role in helping mitigate the economic consequences of COVID-19, by rapidly and effectively issuing state-guaranteed loans to firms that need them the most, particularly the small and medium-sized firms.

A direct extension of this work is the replication of the analysis for regional industries, which can be carried out to assess the sectoral impacts in each region, thus informing regional protection and recovery strategies. Likewise, a better understanding of the different types of contractual arrangements that are predominant in each productive sector as well as the sensitivity of jobs to economic shocks under these different contractual arrangements, can shed light on more specific employment protection policies for this crisis and other future crises. Finally, this work has not taken into consideration regional interactions which could significantly alter the estimated regional distribution of the impacts of COVID-19. These interactions may stem, for instance, from physical and human capital externalities and regional growth spillover effects (Valdez, 2019), and could be analyzed using spatial econometric methods, inter-regional input–output analysis and/or regional computable general equilibrium models. Global VAR (GVAR) models (Pesaran, Schuermann, & Weiner, 2004), which have been applied to forecasts of regional unemployment in Germany (Schanne, Wapler, & Weyh, 2010), are also a particularly appealing alternative for forecasting and simulating regional impacts of COVID-19. Further studies along these lines can substantially expand our understanding of the mechanisms by which the effects of the pandemic may diffuse across regions.

ACKNOWLEDGEMENTS

The authors are grateful to the two anonymous reviewers for their very helpful comments and suggestions. Any errors are solely the authors' responsibility. Modrego wishes to thank the support of the Chilean National Agency of Research and Development (ANID) through the FONDECYT Initiation in Research Fund 2019 [Grant: 11190112].

Canales is partially supported by FONDECYT Postdoctoral Grant No. 3200650 and by the Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002.

ORCID

Félix Modrego https://orcid.org/0000-0002-7428-0039

Andrea Canales https://orcid.org/0000-0001-7536-0674

Héctor Bahamonde https://orcid.org/0000-0002-1146-6426

REFERENCES

Anríquez, G., & López, R. (2007). Agricultural growth and poverty in an archetypical middle income country: Chile 1987–2003. *Agricultural Economics*, *36*(2), 191–202.

Atienza, M., & Modrego, F. (2019). The spatially asymmetric evolution of mining services suppliers during the expansion and contraction phases of the copper super-cycle in Chile. *Resources Policy*, *61*, 77–87.

Baum, C. F., & Linz, T. (2009). Evaluating concavity for production and cost functions. *The Stata Journal*, *9*(1), 161–165.

Berndt, E. R. (1991). *The practice of econometrics: Classic and contemporary*. Reading, MA: Addison-Wesley Publishing.

Cerda, H. A. (2018). Inversión, stock de capital e infraestructuras en la economía chilena: una aproximación por regiones y actividad económica, 1990–2010. Ph.D. Thesis, Universidad Complutense de Madrid.

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. New York: Cambridge University Press.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic production function. *Econometrica*, *39*(4), 255–256.

Darby, M. R., Haltiwanger, J. C., & Plant, M. W. (1985). Unemployment-rate dynamics and persistent unemployment under rational expectations. National Bureau of Economic Research Working Paper 1558.

Darby, M. R., Haltiwanger, J. C., & Plant, M. W. (1986). The ins and outs of unemployment: The ins win. National Bureau of Economic Research Workingh Paper 1997.

Davis, S. J., & Haltiwanger, J. (1990). Gross job creation and destruction: Microeconomic evidence and macroeconomic implications. *NBER Macroeconomics Annual*, *5*, 123–168.

Davis, S. J., & Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, *107*(3), 819–863.

Diewert, W. E. (1974). Applications of duality theory. In M. D. Intrilligator & D. A. Kendrick (Eds.), *Frontiers of quantitative economics* (Vol. II) (pp. 106–171). Amsterdam: North-Holland Publishing.

Diewert, W. E., & Wales, T. J. (1987). Flexible functional forms and global curvature conditions. *Econometrica*, *55*(1), 43–68.

Finch, W. H., & Finch, M. E. H. (2017). Multivariate regression with small samples: A comparison of estimation methods. *General Linear Model Journal*, *43*, 16–30.

Guerrieri, V., Lorenzoni, G., Straub, L., & Werning, I. (2020). Macroeconomic implications of COVID-19: Can negative supply shocks Ccause demand ahortages?. National Bureau of Economic Research Working Paper 26918.

Hussain, J., & Bernard, J. T. (2018). Flexible functional forms and curvature conditions: Parametric productivity estimation in Canadian and US manufacturing industries. In W. H. Greene, L. Khalaf, P. Makdissi, R. C. Sickles, M. Veall, & M. C. Voia (Eds.), *Productivity and inequality* (pp. 203–228). New York: Springer.

International Monetary Fund (IMF). (2020). World economic outlook, April 2020: The great lockdown. Washington, DC: IMF Retrieved from https://www.imf.org/en/Publications/WEO

Lund, S., Ellingrud, K., Hancock, B., Manyika, J., & Dua, A. (2020). Lives and livelihoods: Assessing the near-term impact of COVID-19 on US workers. McKinsey global institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20Sector/Our%20Insights/Lives%20and%20livelihoods%20Assessing%20the%20near%20term%20impact%20of%20COVID%2019%20on%20US%20workers/Lives-and-livelihoods-Assessing-the-near-term-impact-of-COVID-19-on-US-workers.pdf

Muro, M., Maxim, R., & Whiton, J. (2020). The places a COVID-19 recession will likely hit hardest. The Brookings Institution. Metropolitan Policy Program. Retrieved from https://www.brookings.edu/blog/the-avenue/2020/03/17/the-places-a-covid-19-recession-will-likely-hit-hardest/

Olfert, M. R., Partridge, M., Berdegue, J., Escobal, J., Jara, B., & Modrego, F. (2014). Places for place-based policy. *Development and Policy Review*, *32*(1), 5–32.

Pesaran, M. H., Schuermann, T., & Weiner, S. M. (2004). Modelling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economic Statistics*, *22*, 129–162.

Pollak, R. A., Sickles, R. C., & Wales, T. J. (1984). The CES-translog: Specification and estimation of a new cost function. *The Review of Economics and Statistics*, *66*(4), 602–607.

Ryan, D. L., & Wales, T. J. (2000). Imposing local concavity in the translog and generalized Leontief cost functions. *Economics Letters*, *67*(3), 253–260.
Schanne, N., Wapler, R., & Weyh, A. (2010). Regional unemployment forecasts with spatial interdependencies. *International Journal of Forecasting*, *26*(4), 908–926.
UN News. (2020). COVID-19: impact could cause equivalent of 195 million job losses, says ILO chief. Media release. Retrieved from https://news.un.org/en/story/2020/04/1061322
Valdez, R. I. (2019). Spatial diffusion of economic growth and externalities in Mexico. *Investigaciones Regionales—Journal of Regional Research*, *45*, 139–160.
World Trade Organization (WTO). (2020). Trade set to plunge as COVID-19 pandemic upends global economy. World Trade Organization Press Release 855. Retrieved from https://www.wto.org/english/news_e/pres20_e/pr855_e.htm
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368.

How to cite this article: Modrego F, Canales A, Bahamonde H. Employment effects of COVID-19 across Chilean regions: An application of the translog cost function. *Reg Sci Policy Pract*. 2020;1–17. https://doi.org/10.1111/rsp3.12337

Acta Politica
https://doi.org/10.1057/s41269-020-00174-4

ORIGINAL ARTICLE

Still for sale: the micro-dynamics of vote selling in the United States, evidence from a list experiment

Héctor Bahamonde[1]

Abstract

In nineteenth-century United States politics, vote buying was commonplace. Nowadays, vote buying seems to have declined. The quantitative empirical literature emphasizes vote buying, ignoring the micro-dynamics of vote selling. We seem to know that vote buyers can no longer afford this strategy; however, we do not know what American voters would do if offered the chance to sell their vote. Would they sell, and at what price, or would they consistently opt out of vote selling? A novel experimental dataset representative at the national level comprises 1479 US voters who participated in an online list experiment in 2016, and the results are striking: Approximately 25% would sell their vote for a minimum payment of $418. Democrats and Liberals are more likely to sell, while education or income levels do not seem to impact the likelihood of vote selling.

Keywords Vote buying · Vote selling · Clientelism · List experiments · United States

Vote sellers and vote buyers

Prior research on clientelism usually focuses on whether parties have attempted to buy votes (Vicente and Wantchekon 2009; Vicente 2014; Rueda 2015, 2017; Reynolds 1980; Nichter 2014; de Jonge 2015; Finan and Schechter 2012; González-Ocantos et al. 2014; Diaz-Cayeros et al. 2012; Brusco et al. 2004). Unfortunately, while this is an important question, it overlooks the conditions under which citizens would sell their vote. In fact, Nichter and Peress (2017) explain that studies continue to view clientelism typically as a top-down process, generally overlooking citizens'

Héctor Bahamonde
hector.bahamonde@uoh.cl
https://www.HectorBahamonde.com

[1] Instituto de Ciencias Sociales, O'Higgins University, Rancagua, Chile

Published online: 22 July 2020

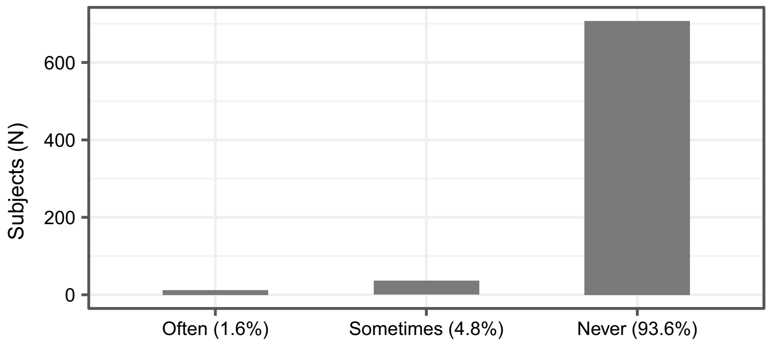


Fig. 1 Frequency of clientelism in the United States (2010). *Note* figure shows the frequency of survey respondents, $N = 755$ *Source:* LAPOP*, 2010 wave for the United States. Question is* `clien1`: "In recent years and thinking about election campaigns, has a candidate or someone from a political party offered you something, like a favor, food, or any other benefit or object in return for your vote or support? Has this happened often, sometimes, or never?"

demands. Since several questions pertaining to vote sellers remain unanswered, a bottom-up reconceptualization is necessary. For instance: *What would voters do if offered the chance to sell their vote? Would they sell it? And at what price?*[1]

To illustrate the issue at hand, Fig. 1 shows responses of US citizens asked whether a candidate or a member of a political party has offered something in exchange for their vote, completely ignoring voters' preferences. The figure begs the question of whether survey respondents who answered "never" *would* still be willing to sell their votes.

It seems that whether studies focus on vote buying or vote selling depends partly on methodological rather than theoretical decisions.[2] On the one hand, historical and/or ethnographically based contributions describe clientelist transactions from the point of view of voters, focusing on the conditions that make vote selling most likely (Posada-Carbó 1996; Sabato 2001; Auyero 2000; Szwarcberg 2013; Borges 2019). On the other hand, statistical, survey, and/or experimentally based work mostly explores issues related to vote buying. For example, using a field experiment in Benin, Wantchekon (2003) stresses the role of incumbency on vote buying. Jensen and Justesen (2014, p. 227) focus on the impact of "poverty on vote buying," while Khemani (2015, p. 84) shows that "vote buying in poor democracies is associated with lower [public] investments." Hence, and except for several important quantitative studies (Corstange 2012; Imai et al. 2015; Nichter and Peress 2017; Hicken et al. 2015, 2018; Michael and Thachil 2018), the emphasis of statistical studies remains on studying vote buying. Importantly, other statistically based studies have explored attitudes toward vote buying (Bratton 2008; Weitz-Shapiro 2012).

[1] It is important to note that clientelism as a practice involves more than just buying or selling votes. Other goods might be involved in the clientelist transaction—for instance, public jobs or public infrastructure, e.g., see for example Dixit and Londregan (1996), Calvo and Murillo (2004), and Khemani (2015). However, this paper's focus is on just vote buying and vote selling.

[2] I thank one of the anonymous reviewers for this comment.

They suggest that a strong stigma is attached to vote buying, which might make voters unwilling to sell their vote. For instance, González-Ocantos et al. (2014, p. 208) designed a list experiment to study attitudes toward vote buying in Latin America. They conclude that most respondents find vote buying "unacceptable when provided with a hypothetical example."

While the quantitative literature has advanced several important avenues of research, it has overlooked many important questions. The wording of the Latin American Public Opinion Project (LAPOP) question illustrates part of the issue. By focusing on vote buying, it gives the falsely optimistic impression that US voters systematically "oppose" vote buying, "thus" rarely engaging in clientelism (as Fig. 1 strongly suggests). Furthermore, most quantitative studies were conducted primarily in developing countries, seriously narrowing the scope of our inferences. In part, this is because the clientelism literature usually focuses on realized behaviors only—that is, actual clientelist transactions. Unfortunately, by ignoring attitudes of potential vote sellers, particularly when it comes to the willingness to sell, selection bias seriously threatens causal inferences.

This paper makes both methodological and substantive contributions to the literature by leveraging a list experiment on hypothetical vote selling in a consolidated democracy. We believe that studying hypothetical behaviors—such as the willingness to sell—is a valuable exercise. Geddes (1990, p. 131) explains the well-known selection issues of studying "only cases that have achieved the outcome of interest." Hence, if we are interested in understanding the micro-dynamics of clientelism—particularly as a supply-and-demand issue—we should incorporate the preferences of both sellers and buyers, potential and/or actual. Since the focus of this paper is on the willingness to sell, we believe that we can also learn from *unrealized* clientelist transactions. Following the lead of González-Ocantos et al. (2014), this paper presents experimental evidence of hypothetical vote selling in the United States.

In 2016, a novel dataset representative at the national level was collected. A total of 1479 US voters participated in a list experiment between March 2 and March 6. This experiment made possible both the identification of the demographic factors that would make US voters more likely to sell their vote, and at what price, and the investigation of whether they would systematically lie about selling their vote. The results are striking. The data suggest that a sizable portion of US voters are willing to sell their vote (approximately 25%), would sell it for at least $418, and would systematically lie about it (approximately 8%). Given that these data are representative at the national level (i.e., this is not a convenient sample), these findings are surprising. Democrats and Liberals are systematically more likely to sell than Republicans. Education and income levels do not seem to have a systematic impact on the willingness to sell.

While this paper essentially describes the phenomenon, it leaves for future research further consideration of the causes of hypothetical vote selling in the United States. Ultimately, this paper attempts to bring voters back into the quantitative study of clientelism, particularly by studying their willingness to sell.

The United States as a case

At first, many advanced democracies were clientelist political systems. For instance, Stokes et al. (2013, p. 200) explain that in the nineteenth-century United States, "vote buying was commonplace" and "the major urban political institution in the late nineteenth century" (Erie 1990, p. 2). In Chicago, New York City, Newark, and other large American cities, votes were exchanged for "cash, food, alcohol, health care, poverty relief, and myriad other benefits" (Stokes et al. 2013, p. 200). The street price of the right to vote freely was low. Bensel explains that "[voters] handed in a party ticket in return for a shot of whiskey, a pair of boots, or a small amount of money" (Stokes et al. 2013, p. 227). In general, students of American political development have analyzed vote buying in detail, confirming both its early development and its generalized practice (Bensel 2004; Campbell 2005).[3]

However, vote buying currently seems to have declined considerably, for two competing reasons. Stokes et al. (2013, p. 201) show that industrialization drove up the electorate's median income, making vote buying more expensive for party machines. However, Kitschelt and Wilkinson (2006, p. 320) disregard the industrialization hypothesis, focusing on the lower levels of "[s]tate involvement in the public sector."

Regardless, clientelist linkages are now rare. Figure 1 suggests that 93.6% of US respondents have never received a clientelist offer from a political party. While only a very small percentage (4.8%) report receiving such an offer from a political party, we do not know whether survey respondents *would* sell their votes. This paper presents systematic evidence that they would. Consequently, the counterintuitive results presented in this paper make our descriptive efforts worth pursuing. Representing the United States as a "crucial case," both the narrative and the findings follow a "least-likely" design approach. As Levy (2008, p. 12) explains, "[i]nferential leverage from a least likely case is enhanced if our theoretical priors for the leading alternative explanation make it a most likely case for that theory." The vote-buying literature mostly considers developing countries and describes vote sellers as poor (Weitz-Shapiro 2014, p. 12), uneducated (González-Ocantos et al. 2014), and undemocratic (Carlin and Moseley 2015). Thus, previous literature implies that the willingness to sell votes in the United States should be low, making it a difficult case study on vote selling.

The evidence that this paper presents may be associated with a probable erosion of American democracy.[4] In a highly controversial pair of articles, Foa and Mounk (2016, p. 7) document a deep "crisis of democratic legitimacy [that] extends across a [...] wider set of indicators" in the United States. They find that 26% of millennials declare that it is "unimportant" in a democracy for people to "choose their leaders in free elections" (Foa and Mounk 2016, p. 10, and Foa and Mounk 2017). These findings raise many (unanswered) questions regarding the actual value that American electoral institutions hold for citizens, possibly undermining the legitimacy of

[3] For the British case during the Victorian era see Kam (2017).

[4] Relatedly, see Levitsky and Ziblatt (2018).

the integrity of voting. Is voting unimportant enough to lead US citizens to sell their votes if offered the possibility?

The next section gives a historical account of vote buying and vote selling in the United States. The section also attempts to situate both within a historical context. It particularly shows how vote buying and vote selling transitioned from their status as an important institution in American elections to a scarcely practiced electoral method. The following section explains the experimental design. Immediately thereafter, the paper presents the statistical analyses of the experimental data. The last section offers some working hypotheses and possible lines for future research.

Vote selling and patronage in the United States: a brief historical account

While all US states made bribery of voters illegal early in US history, these laws were purposely ignored. Well before the Gilded Age (1877–1896), several norms aimed to prohibit bribery, clientelism, and patronage. For instance, as early as 1725, the New Jersey legislature had already outlawed many electoral malpractices (Bensel 2004, p. 59). However, these restrictions were systematically bypassed. To circumvent property qualifications, for instance, office-seekers (and their supporters) commonly bought "freeholds for landless men in return for their vote" (Campbell 2005, p. 6), a practice known as "fagot voting." Since it was a coercive bribe, after "the election, the land was simply returned to the original owner" (p. 6).

Weak institutions, poor bureaucracies, and bad-quality record-keeping helped to foster electoral malpractice.[5] First, most states did not have actual registration laws, making voter eligibility difficult to determine (Argersinger 1985, p. 672). Historians frequently report that judges at polling places had a hard time determining not only the age of the potential voter,[6] but also whether the prospective voter was a US citizen, especially in cases that involved newly naturalized immigrants with strong foreign accents (Bensel 2004, p. 20). Consequently, it was often up to the judge's discretion whether to let prospective voters cast a ballot. Since judges were party appointees (Argersinger 1985, p. 672), their discretionary powers were systematically used to shape electoral outcomes.

Low literacy levels also helped to sustain vote selling in the United States. For example, in Kentucky and Missouri, the law required voters to verbally announce their choices at the polling places, instead of using party tickets (Bensel 2004, p. 54). Of course, the viva voce method was convenient for party workers who usually swarmed around the polling places. However, the ticket system eventually supplanted this method.

[5] The U.S. Bureau of the Census did not exist. Consequently, it was relatively easy to invent names, "repeat," or use any other subterfuge to "stuff the ballot box." In fact, "a St. Louis politician admitted registry fraud but argued that there was no proof that the names he copied into the registry were of real people and, therefore, no crime had been committed" (Argersinger 1985, p. 680).

[6] Judges used as a rough proxy whether the prospective voter had the ability to grow a beard (Bensel 2004, p. 20).

The "party strip" or "unofficial" ballot system also permitted all sorts of fraudulent election practices. The parties themselves produced party tickets. Since tickets varied by size and color, it made "the voter's choice of party a public act and rendered voters susceptible to various forms of intimidation and influence while facilitating vote buying" (Argersinger 1985, p. 672). Similarly, Rusk (1970, p. 1221) explains that distinctive ticket colors and shapes "assured instant recognition of the ballot by the voters [and] party workers." Reynolds and McCormick (1986, p. 836) present similar evidence. Consequently, party workers hired to monitor the voting window (Argersinger 1985, p. 672) had ample opportunity to punish or reward voters accordingly.

The ticket system required very strong party machines, which, in turn, required considerable economic resources to make the system work. However, political machines were oiled not only with money. On the one hand, many "ticket peddlers" (Argersinger 1985, p. 672) were volunteers (Bensel 2004, p. 17), saving some of the costs needed to maintain the machine. Most of these volunteers "enjoyed the patronage of elected party officials by holding government jobs, drawing public pensions, servicing government contracts, or enjoying special licensing privileges" (Bensel 2004, p. 17). On the other hand, political appointees "from janitor to secretary of state" and some corporations donated annually part of their salaries and revenues (Reynolds 1980, p. 197). Thus, parties amassed huge amounts of money.

With all these resources flooding the polls on election day, voting was truly an interesting spectacle. On that day, party agents would offer voters plenty of liquor as an incentive to vote the party ticket. Hence, "the street or square outside the voting window frequently became a kind of alcoholic festival in which many men were clearly and spectacularly drunk [to the point that] some could not remember whether or not they had voted" (Bensel 2004, p. 20). Even before the Gilded Age, American elections were engineered according to these "principles." When running for the Virginia House, a young George Washington "spent nearly 40 pounds—a considerable sum for the day—on gallons of rum, wine, brandy, and beer; all used to win over the votes of his neighbors" (Campbell 2005, p. 5).[7]

The Australian ballot system significantly reduced the frequency of most of this malpractice (Rusk 1970, p. 1221). However, as vote selling and vote buying were so embedded in what was considered normal, the immediate effect of the Australian system was to reduce turnout (Reynolds and McCormick 1986, p. 851).

Today, the modus operandi of clientelism has changed, and both the frequency of vote buying/selling and the importance of party machines have declined. Scholars have pointed out that "party machines are a thing of the past" (Stokes et al. 2013, p. 230). However, some contemporary accounts remain of vote buying and selling in American elections. For instance, Campbell (2005, pp. 243–244) explains how a Democratic leader in Logan County, West Virginia, accepted $35,000 in cash to support Senator Kennedy. As the Democratic leader explained, "this money was for one purpose: 'We bought votes with it [...] that's the way real politics works.' " Other examples are the famous primary election in March 1972 in Chicago (p. 262)

[7] $1250 in 2017 US dollars. Conversion based on Williamson (2018).

and the elections in the coal-rich Appalachian Mountains during the 1980s (p. 275). Similarly, nonacademic sources find that during the 2010 elections, "selling votes [was a] common type of election fraud" (Fahrenthold 2012). Others find that "[v]ote-buying is extremely common in *developed* [...] countries" (Leight et al. 2016, p. 1). If vote buying is "a thing of the past," why do we still see it? How common is vote selling? The next two sections attempt to quantify—in an unbiased way—the willingness to sell votes among a representative sample of US voters.

Experimental design

The study of individual preferences depends on truthful answers. However, under certain circumstances, individuals might not want to answer truthfully, due to social pressure. For instance, to avoid having the interviewer judge them, individuals might not want to reveal having done something illegal, such as selling one's vote. Failing to consider this systematic source of bias will pose threats to causal inference.

Since list experiments administer two lists of items (one to the control group, one to the treated group), list experiments are well suited to eliciting truthful answers (Blair 2015). Both lists look identical (e.g., each containing the same three items); however, the treatment list traditionally includes a fourth item, the sensitive item related to some socially condemned behavior. Respondents are asked how many items on the list they would endorse, not which ones. For instance, if an experimental subject answers "2," the interviewer will not know whether that number includes the sensitive item. Consequently, if the survey respondent wants to endorse the sensitive item, the answer will be "masked" by the other items in the list. This concealment makes this technique suitable for studying socially condemned behaviors, such as vote buying (Corstange 2008; González-Ocantos et al. 2012; Corstange 2012; Blair and Imai 2012), drug use (Druckman et al. 2015), sexual preferences (LaBrie and Earleywine 2000), and attitudes toward race (Kuklinski et al. 1997; Redlawsk et al. 2010).

Given that both lists are assigned randomly, the mean number of nonsensitive activities that respondents endorse should be equal across the two lists. However, if there are any differences in means between the two groups, the differences should be attributed only to the presence of the sensitive item.

Blair and Imai (2012) and Imai et al. (2015) provide a statistical framework to analyze list data efficiently.[8] They formalize two assumptions, namely, that there are (1) "no design effects" (i.e., the inclusion of a sensitive item has no effect on respondents' answers to control items), and (2) "no liars" (i.e., respondents give truthful answers for the sensitive item). When the two assumptions hold and the item counts for types $y = 0$ and 4 are fully observed,[9] experimental subjects with

[8] While list experiments are common, researchers unfortunately "[utilize] only a difference in means estimator, and [do] not provide a measure of the sensitive item for each respondent" (Glynn 2013, p. 159).

[9] For a hypothetical treatment list of four items.

item-count types $y = 1$, 2, and 3 can be inferred using multivariate techniques that allow for inferring who answered "yes" to the sensitive item. In addition, the statistical analyses permit studying the relationship between preferences over the sensitive item (i.e., vote selling) and an individual's characteristics, such as income and party identification. Also, the design includes a "direct" question on the sensitive item, also making possible an estimation of the amount of social-desirability bias.

Collected in 2016, the data ($N = 1479$) are representative at the national level.[10] Figure 6 shows the geographical distribution of survey respondents, grouped by party identification. The experiment was framed as a study about crime in the United States, not as a study about vote selling.[11] While pretesting the study, it was decided that the experiment needed to mask a very serious felony (selling one's vote) among other equally serious felonies (such as stealing) and other less serious crimes (such as speeding or downloading music illegally from the Internet). Otherwise, the vote-selling item would have stood out among the other items, making it seem totally negative and undoable, and/or making the true purpose of the study obvious.

Before splitting the subject pool into the subjects' respective experimental conditions, participants were asked to read an excerpt describing four illegal activities (including vote selling).[12] All were formatted as news pieces. The idea was to explain "vote selling" to "newsreaders."

As Fig. 2 suggests, to prevent possible priming effects,[13] the order in which experimental subjects answered the direct question[14] and the list experiment were randomly assigned. To be sure, all subjects answered both the direct question and the list experiment. To further prevent the possibility of biased answers when asking the direct question to individuals in the treated group, the direct question stated that the hypothetical possibility of doing one of the illegal things mentioned previously in the excerpt would be randomly assigned. However, all participants were directly asked whether they would be interested in selling their vote. Direct answers were then used to estimate the proportion of "liars."

As a follow-up, subjects answering "yes" to the direct question answered a pricing test that asked them to indirectly put a price on their votes. Following standard practice in marketing research, participants slid a handle indicating which price was considered "too cheap" for one's vote. The slide ranged from \$0 to \$1000, in one-dollar increments. The idea was to capture the respondent's willingness to sell. The pricing test particularly measures the lowest bound at which the participant would perceive the least economic benefit that was still enough to make selling his/her vote interesting. Moving forward, the list experiment contemplated one control and two

[10] *Research Now SSI* collected the data between March 2 and March 6. Survey respondents belong to the online panel owned and administered by SSI. Notice of IRB Exemption Protocol #E16-292 is kept on file at the Office of Research and Regulatory Affairs of Rutgers University.

[11] To isolate the risks and costs associated with engaging in any illegal activity, the next phrase was included: "assuming you would not go to jail."

[12] See "Appendix" for wording.

[13] Blair and Imai (2012, p. 54) explain that asking the direct question to individuals in the treated group might bias the results.

[14] See "Appendix" for wording.

possible treatments, each with different vote-selling prices. Since pricing a vote is difficult and based on several pretests, an arbitrarily low ($100) or an arbitrarily high ($500) price appeared. Participants randomly assigned to the treatment condition answered only one of these two treatments. More than hard-pricing tests, these two treatments account for possible different elasticities that might have interacted with individual socioeconomic backgrounds, mainly defined by income.[15]

Subjects randomly assigned to the control condition answered the following question:

```
Now, you will have to type HOW MANY, if any, of the following illegal activities
    you might engage in, assuming you would not go to jail.

(1) steal an iPod from a large department store
(2) speed on the highway because you're late for work/school
(3) download your favorite music from the internet illegally

Type in HOW MANY (NOT WHICH), if any, of these things you would do.
```

In turn, subjects randomly assigned to the treatment condition answered one of the following two treatments. Subjects assigned to the "low" price condition read the same vignette, but the next line was added in the third place: `(3) sell your vote to a candidate for $100`. Those assigned to the "high" price condition read the following line: `(3) sell your vote to a candidate for $500`.[16]

Figure 3 shows the distinct frequencies of participants declaring how many (if any) illegal things they would do. Notice that the figure shows the total number of items, not which ones. For instance, a frequency of "3" does not mean the frequency of the third item, but describes the total number of individuals answering that they would do three of the illegal activities described in the vignette.[17] The order of the items was not randomized, to avoid violating the stable unit treatment value assumption (SUTVA).[18]

Showing that the probability of being assigned to any condition is not associated with individual covariates is important. Table 1 shows a multinomial logistic model. The dependent variable is the treatment condition (high treatment, low treatment, and control). The independent variables are observable characteristics captured

[15] Holland and Palmer-Rubin (2015, p. 1189) explain that "the poor are thought to be more susceptible to vote buying."

[16] Since one of the two sentences was added, item `(3) download your favorite music from the Internet illegally` was moved to the fourth place.

[17] The experimental design passes the standard tests for design effects (floor and ceiling effects). See Table 3.

[18] Morton and Williams (2010, p. 98) explain that the treatment should be invariant or "stable."

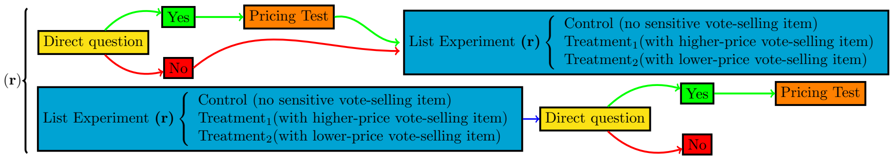


Fig. 2 Experimental flow of the list design. *Note* this figure shows the flow of the list experiment. Notice that (1) the order in which experimental subjects answered both the direct question and the list experiment was randomized; (2) there are two treatments, one with a selling price of \$100 ("low") and one with a selling price of \$500 ("high")

by a short questionnaire included in the study. Four variables were used: income, education, party identification, and political ideology. These were the same set of variables used when estimating likely vote sellers (below). Conveniently, the base category in the multinomial logistic regression is the control condition. The coefficients in the table are all zeros (and statistically nonsignificant). Consequently, these results show no observable differences between the "high" treatment condition and the control group. The same applies to the "low" condition.[19]

The paper acknowledges that considerable friction and transaction costs in the real world might mean that creating a market for vote selling would not be easy. For instance, party identification might increase (or decrease) the cost of selling one's vote, presumably preventing (or fostering) the transaction. If the party of both sellers and buyers should match, fostering vote selling might represent a win–win situation for both. This experimental design does not consider blocking on party identification, as that might have increased considerably the number of cells.

Statistical analyses

Would US citizens sell their vote?

Table 2 shows a simple difference-in-means analysis between each treated group and the control group. On average, the control group would do 1.116 things on the list. Subjects treated under the "low" condition (\$100) would do 1.182 things on the list, while subjects in the "high" condition (\$500) would do 1.189 things.

Three important points characterize this bivariate analysis. First, the mean differences between treated groups (i.e., "low" and "high" treatments) are statistically zero, implying that neither treatment should introduce design bias into the experiment. Second, while treated subjects do have slightly higher means when compared to the control group (indicating some vote-selling propensity), these differences are not statistically significant. Third, while not statistically significant, $0.066 \times 100 = 6.6\%$ of subjects would sell their vote under the "low"

[19] I thank the anonymous reviewer at *Acta Politica* for this suggestion.

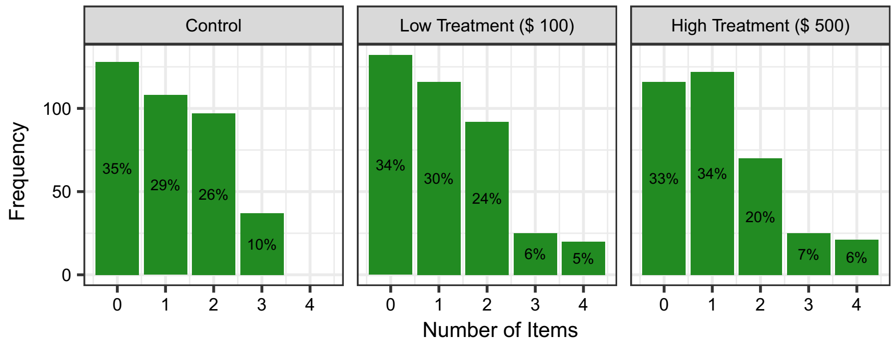


Fig. 3 Frequency and percentages of subjects declaring how many (if any) illegal things they would do. *Note* notice that the *X*-axis denotes the number of items, not which ones. Percentages show proportions per condition

Table 1 Covariate balance: multinomial logistic regression for both treatment conditions

	High	Low
Ideology	0.019 (0.068)	−0.031 (0.067)
Party Id.	−0.125 (0.083)	0.022 (0.080)
Income	−0.021 (0.022)	0.006 (0.021)
Education	0.049 (0.048)	−0.008 (0.047)
AIC	2449.471	2449.471
BIC	2499.583	2499.583
Log likelihood	−1214.736	−1214.736

The table shows a multinomial logistic regression. The dependent variable is the treatment condition (high, low, control). In both models, the base category is the control condition. The independent variables are observable characteristics captured by a short questionnaire included in the study. This set of covariates is the same as the one used in the statistical analyses of the list experiment. Since all estimated coefficients are close to zero and statistically nonsignificant, we can safely assume that the randomization mechanism worked as expected, i.e., there are no observable differences across the different treatment conditions. Reference category is control condition. Intercept was excluded from the table. $N = 1479$

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1.$

condition, while $0.073 \times 100 = 7.3\%$ of subjects would sell their vote under the "high" condition. While these estimations score substantially under what is found through the multivariate approach used in this study, as shown below, they are also highly inefficient.

Bivariate calculations are statistically inefficient; hence, the data should be analyzed using multivariate techniques instead. Following the advice of Blair and Imai

Table 2 Differences in means between Treatments (high and low) and the Control group

Condition	Mean	Difference with control condition	Confidence intervals	t	df	p value
Low ($100)	1.182	1.182–1.116 = 6.6%	[−9%, 22%]	0.846	748	0.398
High ($500)	1.189	1.189–1.116 = 7.3%	[−8%, 23%]	0.913	700	0.361

The table shows two-tailed t tests between each experimental treated unit ("low" and "high" conditions) and the control group. The table shows that $0.066 \times 100 = 6.6\%$ of subjects would sell their vote under the "low" condition, while $0.073 \times 100 = 7.3\%$ of subjects would sell their vote under the "high" condition. Also, 95% confidence intervals are shown. It is evident that they are quite wide and not statistically significant

(2012) and Blair (2015), we took a statistical multivariate approach.[20] Exploiting the "low" and "high" treatments, we estimated two identical statistical models. In both models, the outcome variable is the item count of things that subjects would do. The idea is to estimate what we cannot observe (i.e., vote selling), using information that we do observe (i.e., socioeconomic and political variables captured by the questionnaire). The model considers the most common covariates studied in the vote-buying literature (Calvo and Murillo 2004; Stokes 2005; Kitschelt and Wilkinson 2006; Nazareno et al. 2008; Weitz-Shapiro 2012; González-Ocantos et al. 2014; Oliveros 2016; Bahamonde 2018)—that is, income, education, party identification, and political ideology.

Leveraging this multivariate approach makes estimating the proportion of hypothetical vote sellers possible. For both the "low" and "high" treatments, Fig. 4 shows the proportions of declared vote sellers ("Direct Question"), predicted vote sellers ("List Experiment"), and the difference between the two ("Social Desirability").[21] Substantively, the figure suggests that after combining the estimates of the "low" and "high" treatments, approximately 25% of the nationally representative sample would be willing to sell their vote.[22] While a considerable proportion answered the direct question affirmatively (18%),[23] the analyses still suggest that survey respondents systematically underreported their true answers—that is, approximately 8% of the nationally representative sample would have lied.[24]

The difference-in-means approach in Table 2 suggests that between 6.6 and 7.3% would be willing to sell their votes. However, the multivariate approach in Fig. 4 suggests that 25% would be willing to do so. While at first these differences might seem huge, they are not. As the literature suggests, multivariate approaches to analyzing list experiment data are far more efficient (Blair and Imai 2012; Blair 2015).

[20] The `R` package `list` was used (Blair 2015). The estimation method used was the "ml" and the maximum number of iterations was 200,000. The remaining arguments of the package were left at their default values.

[21] Since the estimated quantities do not vary across the different treatments ("low" and "high"), it is reasonable to think that there are no specific concerns associated with the (arbitrarily) chosen prices.

[22] This number was calculated averaging over the "high" (27%) and "low" (23%) estimates.

[23] This number was calculated averaging over the "high" (19%) and "low" (17%) estimates.

[24] This number was calculated averaging over the "high" (8%) and "low" (7%) estimates.

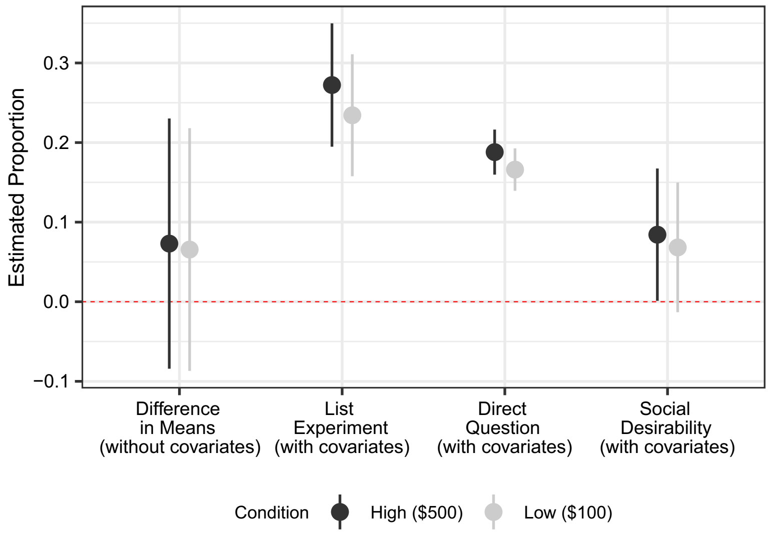


Fig. 4 List experiment data: declared and predicted vote sellers. *Note* figure summarizes Table 2 by showing the simple difference in means (without covariates). It also shows the proportion of declared ("Direct Question") and predicted ("List Experiment") hypothetical vote sellers, and the difference ("Social Desirability"). The three sets of main estimates were obtained via a multivariate procedure (including covariates). Combining both "low" and "high" treatments, 25% would be willing to sell their votes. And of those who answered affirmatively when asked directly (18%) an estimated additional 8% lied about it. "Liars" answer the direct question negatively, but they are likely sellers. The figure shows 95% confidence intervals. There are two arbitrarily "low" and "high" vote-selling prices. The reason for having both was to control for possible price elasticities. The figure suggests some small differences that are not statistically significant. Consequently, these arbitrary pricing decisions do not threaten the experimental design

Within the framework of regression analysis, the difference-in-means approach is just a bivariate lineal model.[25] Instead, the multivariate approach is also a lineal model, but it incorporates covariates. We claim that due to the multivariate's greater efficiency than that of the difference-in-means approach, the former is a far better approach than the latter. One way of showing the efficiency of a statistical model is by examining its standard errors (King 1986, p. 676): the worse the data's fit is, the greater the standard errors are, the more imprecise the model is, and the wider are the confidence intervals. Considering the statistical uncertainty of both methods (depicted in Fig. 4), it is easy to see that the multivariate approach is far more efficient than the difference-in-means approach. Since it uses more information when fitting the data (the covariates), it gives more precise estimates (narrower confidence intervals). Furthermore, going beyond efficiency issues, the estimates of both methods are statistically indistinguishable. Since the confidence intervals of

[25] With just a constant 1 on the right-hand side of the equation.

both approaches overlap, it is not possible to say that the estimated 7.3% and 6.6% are "smaller" than the estimated 25%.[26]

Moving forward, the estimated proportion of vote sellers—"List Experiment" in Fig. 4—is calculated using information from subjects with fully observable preferences, i.e., subjects with an item count of 0 or 4. We know that the former would not do anything, and the latter would do all things mentioned in the list (including the sensitive item). Using the identified covariates (income, education, party identification, and political ideology), a model is fitted to predict all subjects with 0's and 4's on the left-hand side. Using this information makes obtaining individual-level vote-selling predictions possible, i.e., participants who would do 1, 2, or 3 things on the list (shown in Fig. 7 in the "Appendix"). Then, these individual-level predictions are compared with the direct question that all experimental subjects answered. If a subject is a predicted vote seller but answers the direct question negatively, it is inferred that due to concerns of social desirability, she might have chosen to lie.

What is the price for which US citizens would sell their vote?

Participants were also asked to declare which price they considered "too cheap" for their vote. The intention was to capture the respondent's willingness to sell. The test measures the lowest bound at which participants would perceive the least possible economic benefit but enough to make them sell. Since it is the lowest threshold, the understanding is that a higher price will still be economically attractive.

The results indicate that the average survey respondent would sell his/her vote for \$418 ($N = 189$), a very expensive price. These results are not unrealistic. While the selling price is very high, it matches what others have found. Bahamonde (2018, p. 52) finds that clientelist political parties in Brazil do target affluent voters at considerably higher prices. Part of the argument is that higher levels of economic development not only raise personal income, but also shift the broker's vote-buying capacity upward.[27] That is, higher income does not necessarily stop vote buying; it just makes it more expensive.[28]

Stokes et al. (2013) analyze the (im)possibility of expensive vote selling. Industrialization has driven up the median income of the electorate, increasing the selling price while turning vote buying into an increasingly expensive strategy for winning elections. Thus, from the demand-side (parties), vote buying is no longer an efficient mass strategy for party machines. Evidently, with the selling price so expensive, political parties cannot catch up with the supply-side, making vote buying in the United States a rare event (as Fig. 1 suggests). This situation has forced party machines to turn to other, less prohibitively costly alternatives. Thus, these results suggest that from the supply-side (i.e., voters), the vote is still up for sale, only for a very high price that party machines cannot afford.

[26] I thank the two anonymous reviewers of *Acta Politica* for stimulating this discussion.

[27] Similarly, see Abramo and Speck (2001, p. 14). For the Philippine case, see Schaffer (2004).

[28] In fact, there is some anecdotal evidence suggesting that a broker purchased one man's vote for \$800 during the 2010 elections in eastern Kentucky (Shawn 2012, p. 6).

Since the pricing test is based on the direct question, its results require a word of caution. The list experiment does suggest that some respondents lied when directly asked if they would sell their vote. Consequently, we should expect the pricing test to be biased to some degree. Also, only a small proportion of respondents answered the direct question affirmatively. In addition, prices are the product of supply-and-demand dynamics. In this context, prices result from the interaction between parties (buyers) and voters (sellers). This research design observes only the sellers' side. Hence, we limit our inferences even more by thinking about these results as only suggestive of some willingness to sell. Hence, more than acting as definitive and final pricing tests, these findings do seem to suggest that the vote-selling price is high enough to deter political parties from engaging in vote selling. Finally, future research should design and conduct more complex studies where the design incorporates supply-and-demand dynamics.

Who are the most-likely vote sellers?

The proportion of likely vote sellers was estimated using a multivariate approach. The variables used were the most common explanatory factors studied in the clientelism literature. Ultimately, this procedure allows for profiling participants into likely vote sellers. Figure 5 shows estimated vote-selling probabilities at different levels of all variables used in the multivariate approach.

The analyses suggest that Democrats and Liberals are more likely to sell. These findings are in line with research that studies the different constitutive values of Liberals and Conservatives. Political psychologists have found that compared with Conservatives, Liberals construct their moral systems primarily upon narrower psychological foundations. Particularly, Liberals consider less important both the authority/respect and the purity/sanctity dyads (Graham et al. 2009, p. 1029). This might lead Liberals to engage more frequently in behaviors that might be considered "wrong," such as vote selling. In fact, Gray et al. (2014, p. 7) explain that Conservatives "see impure violations as relatively more wrong."

Unlike the conventional wisdom (Kitschelt 2000; Calvo and Murillo 2004; Weitz-Shapiro 2012; Carlin et al. 2015), Fig. 5 shows that education and income levels do not make vote selling more likely. Poverty has long been associated with vote selling. Brusco et al. (2004), Stokes et al. (2013), and Nazareno et al. (2008) explain that since the poor derive more utility from immediate transfers relative to returns associated with future (and uncertain) policy packages, clientelist political parties only target the poor. For instance, Weitz-Shapiro (2014, p. 12) explains that "[a] lmost *universally*, scholars of clientelism treat and analyze [this] practice as an exchange between politicians and their poor clients."[29] The evidence presented in this paper aligns with that of others who have recently questioned the importance of this canonical predictor. Szwarcberg (2013) "challenges the assumption [that brokers] will always distribute goods to low-income voters in exchange for electoral

[29] My emphasis.

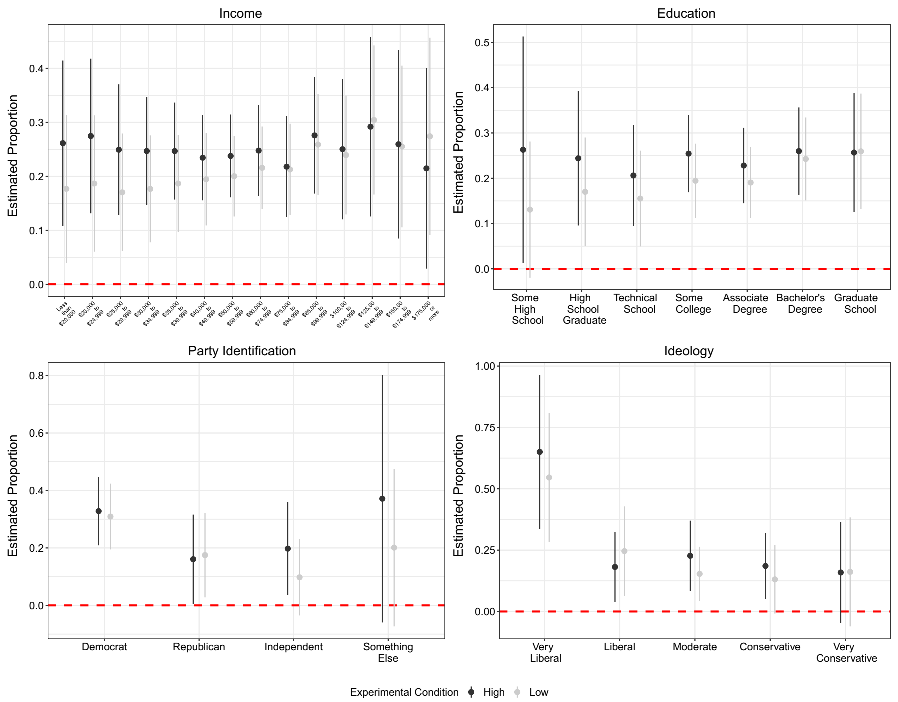


Fig. 5 List experiment: covariates used to estimate likely vote sellers. *Note* these variables were used in the multivariate statistical model to estimate individual-level probabilities of vote selling. The figure shows the predicted probabilities and their corresponding 95% confidence intervals for income, education, party identification, and ideology. Since the vote-selling prices were set arbitrarily, the reason for two experimental conditions ("low" and "high") was to control for possible price elasticities. While there are some perceptible changes, they are not statistically significant. Consequently, these arbitrary decisions do not threaten the identification strategy

support," while González-Ocantos et al. (2012) and Holland and Palmer-Rubin (2015) find that income had little or no effect on vote buying.[30] Notably, Bahamonde (2018) explains that brokers target individuals when they are identifiable and groups when brokers need to rely on the spillover effects of clientelism. Both mechanisms occur regardless of individual levels of income.

There do seem to be important substantive differences between the "low" and "high" vote-selling treatments. That is, factors that heavily determine economic status (income and education) seem to be more elastic to marginal increments in the buying price. As Fig. 5 shows, low-income and less-educated individuals are willing to sell their votes in a similar proportion to wealthier and more-educated respondents. However, poorer and uneducated individuals are more willing to sell their votes, conditional on higher prices. This might indicate that for them, behaving

[30] Relatedly, González-Ocantos et al. (2014, p. 205) and Corstange (2012, p. 494) also find very weak results for education in Peru and Nicaragua, and in Lebanon, respectively.

illegally is worthwhile but only when the payoff is "large enough." These results are in line with those of experimental and applied economists who argue that "risk aversion decreases as one rises above the poverty level and decreases significantly for the very wealthy" (Riley and Chow 1992, p. 32). In other words, less-educated and low-income individuals, who are more fragile and precarious, tend to avoid risks and, hence, illegal activities. On the contrary, higher-income and more-educated individuals seem unaffected by the different stimuli and sell their vote in the same proportion, regardless of the price. For instance, highly educated individuals (graduate school level) sell their vote in the same proportion, under both the "low" (26%) and "high" (26%) conditions.

General discussion

Two conflicting pictures emerge. On the one hand, leaving aside concerns about social-desirability bias, we "know"—using nonexperimental data—that most people have never been offered the possibility to sell their vote (as per Fig. 1). On the other hand, the results presented here strongly suggest that they *would*. While buyers (e.g., parties) are not buying, a large proportion of latent vote sellers is willing to sell their vote.

While vote buying/selling in the United States was commonplace during the nineteenth century, higher median incomes have increased the cost of this strategy as a feasible tool to win elections, in turn, making vote buying rare in the United States. The paper confirms this hypothesis by suggesting that an important estimated proportion of US voters—25%—is very much willing to sell their vote, but for an estimated very expensive price—$418. Overall, these results are striking, and the author is not aware of any other experimental design in which subjects in an industrialized democracy are asked whether they would sell their votes, and, moreover, which produces positive results. The paper began by establishing the tension between supply and demand sides within a clientelist relationship and noting that qualitative research usually focuses on vote selling, while quantitative studies usually focuses on vote buying. Furthermore, most of the literature concentrates its efforts on studying developing countries, mostly paying attention to realized clientelist transactions. As discussed, both aspects pose threats of selection bias to our inferences. This paper tries to fill these gaps by studying hypothetical vote selling via an experimental design implemented in an advanced democracy.

While the paper is rather descriptive, the author believes that the exercise was worth pursuing. The experimental evidence of a large critical mass willing to sell their votes in a developed country is novel. It is hoped that the paper sets the stage for future research and encourages other scholars to field the experimental design presented here in a comparative setting, to include both developed and developing countries. Future research should also consider different values placed on different offices.[31] It is reasonable to think that presidential, Senate, House, state-legislature,

[31] I owe this point to Christopher Chambers-Ju.

mayoral, and city-council elections produce different incentives and constraints regarding buying and selling votes. Also, future research should consider blocking party identification—for example, designing a more complex experiment, in which not only the price varies but also the vote-selling treatment is partisan.

Acknowledgements I thank Virginia Oliveros, Richard Lau, David Redlawsk, Christopher Chambers-Ju, Jessica Price, Maria Akchurin, the 2016 Experimental Research Group in Political Psychology at Rutgers University—New Brunswick, the Social Sciences Seminar at O'Higgins University, and the two anonymous reviewers at *Acta Politica* for their comments. This project was funded by the Center for the Experimental Study of Psychology and Politics at Rutgers University—New Brunswick. Bastián Garrido provided excellent research assistance. Usual caveats apply.

Appendix

Experimental manipulations and vignettes

Distractor paragraph. The next paragraph was used to distract subjects from the main purpose of the study, and also to define vote selling.

```
Washington, D.C.- A department store downtown had a robbery incident last week,
    reporting several missing iPods from their inventory. Authorities also inform
    that a group of local residents are trying to ``sell'' their votes to
    political candidates ahead of a local election for city council. Residents
    approached some of the candidates running for office and offered to vote for
    that candidate in return for monetary compensation. In a different subject
    matter, the local police station released a report on driving habits and
    behaviors in the Capitol district last week. Finally, cyber-crime has become
    an increasingly serious issue in the area in the past few year.
```

Direct Question. All subjects read the next paragraph, and then *all* answered the direct question:

```
Now you will be entered into a random lottery for the opportunity to do ONE of
    the illegal things you just read before. This means that you might be randomly
     offered to hypothetically do ANY of the activities mentioned before.
```

```
After a random assignment, you have been selected for the opportunity to
    hypothetically sell your vote. This means that you will have the hypothetical
    opportunity to accept money from a candidate for your vote. Would you be
    willing to accept the offer, assuming you would not go to jail? By selecting
    ``Yes,'' you could earn up to $1,000.
```

Testing for design effects

See Table 3.

Table 3 Test for list experiment design effects

Respondent types	Low condition		High condition	
	Estimate	Standard error	Estimate	Standard error
$(y = 0, t = 1)$	0.0031	0.0346	0.0183	0.0351
$(y = 1, t = 1)$	−0.0063	0.0349	−0.0345	0.0353
$(y = 2, t = 1)$	0.0169	0.0226	0.0299	0.0237
$(y = 3, t = 1)$	0.0519	0.0113	0.0593	0.0126
$(y = 0, t = 0)$	0.3429	0.0242	0.3277	0.0249
$(y = 1, t = 0)$	0.2982	0.0347	0.3264	0.0351
$(y = 2, t = 0)$	0.2453	0.0299	0.2322	0.0307
$(y = 3, t = 0)$	0.0481	0.0193	0.0407	0.02

Since the Bonferroni-corrected p values of the *low* (0.8567) and *high* (0.3298) conditions are above the specified α (0.05), I fail to reject the null of no design effects

Geographical distribution of survey respondents

See Fig. 6.

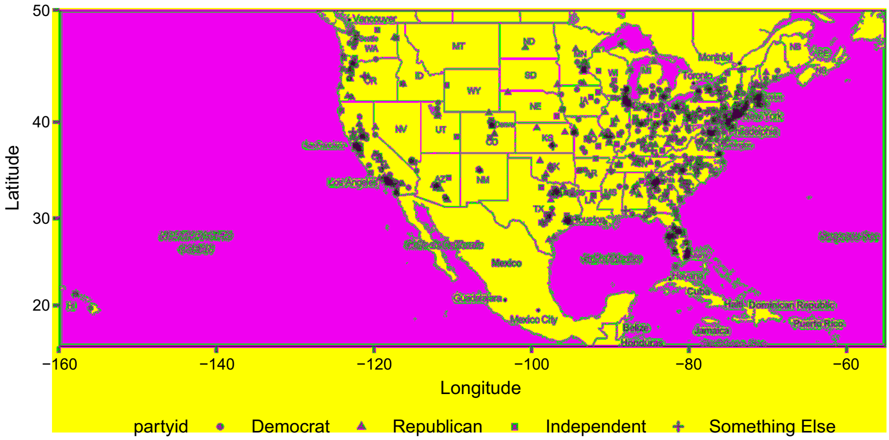


Fig. 6 Geographical distribution of survey respondents by party identification

Individual predictions

The vertical axis of Fig. 7 shows the estimated probabilities of the entire experimental sample, sorted across the horizontal axis. The figure is relevant as it openly shows the amount of uncertainty of the statistical estimates. Ultimately, these individual-specific predictions will be used to profile likely vote sellers.

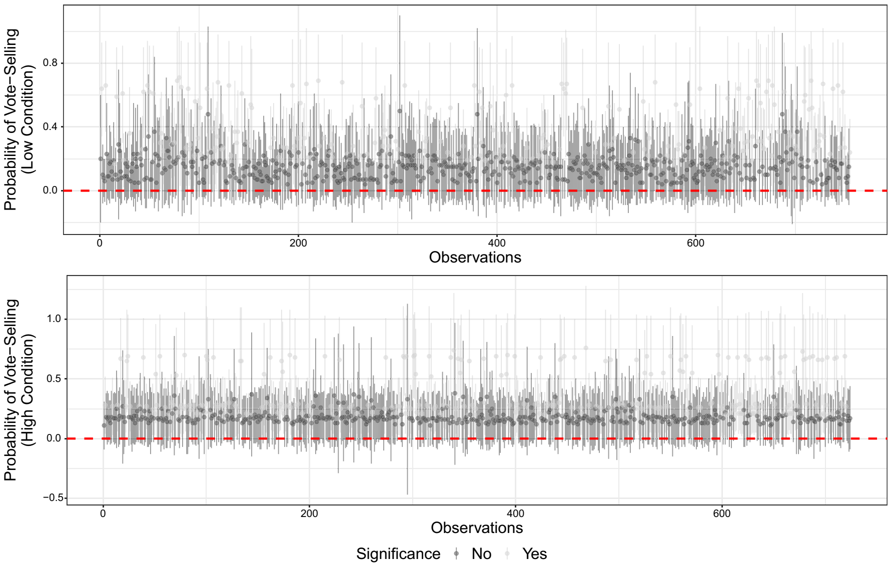


Fig. 7 Individual estimated probabilities of vote selling. *Note* figure shows the individual probabilities of vote selling ($N = 1479$) under the "low" and "high" conditions. After fitting the model, and following the advice of Blair and Imai (2012) and Imai et al. (2015), individual probabilities of vote selling under the "low" and "high" conditions were estimated. The figure also shows 95% confidence intervals

References

Abramo, Claudio, and Bruno Speck. 2001. *Report on Brazil*. Technical report.
Argersinger, Peter. 1985. New Perspectives on Election Fraud in the Gilded Age. *Political Science Quarterly* 100 (4): 669–687.
Auyero, Javier. 2000. The Logic of Clientelism in Argentina: An Ethnographic Account. *Latin American Research Review* 35 (3): 55–81.
Bahamonde, Hector. 2018. Aiming Right at You: Group Versus Individual Clientelistic Targeting in Brazil. *Journal of Politics in Latin America* 10 (2): 41–76.
Bensel, Richard. 2004. *The American Ballot Box in the Mid-nineteenth Century*. Cambridge: Cambridge University Press.
Blair, Graeme. 2015. Survey Methods for Sensitive Topics. *Comparative Newsletter* 12: 12–16.
Blair, Graeme, and Kosuke Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis* 20 (1): 47–77.
Borges, Mariana. 2019. *When Voters Help Politicians: Understanding Elections, Vote Buying, and Voting Behavior Through the Voters' Point of View*. Evanston: Northwestern University.
Bratton, Michael. 2008. Vote Buying and Violence in Nigerian Election Campaigns. *Electoral Studies* 27 (4): 621–632.
Brusco, Valeria, Marcelo Nazareno, and Susan Stokes. 2004. Vote Buying in Argentina. *Latin American Research Review* 39 (2): 66–88.
Calvo, Ernesto, and María Victoria Murillo. 2004. Who Delivers? Partisan Clients in the Argentine Electoral Market. *American Journal of Political Science* 48 (4): 742–757.
Campbell, Tracy. 2005. *Deliver the Vote: A History of Election Fraud, an American Political Tradition: 1742–2004*. New York: Carroll and Graf.
Carlin, Ryan, and Mason Moseley. 2015. Good Democrats, Bad Targets: Democratic Values and Clientelistic Vote Buying. *The Journal of Politics* 1 (77): 14–26.
Carlin, Ryan, Matthew Singer, and Elizabeth Zechmeister (eds.). 2015. *The Latin American Voter: Pursuing Representation and Accountability in Challenging Contexts*. Ann Arbor: University of Michigan Press.
Corstange, Daniel. 2008. Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis* 17 (1): 45–63.
Corstange, Daniel. 2012. Vote Trafficking in Lebanon. *International Journal of Middle East Studies* 44 (3): 483–505.
Diaz-Cayeros, Alberto, Federico Estévez, and Beatriz Magaloni. 2012. Strategies of Vote Buying: Democracy, Clientelism, and Poverty Relief in Mexico: 1–381.
Dixit, Avinash, and John Londregan. 1996. The Determinants of Success of Special Interests in Redistributive Politics. *The Journal of Politics* 58 (4): 1132–1155.
Druckman, James, Mauro Gilli, Samara Klar, and Joshua Robison. 2015. Measuring Drug and Alcohol Use Among College Student-Athletes. *Social Science Quarterly* 96 (2): 369–380.
Erie, Steven. 1990. *Rainbow's End: Irish-Americans and the Dilemmas of Urban Machine Politics, 1840–1985*. Berkeley: University of California Press.
Fahrenthold, David. 2012. Selling Votes is Common Type of Election Fraud. The Washington Post. https://www.washingtonpost.com/politics/decision2012/selling-votes-is-common-type-of-election-fraud/2012/10/01/f8f5045a-071d-11e2-81ba-ffe35a7b6542_story.html.
Finan, Rederico, and Aura Schechter. 2012. Vote-Buying and Reciprocity. *Econometrica* 80 (2): 863–881.
Foa, Roberto, and Yascha Mounk. 2016. The Danger of Deconsolidation: The Democratic Disconnect. *Journal of Democracy* 27 (3): 5–17.
Foa, Roberto, and Yascha Mounk. 2017. The Signs of Deconsolidation. *Journal of Democracy* 28 (1): 5–15.
Geddes, Barbara. 1990. How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2 (1): 131–150.
Glynn, Adam. 2013. What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. *Public Opinion Quarterly* 77 (S1): 159–172.
González-Ocantos, Ezequiel, Chad de Jonge, Carlos Meléndez, Javier Osorio, and David Nickerson. 2012. Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua. *American Journal of Political Science* 56 (1): 202–217.

González-Ocantos, Ezequiel, Chad Kiewiet de Jonge, and David Nickerson. 2014. The Conditionality of Vote-Buying Norms: Experimental Evidence from Latin America. *American Journal of Political Science* 58 (1): 197–211.

Graham, Jesse, Jonathan Haidt, and Brian Nosek. 2009. Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology* 96 (5): 1029–1046.

Gray, Kurt, Chelsea Schein, and Adrian Ward. 2014. The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion from Sin to Suffering. *Journal of Experimental Psychology* 143 (4): 1600–1615.

Hicken, Allen, Stephen Leider, Nico Ravanilla, and Dean Yang. 2015. Measuring Vote-Selling: Field Evidence from the Philippines. *American Economic Review* 105 (5): 352–356.

Hicken, Allen, Stephen Leider, Nico Ravanilla, and Dean Yang. 2018. Temptation in Vote-Selling: Evidence from a Field Experiment in the Philippines. *Journal of Development Economics* 131: 1–14.

Holland, Alisha, and Brian Palmer-Rubin. 2015. Beyond the Machine: Clientelist Brokers and Interest Organizations in Latin America. *Comparative Political Studies* 48 (9): 1186–1223.

Imai, Kosuke, Bethany Park, and Kenneth Greene. 2015. Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models. *Political Analysis* 23 (02): 180–196.

Jensen, Peter Sandholt, and Mogens Justesen. 2014. Poverty and Vote Buying: Survey-Based Evidence from Africa. *Electoral Studies* 33: 220–232.

Kam, Christopher. 2017. The Secret Ballot and the Market for Votes at 19th-Century British Elections. *Comparative Political Studies* 50 (5): 594–635.

Khemani, Stuti. 2015. Buying Votes Versus Supplying Public Services: Political Incentives to Under-invest in Pro-poor Policies. *Journal of Development Economics* 117: 84–93.

de Jonge, Chad Kiewiet. 2015. Who Lies About Electoral Gifts? *Public Opinion Quarterly* 79 (3): 710–739.

King, Gary. 1986. How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science* 30 (3): 666–687.

Kitschelt, Herbert. 2000. Linkages Between Citizens and Politicians in Democratic Polities. *Comparative Political Studies* 33 (6–7): 845–879.

Kitschelt, Herbert, and Steven Wilkinson (eds.). 2006. *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition*, vol. 392. Cambridge: Cambridge University Press.

Kuklinski, James, Paul Sniderman, Kathleen Knight, Thomas Piazza, Tetlock Philip, Gordon Lawrence, and Barbara Mellers. 1997. Racial Prejudice and Attitudes Toward Affirmative Action. *American Journal of Political Science* 41 (2): 402–419.

LaBrie, Joseph, and Mitchell Earleywine. 2000. Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique. *Journal of Sex Research* 37 (4): 321–326.

Leight, Jessica, Rohini Pande, and Laura Ralston. 2016. *Value for Money in Purchasing Votes? Vote-Buying and Voter Behavior in the Laboratory*.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die*. New York: Crown.

Levy, Jack. 2008. Case Studies: Types, Designs, and Logics of Inference. *Conflict Management and Peace Science* 25 (1): 1–18.

Michael, Adam, and Tariq Thachil. 2018. How Clients Select Brokers: Competition and Choice in India's Slums. *American Political Science Review* 112 (4): 775–791.

Morton, Rebecca, and Kenneth Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.

Nazareno, Marcelo, Valeria Brusco, and Susan Stokes. 2008. *Why Do Clientelist Parties Target the Poor?*.

Nichter, Simeon. 2014. Conceptualizing Vote Buying. *Electoral Studies* 35: 315–327.

Nichter, Simeon, and Michael Peress. 2017. Request Fulfilling: When Citizens Demand Clientelist Benefits. *Comparative Political Studies* 50 (8): 1086–1117.

Oliveros, Virginia. 2016. Making it Personal: Clientelism, Favors, and Public Administration in Argentina. *Comparative Politics* 48 (3): 373–391.

Posada-Carbó, Eduardo. 1996. *Elections Before Democracy: The History of Elections in Europe and Latin America*. London: Palgrave Macmillan.

Redlawsk, David, Caroline Tolbert, and William Franko. 2010. Voters, Emotions, and Race in 2008: Obama as the First Black President. *Political Research Quarterly* 63 (4): 875–889.

Reynolds, John. 1980. The Silent Dollar': Vote Buying in New Jersey. *New Jersey History* 98 (3): 191–211.

Reynolds, John, and Richard McCormick. 1986. Outlawing ‘Treachery’: Split Tickets and Ballot Laws in New York and New Jersey, 1880–1910. *The Journal of American History* 72 (4): 835–858.

Riley, William, and Victor Chow. 1992. Asset Allocation and Individual Risk Aversion. *Financial Analysts Journal* 48 (6): 32–37.

Rueda, Miguel. 2015. Buying Votes with Imperfect Local Knowledge and a Secret Ballot. *Journal of Theoretical Politics* 27 (3): 428–456.

Rueda, Miguel. 2017. Small Aggregates, Big Manipulation: Vote Buying Enforcement and Collective Monitoring. *American Journal of Political Science* 61 (1): 163–177.

Rusk, Jerrold. 1970. The Effect of the Australian Ballot Reform on Split Ticket Voting: 1876–1908. *The American Political Science Review* 64 (4): 1220–1238.

Sabato, Hilda. 2001. On Political Citizenship in Nineteenth-Century Latin America. *The American Historical Review* 106 (4): 1290.

Schaffer, Schaffer. 2004. Vote Buying in East Asia. Unpublished Manuscript.

Shawn, Eric. 2012. Drug Money Funds Voter Fraud in Kentucky. *Fox News* 1–9.

Stokes, Susan. 2005. Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina. *American Political Science Review* 99 (3): 315–325.

Stokes, Susan, Thad Dunning, Marcelo Nazareno, and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge: Cambridge University Press.

Szwarcberg, Mariela. 2013. The Microfoundations of Political Clientelism. Lessons from the Argentine Case. *Latin American Research Review* 48 (2): 32–54.

Vicente, Pedro. 2014. Is Vote Buying Effective? Evidence from a Field Experiment in West Africa. *The Economic Journal* 124 (574): F356–F387.

Vicente, Pedro, and Leonard Wantchekon. 2009. Clientelism and Vote Buying: Lessons from Field Experiments in African Elections. *Oxford Review of Economic Policy* 25 (2): 292–305.

Wantchekon, Leonard. 2003. Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics* 55 (April): 399–422.

Weitz-Shapiro, Rebecca. 2012. What Wins Votes: Why Some Politicians Opt Out of Clientelism. *American Journal of Political Science* 56 (3): 568–583.

Weitz-Shapiro, Rebecca. 2014. *Curbing Clientelism in Argentina: Politics, Poverty, and Social Policy*. Cambridge: Cambridge University Press.

Williamson, Samuel. 2018. Seven Ways to Compute the Relative Value of a U.S. Dollar Amount, 1774 to Present. MeasuringWorth.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.