

BOLSA FAMÍLIA AND THE SHIFT IN LULA'S ELECTORAL BASE, 2002–2006

A Reply to Bohn

Cesar Zucco

Rutgers University and Fundação Getúlio Vargas–EBAPE

Timothy J. Power

University of Oxford

Abstract: In a recent article published in the *Latin American Research Review*, Simone Bohn analyzed electoral results and survey data from Brazil to contest several theses concerning the reelection of President Luiz Inácio Lula da Silva in 2006. In particular, Bohn asserted that beneficiaries of Bolsa Família, a conditional cash transfer program that was reaching eleven million families at the time of the 2006 election, were already supporters of Lula in 2002, and therefore the program could not have contributed to the change in Lula's constituency between his election in 2002 and his reelection in 2006. We show that these claims are based on voter recall data collected between nine and fifty-seven months after the elections, and that these data grossly overestimate actual electoral support for Lula—probably as a result of well-known reporting biases. Reanalysis of Bohn's data as well as analysis of more reliable surveys suggest that there were indeed significant changes in voting patterns between 2002 and 2006, and that Bolsa Família did play an important role in the 2006 elections.

In a recent article published in the *Latin American Research Review*, "Social Policy and Vote in Brazil: Bolsa Família and the Shifts in Lula's Electoral Base," Simone Bohn (2011) analyzed electoral results and survey data from Brazil to contest several theses concerning the reelection of President Luiz Inácio Lula da Silva in 2006. Although we have many disagreements with Bohn's article, we take issue with one claim in particular, that Bolsa Família did not contribute to the change in Lula's constituency between 2002 and 2006.

What we consider to be the conventional wisdom surrounding Lula's election and reelection follows roughly these lines: Lula, founder of the Partido dos Trabalhadores (PT), ran for president and lost in 1989, 1994, and 1998. Taking advantage of an economic crisis and an uncharismatic opponent, Lula was first elected president in 2002, and he was reelected four years later. Although Lula's national vote share was almost identical in his victorious 2002 and 2006 campaigns (he

We thank Wendy Hunter, David Doyle, David Samuels, and the anonymous *LARR* reviewers for comments, the Latin American Public Opinion Project (Vanderbilt University) for conducting and making available its surveys, and the Center for the Study of Public Opinion at the University of Campinas (CESOP-UNICAMP) for cataloging and making available a wealth of public opinion data. Replication materials and a web appendix to this article with extended results are available at <http://hdl.handle.net/1902.1/18350>. The usual disclaimer applies.

Latin American Research Review, Vol. 48, No. 2. © 2013 by the Latin American Studies Association.

received 61 percent of the vote in the respective runoffs), there is evidence that his constituency changed considerably in the intervening years. While the “old” Lula was not particularly strong in the less developed regions of Brazil and was backed (predominantly) by the urban middle classes, the Lula of 2006 was reelected with massive support in poorer areas.

Claims of a shift in Lula’s electoral base have been put forth in several recent articles written independently and using various methods of aggregate data analysis (e.g., Hunter and Power 2007; Nicolau and Peixoto 2007; Soares and Terron 2008; Zucco 2008). Several scholars have provided evidence suggesting that one of the causes of this shift was the considerable expansion of Bolsa Família (BF), a conditional cash transfer program that had reached eleven million families by the eve of the 2006 election. The authors in question have not claimed that BF is the only driver of this change. Renewed (and socially inclusive) economic growth, rising real wages among the poor, incumbency advantages, and differential attentiveness to corruption allegations have all been suggested as possible causes of the geographic and demographic changes in Lula’s electoral profile. But there is a broad consensus in the political science literature that BF did play a significant and independent role in this shift.

Bohn strongly contests this view. Using survey data from the Latin American Public Opinion Project (LAPOP), she claims to show that Bolsa Família beneficiaries in 2006, at the time of Lula’s reelection, were already Lula supporters in 2002, when he first won the presidency.¹ Hence, Bohn concludes that there was no major shift in Lula’s constituency between 2002 and 2006, and that consequently, BF could not have had any meaningful effect on the 2006 presidential election. In this response to her article, we show that the results Bohn finds for the impact of BF on Lula’s electoral performance do not stand up to scrutiny. Hence, her article offers no factual basis on which to reject the current interpretation concerning the electoral role of Bolsa Família.

We note, from the outset, that we do not contest Bohn’s more general points that the analysis of aggregate electoral data can be subject to ecological fallacies and that to advance knowledge, it is important to analyze individual-level data as well. In fact, several of the articles that Bohn criticizes have relied on a combination of survey and aggregate data analysis, a point she does not acknowledge.² We

1. In the interest of precision here, Bohn understands as 2006 beneficiaries those who reported receiving a BF stipend during the LAPOP survey conducted in July–August 2007, about nine or ten months after the election. However, since the program had already reached nearly 100 percent coverage of eligible families by the time of the presidential election in October 2006, and because post-rollout changes to the BF registry (e.g., deaths, purges) are few and slow to be implemented, this operationalization is defensible.

2. We note that other authors had already applied individual-level data to these questions. Using survey data, Balbachevsky and Holzhaecker (2007) found that Lula’s 2002 support was better explained by ideology than by social class, but that by 2006 social class had become extremely important and ideology had diminished in explanatory power. Lício, Rennó, and Castro (2009) used LAPOP data from a later wave in 2008 to show that BF was a major contributor to presidential approval among recipients. Their analysis shows sensitivity to the problem of using recall data for analyzing electoral preferences in 2006.

also note that our present article is not intended to be an exercise in quantitative methods, regression coefficients, or any other type of "statistical hairsplitting." The essence of our criticism is much simpler: we hold that opinion surveys are not *unconditionally* better than aggregate electoral data and that not all surveys serve all purposes. In general, voter recall data (i.e., asking survey respondents to report how they voted in the past) collected months or years after the relevant election should not be used to analyze the determinants of voting behavior.

In terms of substance, one might question whether it is really that important to know whether Lula's electoral base shifted from 2002 to 2006. Does the 2006 Brazilian presidential election really merit all the number crunching and spilled ink that have been expended on it already? We believe that the answer to both questions is yes. Most observers of Brazilian politics hold that the 2006 election was much more than a routine reelection of a popular incumbent, and they take this view precisely *because* Lula's electoral base shifted after 2002. For example, in a series of influential recent essays, André Singer (2009, 2010, 2012) has interpreted the last three presidential elections in Brazil as an electoral "realignment" that is emplacing a sort of a Rooseveltian New Deal in Brazil—a progressive multiclass political coalition captained by the PT, intent on pro-poor policies and growth, and with a renewed role for the state. Paraphrasing Singer, Perry Anderson (2011, 7) recently wrote, "Lula's victories in 2002 and 2006 can be mapped with uncanny closeness onto Roosevelt's of 1932 and 1936." If, following the New Deal parallel, the PT's victory in 2002 was a "breakthrough" election and the victory in 2006 was a "realigning" election, then it is clearly incumbent upon scholars to document the intervening changes in Lula's electoral constituency. The victory of Lula's chosen successor, Dilma Rousseff, in 2010 makes a sustained evaluation of the Singer realignment thesis all the more necessary. Although our article does not resolve the issue of whether this is the correct interpretation of recent events, it does help get the basic facts straight.

We first attempt to replicate the analysis Bohn presents in table 3 of her article, which uses "recall" data collected in 2007 to analyze the two preceding electoral cycles. We report being able to replicate her results for the 2002 election but not for the 2006 elections, and we also present alternative specifications that lend further credence to our efforts at direct replication. Next, we take a step back to consider whether recall data should be used for this type of analysis at all. We show that voter recall data from LAPOP's 2007 Brazil survey do not resemble the actual electoral results, nor do they coincide with other surveys taken closer in time to both elections—two problems that should be expected given what social scientists know about recall data. Frankly, some heroic assumptions would have to be made in order to use these data to analyze interelection shifts in individual behavior. We examine briefly the limited information that more trustworthy surveys can provide about the shift in support for Lula between 2002 and 2006. Finally, we conclude with a brief analysis of the role of BF in the 2006 elections, using a different survey that was taken closer to the day of the election and that better postdicted the official results. Our results corroborate the view that Bolsa Família beneficiaries voted for Lula at higher rates than similarly situated nonbeneficiaries in 2006.

REPLICATION AND ALTERNATIVE SPECIFICATIONS

Direct Replication

We first attempted to replicate the results presented in Bohn's table 3, in which the author estimates the association between being a Bolsa Família (BF) beneficiary at the time of the LAPOP survey (July–August 2007) and recalling a prior vote for Lula in 2002 and 2006. Bohn finds a positive and significant effect of BF on the Lula 2002 vote, but no significant effect in 2006, when the coefficient is actually negative. We note that in the case of nonlinear models, one should typically report relevant quantities of interest rather than a table of coefficients (King, Tomz, and Wittenberg 2000). However, for the sake of direct comparability with the results presented by Bohn, we present our replication coefficients in the first two columns of table 1, with additional specifications in the rightmost columns.³

For the 2002 elections, we found a coefficient for BF participation that is fairly close to the one reported in Bohn's table 3. We obtained this by using the respondent's declared vote for Lula in the second round of 2002 election as the dependent variable (scored as 1), and we lumped together those that claimed to have cast invalid ballots or to have voted for a candidate other than Lula (scored as 0). The main independent variable was operationalized as a binary indicator, taking the value of 1 when the respondent claimed to personally receive a BF stipend from the federal government. We considered nonrespondents and nonvoters as missing observations.

Although we attempted to follow the specifications in the Bohn article, our results for several predictors of the 2002 Lula vote are slightly inconsistent with hers: neither age nor the dummy for the Northeast region are significantly different from zero. These discrepancies may be due to different treatment of missing observations on the vote variable, given that our results are reported from a subsample with an N of 531 (larger than the 448 reported by Bohn).⁴

Presentation is important here. Because Bohn's table 3 focuses only on coefficients, one important issue that goes unnoticed in her article is that the predicted probabilities of voting for Lula are *extremely* high. Keeping in mind that our sample was not exactly the same as the one used by Bohn, about 75 percent of the observations of the dependent variable in our model were scored as 1—in other words, three-quarters of the LAPOP respondents who remembered having cast a valid ballot in the second round of the 2002 election claimed to have done so for Lula. This is quite at variance with the actual results of the 2002 elections, in which Lula received 61.3 percent of the valid vote in the runoff.

Turning to models estimating the 2006 results, we were never able to find a negative coefficient for BF, which is what Bohn reports in the second column of

3. All analysis in this article was performed in R, using commands available in the Zelig package, by Kosuke Imai, Gary King, and Olivia Lau (Zelig: Everyone's Statistical Software, R package version 3.4–8). Complete replication code is available at the website Cesar Zucco Dataverse, <http://hdl.handle.net/1902.1/18350>.

4. In the web appendix we provide a very detailed explanation of how we arrived at this sample.

her table 3. In fact, throughout several alternative specifications, we found a positive and (marginally) significant “effect” of BF participation on support for Lula in 2006. In contrast to Bohn’s results, neither gender nor the dummy variable for the South are significant. However, we do find significant coefficients for income, education, and the Northeast region, which are consistent with the results Bohn reported.

Once again, our N was larger than the one reported in the Bohn article (878 instead of 800), and it is possible that with exactly the same sample our results would have converged more closely with Bohn’s. Our initial attempt at direct replication suggests, however, that the results for the BF effect on Lula support seem to depend on very specific coding decisions that we could not reproduce.⁵

Alternative Specifications

Momentarily suspending judgment on the quality of the survey data itself—an issue to which we return later—it is both possible and justifiable to attempt alternative specifications of the models. Age, income, and education, for instance, are ordered categorical variables as defined by Bohn but were treated as continuous variables in her regression. This procedure treats differences between arbitrarily defined categories as being the same and assumes that the independent variable in question is linearly related to the dependent variable (given the link function). We correct this by converting education and age to continuous scales using actual years of age and years of schooling (available in the survey).

We treated income as the categorical variable that it is, entering it into our regression as a series of dummy variables.⁶ We also code income differently than the procedure used by Bohn. Instead of following Bohn and creating a separate category for the eighteen respondents earning twenty or more minimum wages, we merged all those respondents with self-reported incomes of greater than ten minimum wages into a single category.⁷

When we estimate the model with our revised categories of income and our alternative treatment of age and education, we find that BF participation in 2007 continues to be a positive predictor of the Lula vote in 2002 and 2006, although it falls short of statistical significance in the latter year. This is consistent with what we found in our replication attempt without the corrections for the categorical variables, but it is different from what Bohn reports in her article. Under

5. Because we could not replicate Bohn’s results for 2006, in the web appendix we document in detail how we arrived at the 878 observations in our sample. In essence, the 1,214 responses are reduced to 930 after eliminating nonvoters and invalid answers. Of these, we lose 52 observations because of missing responses in the income variable and to a lesser extent in BF participation, as well as age. This leaves us with 878 usable observations. Although we could have imputed these missing data (King et al. 2001), our focus here is on replicating the results Bohn obtained.

6. One could have followed the latter approach for all these variables. However, given that a natural linear unit exists for both age and education, we opted to simplify the analysis.

7. The original coding decision is especially problematic because Bohn treated the variable as a continuous five-point scale. The result implies the placing of much more weight on the few members of the wealthiest category in detriment to the rest of the sample.

this specification, it is noteworthy that having controlled for the other variables, higher-income respondents are not significantly less likely to have voted for Lula than respondents in the lowest income bracket.

We also estimated exactly the same model after preprocessing the sample to correct for covariate imbalance. This is particularly important because eligibility for the BF program is contingent on family income, which in turn correlates with several other variables included in our estimations. Bolsa Família recipients are different from nonrecipients in many respects, and not only because of the monthly benefit they receive. Ideally, for any analysis conducted at the individual level, correct inference of a causal BF effect on the Lula vote should not be based solely on the comparison between beneficiaries and nonbeneficiaries, but rather on a comparison between Brazilian voters who are similar in many ways but who differ only in their receipt of a BF stipend. Although an experimental design would be ideal for this task, it is not possible. A second-best approach is provided by nonparametric matching (Ho et al. 2007), which basically entails discarding data that do not contribute to the identification of the causal effect.⁸ Using this procedure, we find that the coefficient for BF is also positive and significant in both election years (2002 and 2006).⁹

Thus, we have undertaken three types of analysis using the LAPOP 2007 survey: (1) a direct attempt to replicate Bohn's results, (2) a correction for the use of categorical independent variables in the regressions, and (3) a new estimation using nonparametric matching to address covariate imbalance. All three analyses point in the same direction: participation in the Bolsa Família program appears to be positively associated with individual-level support—or recollection thereof—for Lula's election in both 2002 and 2006.

Again, we believe that it is important to highlight predicted probabilities. Our results for 2006, for the three models reported in table 1, show that BF participation leads to a relative increase in the probability of voting for Lula of 15 percent, 13 percent, and 17 percent, respectively, over a baseline probability of approximately 0.70. The last of these estimates is derived from the model estimated with the preprocessed data set, which is the most appropriate method for isolating a causal effect of Bolsa Família participation on voting behavior in 2006, at the individual level, using the present survey.

In summary, our reanalyses of the LAPOP 2007 survey corroborate the more conventional view of Lula's reelection story in 2006, contradicting the results reported by Bohn. Yet the reader will note that our reanalyses also apparently confirm Bohn's boldest claim: that (future) "Bolsa Família recipients were already Lula voters in 2002" (70). This last claim, if credible, would seriously defy the conventional wisdom that Lula's constituency changed between 2002 and 2006

8. We matched beneficiaries to nonbeneficiaries on income and region (exact matches), as well as on age, gender, and education (nearest-neighbor matches). We allowed for matching with replacement and a control-to-treatment ratio of 2. Matching was performed using MatchIt for R (Ho et al. 2011). The exact matching means that the sample of beneficiaries looks exactly like the sample of nonbeneficiaries with respect to income and region. On the other variables, the two groups are considerably more similar after matching.

9. After matching, we are left without respondents in the higher income bracket.

Table 1 Regression (probit) estimates for vote for Lula in 2002 and 2006 using LAPOP 2007

	Our attempt at replication		Correction for categorical variables		Balanced data set	
	2002	2006	2002	2006	2002	2006
Bolsa Família (beneficiary = 1)	0.512	0.322	0.498	0.288	0.450	0.371
<i>p</i> value	0.027	0.074	0.034	0.112	0.094	0.076
Gender (male = 1)	0.007	0.119	0.011	0.117	0.023	-0.069
	0.956	0.213	0.931	0.220	0.929	0.741
Age ^a	-0.017	-0.096	-0.004	-0.010	0.004	-0.009
	0.753	0.015	0.388	0.005	0.704	0.243
Income (5 point)	0.060	-0.101				
	0.405	0.071				
Income (2–5 MW)			0.043	-0.040	0.059	-0.278
			0.787	0.749	0.843	0.219
Income (5–10 MW)			0.151	-0.23		0.032
			0.464	0.133		0.961
Income (10 MW or more)			0.212	-0.247		
			0.394	0.206		
Schooling ^a	-0.078	-0.308	-0.019	-0.060	-0.003	-0.038
	0.391	<0.001	0.280	<0.001	0.937	0.216
Center-West	-0.243	0.240	-0.248	0.208	0.322	0.833
	0.282	0.180	0.272	0.245	0.534	0.083
North	0.383	0.416	0.382	0.411	-0.024	-0.171
	0.188	0.072	0.191	0.075	0.973	0.747
Northeast	0.243	0.406	0.25	0.407	0.190	0.494
	0.119	0.001	0.113	0.001	0.560	0.039
South	0.034	-0.121	0.054	-0.095	0.591	0.642
	0.852	0.355	0.773	0.471	0.318	0.054
Intercept	0.697	1.716	0.824	1.383	0.419	1.154
	0.053	<0.001	0.013	<0.001	0.516	0.014
<i>N</i>	531	878	531	878	145	242

Source: LAPOP Brazil 2007 data set.

Note: Dependent variable is a binary indicator of the recalled vote in the runoff of the 2002 and 2006 presidential elections, where 1 equals a declared vote for Lula and 0 equals a vote for Alckmin or an invalid ballot. The *p* values are reported beneath the coefficients. Income brackets refer to minimum wage (MW).

^aAge and schooling are defined differently in the first two columns, as explained in the text.

because of BF. But although the claim is consistent with the data Bohn used, the data themselves are inconsistent with reality.

We would not stake causal claims on the LAPOP recall data, not even for the part of the findings that supports our position. Our skepticism is not driven by the relatively borderline levels of statistical significance we found in the coefficient of interest. Nor is it driven by the fact that (as we report in the web appen-

dix) when we use first-round results, or when we adopt other definitions of BF participation available in the survey, it is possible to find nonsignificant (though never negative) coefficients on BF. Our main point is simply that there is strong evidence that the LAPOP 2007 survey has little external validity when it comes to voting behavior and that any results for vote recall taken from this survey should not be trusted. This is not, we should stress, an indictment of the survey or of the broader LAPOP project: LAPOP has generated a considerable amount of high-quality data that help illuminate a host of important questions that can be systematically studied in a comparative fashion. But simply put, any survey that includes recall questions so long after the actual elections were held will run into problems. As we detail in the next sections, data-quality issues invalidate Bohn's results, especially those based on vote recall in the 2002 elections, when Lula first captured the presidency.

DISCREPANCIES BETWEEN LAPOP 2007 AND THE REAL WORLD

Surveys provide a glimpse into the mind of voters that is invaluable because we can observe actual voting behavior only at the aggregate level. We agree with Bohn on two fundamental points: we believe that it is very important to use individual-level data to the extent possible, and we recognize the limitations of using aggregate data in drawing causal inferences. However, it is still necessary to examine whether stated preferences bear any resemblance to what can actually be observed in the real world. In the case of surveys that ask voters to recall how they voted in the past, the problems can be rather severe. Here we remind the reader that Bohn's method involves using interviews conducted in July and August 2007 to ask people how they voted not only in October 2006 (nine to ten months earlier) but also in October 2002 (nearly five years in the past), and then drawing inferences about Lula's electoral support based on the intertemporal consistency of their responses.

Yet social scientists have known for decades that such recall data are notoriously unreliable.¹⁰ Numerous studies have documented that a sizable minority of survey respondents misreport or in some cases invent responses to conceal electoral nonparticipation. Himmelweit, Biberian, and Stockdale (1978) have shown that error in reporting past votes is generally not random but tends to be linked to one of three factors: (1) a desire to demonstrate political engagement, given the social desirability of participation in most democracies;¹¹ (2) a wish to show

10. In addition to the citations in the text, see Dex (1995) for an interesting literature review concerning problems with recall data in a number of allied disciplines, including psychology and economics. The tendency of unemployed persons to understate their time out of work and the tendency of parents to exaggerate the earliness at which their children began to walk and talk are common examples that distort employment data and childhood development surveys, respectively. For the weakness of recall data in measuring party identification, see Katz, Niemi, and Newman (1980) and Niemi, Katz, and Newman (1980).

11. We observe that much of the relevant research reviewed by Himmelweit, Biberian, and Stockdale (1978) was conducted in democracies where turnout is voluntary. In Brazil, where voting is mandatory for most social segments, admitting abstention to a survey interviewer could be even more problematic. In addition to the social desirability of voting, we must add the complicating factor of legal sanctions

consistency with behavior at the time of recall; and (3) association with success, meaning that respondents disproportionately wish to identify with recent electoral victors or popular incumbents. In their review of the phenomenon, Presser and Traugott (1992, 78) conclude bluntly that “virtually all the error is in the socially desirable direction.”

To these three factors we can add a fourth complication: the presence of floating voters. Problems with recall reliability were first identified via panel studies in older democracies, which—despite their relatively stable party systems—routinely have error rates in interelection recall ranging from 10 percent to 25 percent (Plumb 1986, 303). The classic British panel study by Benewick and colleagues (1969) found that strong partisans (those who tend to vote for the same party at each election) tend to have highly accurate rates of recall (greater than 90 percent), whereas “floating voters” have error rates of 50 percent or more. The authors attribute the poor recall rates of floating voters to a “reluctance to admit having behaved differently” in the past (Benewick et al. 1969, 186). Their study used data from a panel survey in which the same respondents were interviewed during the 1959 general election in the United Kingdom and then were reinterviewed during the next election in 1964. After calculating the recall error, the authors warn, “If we had not had the 1959 results and had drawn conclusions about the floating vote between elections simply on the basis of the 1964 interviews, these conclusions would have been practically worthless” (Benewick et al. 1969, 186).

Thus, error in recall is generally not random, and the literature on the reliability of recall data suggests at least four reasons we should be cautious about drawing inferences using the LAPOP 2007 survey. First, voters may want to make their recalled 2002 vote appear consistent with their declared 2006 vote. Second, those who abstained or did not cast a valid vote in the two elections may have an incentive to recall a valid ballot, given the social desirability of electoral participation. Third, at the time of the survey Lula had been reelected for four more years as president and enjoyed overwhelming public approval,¹² which could bias some respondents into “remembering” a vote for a popular incumbent.¹³ Fourth,

against nonvoting, no matter how weak those sanctions may be in practice (for details, see Power 2009). As Jackman (1987, 409) argues, compulsory voting laws are best described as a “disincentive to nonvoting” and have effects even in the absence of strong enforcement.

12. President Lula was hugely popular at the time of the survey. We examined three national polls taken by DataFolha, IBOPE, and Sensus between June and August 2007, around the time that LAPOP was in the field. The three surveys had almost identical results. Averaging the three, some 49 percent of Brazilians rated Lula's performance as *ótimo* or *bom* (positive evaluations), 35 percent as *regular* (neutral evaluations), and only 15 percent as *ruim* or *péssimo* (negative evaluations). See the website Pesquisas de Opinião, <http://noticias.uol.com.br/politica/pesquisas/>.

13. A favorite classroom anecdote here involves a survey taken by the National Opinion Research Center (University of Chicago) in the days following President John F. Kennedy's assassination in November 1963. In the poll, some 65 percent of respondents recalled having voted for Kennedy in the 1960 election, compared to 30 percent who said they had supported his Republican opponent, Richard M. Nixon—in reality, Kennedy had received a mere 50.08 percent of the two-party vote in that historically close contest. While many are familiar with this story and naturally attribute the poll results to post-assassination sympathy for Kennedy, few now recollect that an earlier survey by the center taken in June 1963 (when JFK was alive and well) showed that respondents were already “remembering” prior support for Kennedy to the tune of 59 percent (Sheatsley and Feldman 1965, 152). In other words, recall

Table 2 Voter recall and actual results for the 2006 presidential election

	First round, October 1, 2006		Second round, October 29, 2006	
	Actual results	LAPOP recall in July–August 2007	Actual results	LAPOP recall in July–August 2007
Lula (PT)	44.5	64.5	57.2	67.6
Geraldo Alckmin (PSDB)	38.1	19.1	36.8	21.1
Heloísa Helena (PSOL)	6.2	3.1	—	—
Cristovam Buarque (PDT)	2.4	0.9	—	—
Others	0.2	0.1	—	—
Blank and/or null	8.4	3.5	6.0	3.7
Missing ^a		8.7		7.6
Abstention ^b	16.8	16.1	19.0	17.1

Note: Electoral results are from the Tribunal Superior Eleitoral, and we report shares relative to total votes cast. Survey results from LAPOP 2007. PT = Partido dos Trabalhadores; PSDB = Partido da Social Democracia Brasileira; PSOL = Partido Socialismo e Liberdade; PDT = Partido Democrático Trabalhista. ^aMissing refers to those respondents who claim that they voted but for whom a preference was not registered or who recalled voting for a candidate not on the ballot. See the web appendix for further details. ^bAbstention is measured relative to enrolled voters for actual electoral results and to the whole sample for survey results.

low levels of party identification could introduce additional error into the survey results. If the magnitude of recall error in survey research is proportional to the number of floating voters in the electorate, as Benewick and colleagues (1969) found for midcentury Britain, then the reliability of recall data may be even more questionable in a society with weak partisanship such as Brazil.¹⁴ Unfortunately, we have no way to test this final idea with the data set in question, since LAPOP Brazil was not a panel study.

With these reservations in mind, we examine how closely the LAPOP and alternative surveys parallel the actual electoral results in 2002 and 2006. For the sake of simplicity, we observe the following conventions to report actual results and survey estimates and to make them comparable. For election results, we report shares relative to total votes cast, which includes votes for candidates and null and/or blank votes—referred to as “invalid.” We also report abstention figures relative to registered voters, but these do not enter any computation of vote shares. For opinion surveys, we treat those who declared not having voted as abstentions and exclude them from the computations. Of the remaining voters, we report as “missing” those who claim to have voted but who either declined to name a candidate or said they did not know or could not remember.

The results for 2006 are reported side by side in table 2. As expected, there is strong evidence of recall error. Lula’s declared support in the second round

of a supposed Kennedy vote in 1960 had already increased more in the first two years of his presidency than it subsequently did in the immediate aftermath of his death.

14. On the nature of mass partisanship in Brazil, see Samuels (2006). On determinants of vote choice using a valuable panel study in two Brazilian cities, see Baker, Ames, and Rennó (2006).

Table 3 Voter recall and actual results in the 2002 presidential elections

	First round, October 6, 2002		Second round, October 27, 2002	
	Actual results	LAPOP recall in July–August 2007	Actual results	LAPOP recall in July–August 2007
Lula (PT)	41.6	41.6	57.6	51.0
José Serra (PSDB)	20.8	9.8	36.4	14.3
Ciro Gomes (PPS)	10.8	1.4	—	—
Anthony Garotinho (PSB)	16.0	0.9	—	—
Others	0.4	0.2	—	—
Invalid	10.4	1.2	6.0	1.5
Missing ^a	—	44.9	—	33.3
Abstention ^b	21.6	21.7	25.7	30.6

Note: Electoral results are from the TSE, and we report shares relative to total votes cast. Survey results from LAPOP 2007. PT = Partido dos Trabalhadores; PSDB = Partido da Social Democracia Brasileira; PPS = Partido Popular Socialista; PSB = Partido Socialista Brasileiro.

^aMissing refers to those respondents who said that they voted but for whom no vote recall was recorded, or who recalled voting for a candidate not on the ballot. See the web appendix for further details

^bAbstention is measured relative to enrolled voters for actual electoral results, and to the whole sample for survey results.

is 10 percent higher than the official election results, and in the first round it is 20 percent higher. The punch line from table 2 is that even if every single respondent in the “missing” category were to suddenly declare a vote for opposition candidate Geraldo Alckmin (Partido da Social Democracia Brasileira, or PSDB), he would still fail to reach even 27 percent of reported votes in the first round, and Lula would still boast 64 percent of retrospective preferences—a major departure from the actual results. The same principle applies in the runoff. Even if we assumed that Alckmin was the choice of everyone who did not remember his or her vote, or who wrongly recalled having voted for a candidate not on the ballot, we would still overstate Lula’s vote share considerably.

If the LAPOP results diverge so strongly from the actual returns in the 2006 election (which had occurred nine or ten months earlier), what to say about voter recall for the 2002 election? A priori, one should have a healthy skepticism about any question that asks respondents to recall a vote cast nearly five years earlier. Moreover, in the intervening years, Lula underwent profound changes in his public image, from being the untested leader of the leftist opposition in 2002 to being a popular centrist incumbent—supported by more than ten parties in Congress—in 2007. The recall results for 2002 are presented in table 3.

At first glance, recall for Lula in 2002 is exactly on the mark for the first round, and table 3 looks quite reassuring if one reads only the “Lula” row horizontally. However, a vertical scan of the table quickly reveals the immense size of the “missing” category, which reaches more than 40 percent of all respondents for the first round of the election, and one-third of respondents in the runoff. How do these

missing responses affect the estimated support for Lula in 2002? To accept that Lula's support in the first round of 2002 was 41.6 percent among LAPOP respondents, one would have to make two giant leaps of analytical faith. First, one would have to assume that *all* respondents who did not recall for whom they voted (the missing category) did *not* vote for Lula. Second, one would have to accept that two potent alternatives in 2002, Ciro Gomes and Anthony Garotinho (who together received more than a quarter of the first-round vote), were supported by a paltry 2 percent of LAPOP respondents. One reason for the huge "missing" category and for the poor recall of eliminated first-round candidates is rather obvious: the vote-recall questions were not "stimulated" items on the LAPOP questionnaire (i.e., interviewers were instructed not to read the names of candidates to respondents). With no stimulation, the "also-rans" in 2002 are virtually absent from collective memory nearly five years later, thus distorting the results.

The conventional treatment of missing data, and what Bohn appears to have done, is to drop respondents for whom vote recall is not recorded.¹⁵ When this is done, Lula's share of valid observations climbs to 77 percent in both rounds. On the basis of this lopsided and implausible sample, Bohn advances her claim that BF participation in 2007 is associated with past support for Lula in 2002.

While the LAPOP surveys are generally well conceived and executed, the 2007 Brazil module, in particular, has some very serious shortcomings when it comes to making inferences about electoral behavior. This does not mean that the survey is "wrong" in any fundamental way, but simply that it is not useful for the task for which Bohn employed it. In a nutshell, given what we know from tables 2 and 3, many more respondents declared having voted for Lula than actually did in both elections.

Bolsa Família and the LAPOP Questionnaire

Returning to the central question of whether conditional cash transfers affected electoral support for Lula, there were two questions related to Bolsa Família in the 2007 LAPOP survey. One asked whether the individual respondent participated in the program, and another asked whether somebody in the family of the respondent or somebody he or she knew participated in BF. This poses some conceptual difficulties from the outset. The program is generally targeted at households, not individuals, so the first question is somewhat restrictive. The second question, however, can lump together someone who is (possibly) in the household with people who are (possibly) not in the household, thus creating a double imprecision. While we accept that "knowing somebody who participates in the program" is a relevant variable, we believe that it should not be lumped together with direct participation. Finally, there is the issue that only persons of voting age (a minority of beneficiaries) could respond to the LAPOP survey. Since the BF benefit is paid to heads of household, and these are almost always voting-

15. We infer this from the contrasting number of observations reported in Bohn's table 3, which is 800 in 2006 but only 448 in 2002.

age adults, LAPOP respondents who reply yes to the first question are more likely than not the actual heads of household themselves (this, at any rate, is how most Brazilian adults would interpret a question about direct participation in the BF program).¹⁶ Thus, a respondent who replies no to the first question but yes to the second question is most likely referring to a member of another household, not his or her own household.¹⁷

Not surprisingly, the first question generates an estimate of program participation that seems too low when compared to the official enrollment in the program in 2006, whereas the second question produces an estimate that seems too high. We know that there were 11.1 million families in the program by the end of 2006, which corresponds to approximately 19.3 percent of the total number of families in Brazil.¹⁸ It follows that even if there were only one voting-age adult per BF family (a conservative assumption), then solid sampling techniques and accurate responses should have yielded a direct participation rate in the vicinity of 19 percent as well, although under more realistic assumptions the number would be somewhat higher.¹⁹ However, in LAPOP 2007, only 10 percent of respondents declared that they were direct beneficiaries of BF. We also know that the Ministry of Social Development works with an estimate of 4.1 persons per BF household, which if correct would suggest that about 24 percent of the entire population lived in BF households in late 2006. In contrast, the second LAPOP question leads us to believe that 45 percent had a family member, friend, or an acquaintance enrolled in the program.²⁰

Clearly, it is difficult to ascertain what is going on behind these discrepant numbers. Discrepancies could be a result of imprecision in the two BF survey questions, sampling issues, misreporting by respondents, or some combination of all three of these factors. In the context of Bohn's quantitative analysis, this amounts to combining a problematic independent variable (the first BF question on LAPOP) with a questionable dependent variable (recall of a Lula vote). For all the known weaknesses of the earlier analyses using aggregate data, the two

16. Brazilians can register to vote (and can participate in LAPOP surveys) from the age of sixteen. A Brazilian in his or her late teens could be a registered voter, answer the LAPOP survey, and yet still be considered a dependent in a BF-supported family, but such cases would be few and far between.

17. This no-yes combination is, in fact, the pattern for more than a third of respondents. For the record, 96 individuals responded yes to both BF questions, and 615 responded no to both. Some 444 respondents responded no to the first question about personal participation but yes to the second about knowing someone benefiting from BF. Another 32 individuals said they were direct beneficiaries but did not know other beneficiaries.

18. The Brazilian census bureau (Instituto Brasileiro de Geografia e Estatística, or IBGE) estimated that there were 57.5 million families in Brazil in 2006 and that the midyear population count was 186.8 million. This works out to an average family size of 3.25 individuals. In estimating the average size of BF families at 4.1, the Ministry of Social Development is estimating these families to be 20 percent larger than the mean family size in Brazil.

19. Assuming one voting-age adult per BF family is very conservative. The real number should be higher because of two-parent families, young dependents of voting age (e.g., sixteen- and seventeen-year-olds), and the dependent elderly in households.

20. These figures are calculated after excluding the few missing responses (0.4 percent to the first BF question and 2.0 percent to the second).

analogous variables in those published studies (Ministry of Social Development statistics on BF penetration and Supreme Electoral Court data on the election results, respectively) are simply far more reliable.

WHO VOTED FOR LULA IN 2002 AND 2006?

Can the LAPOP 2007 data tell us anything concrete about continuity and changes in patterns of support for Lula? We believe not. We have shown that the data do not match what we know to be true about electoral results and BF coverage, and we have conjectured that well-known reporting biases are part of the problem. To strengthen this point, in this section, we compare LAPOP 2007 to other surveys taken closer to the 2006 and 2002 elections, surveys that provide different, albeit limited, insight into the changes in Lula's constituency.

For each election year, we selected three surveys undertaken by different polling organizations. Since only one of these other surveys asked a BF-related question, we momentarily leave Bolsa Família aside and turn first to a simple analysis of the predicted probability of voting for Lula in the first round of both the 2002 and 2006 presidential elections. Our main explanatory variable was income, and our logit regressions controlled for gender, age, schooling, and region. To facilitate the analysis, we present the predicted vote for Lula by income bracket graphically, computed for a female, in the Southeast of the country, with modal education and age.

Results show that while all other surveys generate extremely similar findings, LAPOP 2007 presents some discrepancies for the 2006 electoral results and considerably large discrepancies for 2002. This is not that surprising given what we have presented in tables 1, 2, and 3, but figure 1 is striking.

According to the other surveys, higher-income voters supported Lula at similar rates in both elections, and he gained significant ground among the poorest voters. More specifically, the probability of voting for Lula in the lowest income bracket in each survey ranges from 0.51 to 0.58 in the three other surveys for 2006, which is a considerable and significant increase from the 0.33–0.37 range in 2002. The LAPOP 2007 survey, however, actually shows higher levels of support among low-income voters in 2002 than 2006, though the difference between the two years is not very large. This is particularly consequential if one considers that approximately 75 percent of Brazilian voters had family incomes below five minimum wages.

As we suggested earlier, the puzzling LAPOP results are probably an artifact of respondents attempting to project consistency in their answers when asked in mid-2007 whom they had voted for in the past two elections. Unfortunately, we cannot use these more reliable data to say for sure how the same individuals behaved in the two elections. We can, however, shed some additional light on the changes in voting patterns between 2002 and 2006.

In this simple exercise, we identified the cities and towns in which survey respondents lived and coded those *municípios* as either "high development" or "low development," based on whether their values of the Human Development Index

Table 4 Surveys compared to LAPOP 2007

Survey	Name in Figures	CESOP code	Days from first round	Sample size
2002 Datafolha	2002 DAT	1692	-8	6,030
2002 IBOPE	2002 IBO	1796	-10	3,000
2002 Criterium ^a	2002 CRIT	1823	-4	3,020
2006 Datafolha	2006 DAT	3364	-1	14,798
2006 IBOPE	2006 IBO	2358	-25	2,002
2006 Vox Populi	2006 VOX	2488	+16	2,005

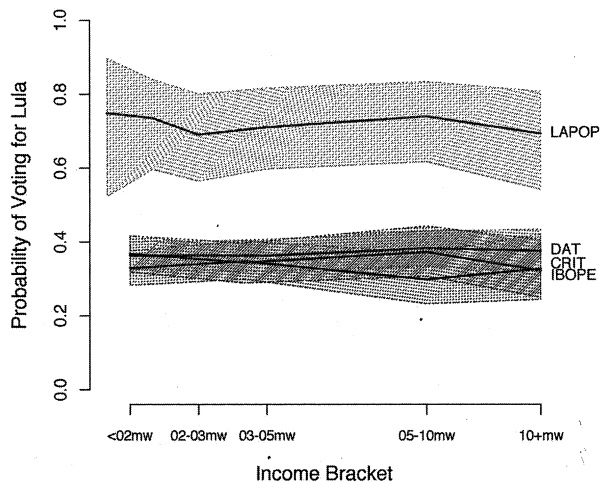
^aWe do not have the exact dates for this survey, but we know it was fielded in October before the first round of the elections, which makes the possible range of dates between -1 and -4. All surveys are deposited and available upon request from the Center for the Study of Public Opinion (CESOP), University of Campinas (<http://www.cesop.unicamp.br>). The CESOP internal codes for each survey are reported to facilitate requests. Some surveys had both stimulated and open-ended vote intention or recall questions, but the question we analyzed was always a stimulated one. We focus on first-round results because those surveys were readily available. All comparisons are with LAPOP's recall questions for the first round of each election.

at the municipal level (HDI-M) were above or below the median value in each sample.²¹ We also singled out respondents who reported family income equal to or lower than two minimum wages. Figure 2 shows vote intention or recall for Lula for the four categories produced by these two binary indicators, with 95 percent confidence intervals surrounding the estimates.

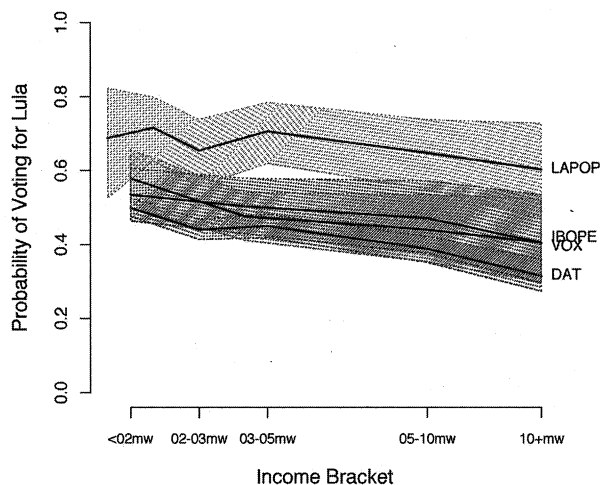
Figure 2 shows that both surveys for each year yield very similar results. Furthermore, there is evidence that Lula's constituency changed between 2002 and 2006. This change is particularly pronounced in the increase in support for Lula among the poorest individuals living in the least developed places, and to a lesser degree in the decline of support for Lula among the nonpoor in more developed places. However, even the lack of significant movement observed in the other two categories is not incontrovertible evidence that no change happened inside these groups as well. As these are not panel data, it is possible that changes in different directions within the group offset one another, giving the impression of no change. Still, some information can be inferred from these figures.

We assume here (and this is a strong, yet we believe justified, assumption) that by averaging the two surveys for each election year, we have arrived at a good estimate of the "real" support for Lula in each subgroup of voters—certainly a far better estimate than is extractable from LAPOP. It follows, then, that at least half of the poorest people in the least developed places who did not vote for Lula in 2002 chose to support him in 2006. Conversely, it also follows that a minimum

21. Surveys conducted by the Brazilian Institute of Public Opinion and Statistics (Instituto Brasileiro de Opinião Pública e Estatística, or IBOPE) did not contain information that allowed for identification of the municipality, so they were not included in this analysis. The HDI-M was computed only once by the IBGE, in 2000, but as it is a combination of indicators of income, health, and education, it is a better measure of development than simply income. The median HDI-M varied from 0.77 to 0.79 across the four surveys. The median value across all municipalities in the country was 0.71.



A. 2002 elections



B. 2006 elections

Figure 1 Comparing LAPOP 2007 to other surveys. The figures show predicted vote for Lula as estimated from different surveys taken close to each election by Datafolha, Critérium, IBOPE, and Vox Populi (for details, see table 4) and as recalled in LAPOP 2007. In all surveys, predicted probabilities were computed from a logit regression of a binary indicator of vote for Lula in the first round of the respective election on family income bracket, education, age, gender, and region. Respondents who did not state a preference or did not vote were dropped. Respondents who cast null or blank votes were coded as non-Lula votes, as were voters who cast a vote for any other candidate on the ballot. Confidence intervals of 95 percent are shown about the estimates.

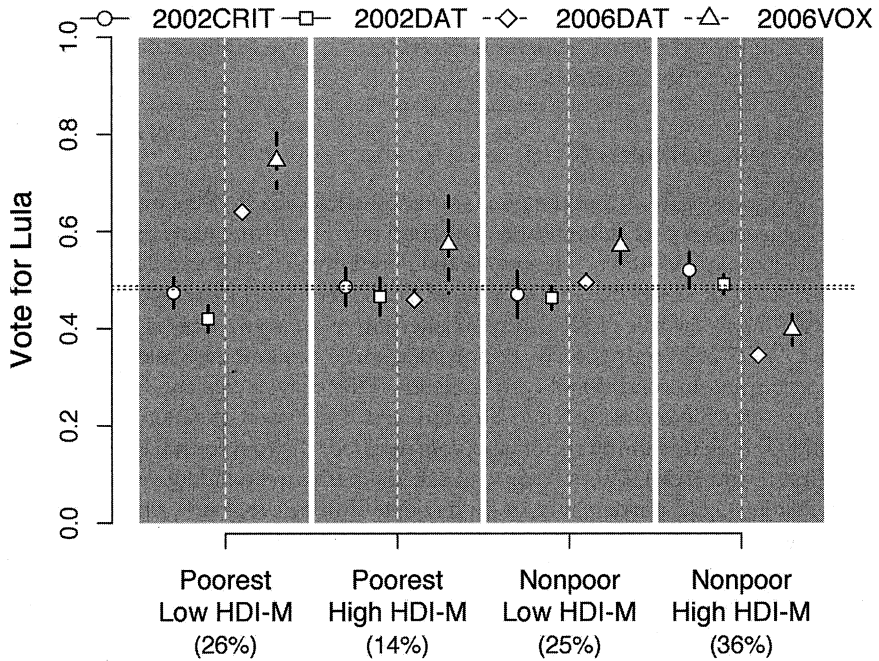


Figure 2 Vote for Lula by income and level of development of the municipality, 2002–2006. The figure reports estimates of vote recall and/or intention for Lula by categories of income and level of development of the municipality in which respondents live. Data are from the four non-IBOPE surveys listed in table 4, which were fielded close to the 2002 and 2006 elections and contain information that identifies respondents' municipality. Some observations (never more than 5 percent) were lost as a result of missing responses in the municipal identifier variable. Horizontal dashed lines represent vote intention or recall for Lula in the full sample. HDI-M = municipal-level Human Development Index.

of slightly less than 30 percent of 2002 Lula supporters among the nonpoor in more developed places deserted him in 2006.²² These are, it should be stressed, the lower bounds on “switches.” For both voter categories, these percentages of floating voters could be considerably higher. Even using these conservative figures, Lula's vote base did shift considerably from 2002 to 2006.

This analysis could be further improved by pooling all surveys taken in each year and examining smaller categories of voters, but we leave this path for future research.²³ For now, the evidence we have is that Lula's support among the poor-

22. These are not estimates but deterministic information contained in the data. We do assume, however, that the data represent the population. For an intuitive discussion of the idea of bounds in ecological inference problems, see King (1997); for tomography plots that support these claims, see the web appendix.

23. For more details, see Gelman and Little (1997) on multilevel regression and poststratification.

est in the poorer municipalities definitely increased, whereas support among the nonpoor in the richest municipalities decreased, and we cannot rule out further changes in the other categories.

THE ROLE OF BOLSA FAMÍLIA IN THE 2006 ELECTIONS

Bohn agrees with us that the poor shifted toward Lula in 2006, but she argues that BF played no part in this shift. Although our preceding analysis shows that Lula greatly expanded his vote among the poor between 2002 and 2006, it is true that not all poor voters benefit from BF, so in this final section, we close the circle by returning to the role of BF. Of the six alternative surveys we obtained to cross-check LAPOP 2007, only the Vox Populi 2006 survey contained a relevant survey item pertaining to conditional cash transfers. This survey was conducted only a few days after the inconclusive first round, and it included questions about vote intention for the upcoming runoff and vote recall for the first round. Here we analyze first-round recall. This survey also asked whether respondents or somebody in their household participated in BF or other similar programs, and then, in a separate question, it asked whether the person knew somebody that participated in these programs. (By 2006, BF had already incorporated the preexisting conditional cash transfer programs, so this wording does not pose too much difficulty.²⁴) The first of these questions estimated coverage to be approximately 22 percent, which is close to what we should expect nationally and a more realistic estimate than the similar question in LAPOP. Following the same criteria as before, we dropped respondents who either did not say for whom they voted or said they did not vote at all. We also lumped together those who cast a null or blank vote and those who said they voted for other candidates. The recall question that asked for which candidate respondents had voted in the first round of the 2006 election gave Lula 51 percent and Alckmin 36 percent, which is reasonably close to the actual results reported in first column of table 2.²⁵ Analysis of this survey (using exactly the same approaches and variables reported in the right-most columns in table 1) reveals a positive and significant coefficient for BF, which is persistent under any specification, and which implies that beneficiaries were 40 percent more likely to vote for Lula than similar nonbeneficiaries (table 5).

As with any observational study, our inference is subject to the existence of unobservable variables that could be associated both with receiving the BF benefit and with voting for Lula. These could be individual-level characteristics that drive self-selection into the program or structural characteristics that make the program more available to some potential beneficiaries than to others. Even though we have matched recipients to nonrecipients on several observable variables that cover most of the plausible selection mechanisms, we cannot rule out

24. For details on how Bolsa Família amalgamated several previous conditional cash transfer programs (including the innovative Bolsa Escola), see Hall (2006) and Fenwick (2009).

25. In the Vox Populi 2006 survey, Lula surpassed his real-world results by 6.5 percentage points and Alckmin underperformed by 2.1 points, for a total deviation of 8.6 points. By way of contrast, the analogous total deviation in the LAPOP 2007 survey was an astounding 39 points.

Table 5 Regression (probit) estimates for vote for Lula in 2006 using Vox 2006

	Full sample	Preprocessed using MatchIt
Bolsa Família (beneficiary = 1)	0.601	0.969
<i>p</i> value	0.000	0.000
Education (grade 5–8)	–0.117	0.018
	0.170	0.932
Education (grade 9–11)	–0.280	–0.306
	0.001	0.158
Education (some college or more)	–0.513	–0.981
	0.000	0.067
Age	0.001	–0.004
	0.532	0.477
Center-West	0.134	0.162
	0.255	0.637
Northeast	0.567	0.906
	0.000	0.000
North	0.078	–0.010
	0.518	0.970
South	–0.166	–0.448
	0.062	0.095
Income (1–5 MW)	–0.200	–0.341
	0.020	0.046
Income (5–10 MW)	–0.264	–0.411
	0.023	0.257
Income (10 or more MW)	–0.369	–0.770
	0.010	0.334
Gender (male = 1)	0.069	–0.067
	0.251	0.670
Intercept	0.058	0.405
	0.690	0.238
<i>N</i>	1925	852

Source: Data from Vox Populi survey 2006.

Note: Dependent variable was a binary indicator of the recalled vote in the first round of the 2006 presidential election, where 1 equals a declared vote for Lula and 0 equals a vote for other candidates or an invalid ballot. The *p* values are reported beneath the coefficients. Income brackets refer to minimum wages (MW).

the possibility that some selection process is still biasing the results. However, sensitivity analysis suggests that any such hidden bias would have to be very strong to reverse these results.

Sensitivity analysis asks, “How would inferences about treatment effects be altered by hidden biases of various magnitudes?” (Rosenbaum 2002, 106). The standard metric is to report how changes in the odds of receiving treatment (expressed as Γ) can affect the *p* value of the inference. With random assignment of treatment, the odds of two observations receiving treatment would be the same, and Γ would be 1. If *p* values were to fall below conventional levels of significance with Γ s close to 1, it would mean that results are sensitive to small deviations from

random assignment. We performed Rosenbaum's sensitivity analysis for binary outcomes and found that results would still be significant with Γ s as large as 1.85.²⁶ This means that omitted variable bias would have to be strong enough to make treated observations almost twice as likely to receive treatment as the correspondent control observation in the matched set for the result not to hold.

Although we believe that further controls and the examination of specific selection biases can be warranted, our results seem not only robust but also compatible with estimates of the BF effect from the aggregate analyses found in other studies. In summary, these results cast further doubt on the interpretation advanced by Bohn.

CONCLUSIONS

If we are to understand electoral shifts (i.e., alignments, realignments, and dealignments), we believe that it is important to combine different sources of data using data triangulation strategies and multimethod approaches. We agree with Bohn that analysis of aggregate electoral data should be complemented with individual-level data, but not any individual-level data will do. Survey data are stated preferences, and there are many reasons that lead people to make biased statements. We should always be concerned with this, and if we are to take triangulation and multimethod approaches seriously, we have to be selective about our data sources. The data that were used in Bohn's recent study are simply not compatible with the real-world results of the 2002 and 2006 presidential elections in Brazil.

With respect to who voted for Lula in his last two campaigns, the preponderance of the evidence points to a change in the rate in which the poorest voters supported Lula in 2002 and in 2006. Although we do not analyze Dilma Rousseff's (PT) victory in 2010, we are confident that the PT's presidential vote bases in 2006 and 2010 will be found to be much more similar to each other than they are to Lula's "breakthrough" base in 2002.

For the present, we agree with Singer (2012) that conditional cash transfers are likely here to stay and that the patterns of support for Lula indicate a change in his constituency. We believe, however, that the jury is still out on the meaning and durability of this new constituency, and there is some evidence that the electoral effects of CCTs might be mostly short term (Zucco 2013). Singer's argument about a new social pact is compatible with the data, but so are other explanations that stress retrospective voting (at least partially driven by BF) or incumbency advantages. Future election results will allow us to distinguish these alternative interpretations.

As for the specific issue of the electoral effects of BF in 2006, the present article will not be the last word on the subject. However, most evidence to date suggests that there is an independent pro-incumbent effect of conditional cash transfers.

26. To perform the analysis, we used *rbounds: Perform Rosenbaum bounds sensitivity tests for matched and unmatched data*, by Luke J. Keele (2011), R package version 0.9, <http://CRAN.R-project.org/package=rbounds>.

Separate analysis of survey data (Zucco 2013) shows estimates of BF electoral effects that are compatible with observational electoral results in 2006 and 2010. But we must come to terms with the fact that we do not have survey data on BF's antecedent programs for 2002, when Fernando Henrique Cardoso (PSDB) still occupied the office of the presidency. The path forward, if any understanding is to be had of the role of social policy in the late Cardoso period, is to make creative use of observational data and to search for better research designs (see, for inspiration, De Janvry, Finan, and Sadoulet 2008).

REFERENCES

- Anderson, Perry
2011 "Lula's Brazil." *London Review of Books* 33, no. 7 (March 31): 3–12.
- Baker, Andy, Barry Ames, and Lucio R. Rennó
2006 "Social Context and Campaign Volatility in New Democracies: Networks and Neighborhoods in Brazil's 2002 Elections." *American Journal of Political Science* 50 (2): 379–396.
- Balbachevsky, Elisabeth, and Denilde Holzacker
2007 "Classe, ideologia e política: Uma interpretação dos resultados das eleições de 2002 e 2006." *Opinião Pública* 13 (2): 283–306.
- Benewick, R. J., A. H. Birch, J. G. Blumler, and Alison Ewbank
1969 "The Floating Voter and the Liberal View of Representation." *Political Studies* 17 (2): 177–195.
- Bohn, Simone R.
2011 "Social Policy and Vote in Brazil: Bolsa Família and the Shifts in Lula's Electoral Base." *Latin American Research Review* 46 (1): 54–79.
- De Janvry, Alain, Frederico Finan, and Elisabeth Sadoulet
2008 "Local Electoral Accountability and Decentralized Program Performance." Unpublished manuscript. <http://are.berkeley.edu/~sadoulet/papers/Bolsa7-08.pdf>.
- Dex, Shirley
1995 "The Reliability of Recall Data: A Literature Review." *Bulletin de Méthodologie Sociologique* 49 (December): 58–90.
- Fenwick, Tracy Beck
2009 "Avoiding Governors: The Success of Bolsa Família." *Latin American Research Review* 44 (1): 102–131.
- Gelman, Andrew, and Thomas C. Little
1997 "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23 (2): 127–135.
- Hall, Anthony
2006 "From Fome Zero to Bolsa Família: Social Policies and Poverty Alleviation under Lula." *Journal of Latin American Studies* 38 (4): 689–709.
- Himmelweit, Hilde T., Marianne Jaeger Biberian, and Janet Stockdale
1978 "Memory for Past Vote: Implications of a Study of Bias in Recall." *British Journal of Political Science* 8 (3): 365–375.
- Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart
2007 "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.
2011 "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42 (8): 1–28.
- Hunter, Wendy, and Timothy J. Power
2007 "Rewarding Lula: Executive Power, Social Policy, and the Brazilian Elections of 2006." *Latin American Politics and Society* 49 (1): 1–30.
- Jackman, Robert W.
1987 "Political Institutions and Voter Turnout in the Industrial Democracies." *American Political Science Review* 81 (2): 405–423.

- Katz, Richard S., Richard G. Niemi, and David Newman
 1980 "Reconstructing Past Partisanship in Britain." *British Journal of Political Science* 10 (4): 505–515.
- King, Gary
 1997 *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.
- King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve
 2001 "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95 (1): 49–69.
- King, Gary, Michael Tomz, and Jason Wittenberg
 2000 "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 341–355.
- Lício, Elaine Cristina, Lucio Rennó, and Henrique Castro
 2009 "Bolsa Família e voto nas eleições presidenciais de 2006: Em busca do elo perdido." *Opinião Pública* 15 (1): 31–54.
- Nicolau, Jairo, and Vitor Peixoto
 2007 "Uma disputa em três tempos: Uma análise das bases municipais das eleições presidenciais de 2006." Paper presented at the annual meeting of the National Association of Research and Postgraduate Studies in Social Sciences (ANPOCS), Caxambu, Minas Gerais, Brazil, October 22–26.
- Niemi, Richard G., Richard S. Katz, and David Newman
 1980 "Reconstructing Past Partisanship: The Failure of the Party Identification Recall Questions." *American Journal of Political Science* 24 (4): 633–651.
- Plumb, Elizabeth
 1986 "Validation of Voter Recall: Time of Electoral Decision Making." *Political Behavior* 8 (4): 302–312.
- Power, Timothy J.
 2009 "Compulsory for Whom? Mandatory Voting and Electoral Participation in Brazil, 1986–2006." *Journal of Politics in Latin America* 1 (1): 97–122.
- Presser, Stanley, and Michael Traugott
 1992 "Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting." *Public Opinion Quarterly* 56 (1): 77–86.
- Rosenbaum, Paul R.
 2002 *Observational Studies*. 2nd ed. New York: Springer.
- Samuels, David
 2006 "Sources of Mass Partisanship in Brazil." *Latin American Politics and Society* 48 (2): 1–27.
- Sheatsley, Paul B., and Jacob J. Feldman
 1965 "A National Survey on Public Reactions and Behavior." In *The Kennedy Assassination and the American Public*, edited by Bradley S. Greenberg and Edwin B. Parker, 149–177. Stanford, CA: Stanford University Press.
- Singer, André
 2009 "Raízes sociais e ideológicas do lulismo." *Novos Estudos* 85 (3): 83–102.
 2010 "O lulismo e seu futuro." *Revista Piauí*, no. 49 (October).
 2012 *Os sentidos do lulismo*. São Paulo: Companhia das Letras.
- Soares, Gláucio, and Sonia Terron
 2008 "Dois Lulas: A geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise geoespacial)." *Opinião Pública* 14 (2): 269–301.
- Zucco, Cesar
 2008 "The President's 'New' Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections." *Journal of Latin American Studies* 40 (1): 29–49.
 2013 "When Payouts Pay Off: Conditional Cash Transfers and Voting Behavior in Brazil, 2002–2010." *American Journal of Political Science*. doi: 10.1111/ajps. 12026.