A Statistical Model of Abstention under Compulsory Voting

Gabriel Katz

Abstract

Invalid voting and electoral absenteeism are two important sources of abstention in compulsory voting systems. Previous studies in this area have not considered the correlation between both variables and ignored the compositional nature of the data, potentially leading to unfeasible results and discarding helpful information from an inferential standpoint. In order to overcome these problems, this paper develops a statistical model that accounts for the compositional and hierarchical structure of the data and addresses robustness concerns raised by the use of small samples that are typical in the literature. The model is applied to analyze invalid voting and electoral absenteeism in Brazilian legislative elections between 1945 and 2006 via MCMC simulations. The results show considerable differences in the determinants of both forms of non-voting; while invalid voting was strongly positively related both to political protest and to the existence of important informational barriers to voting, the influence of these variables on absenteeism is less evident. Comparisons based on posterior simulations indicate that the model developed in this paper fits the dataset better than several alternative modeling approaches and leads to different substantive conclusions regarding the effect of different predictors on the both sources of abstention.

KEYWORDS: compulsory voting, invalid voting, abstention, compositional data

Acknowledgments: I am indebted to Timothy Power and Timmons Roberts for providing their data on Brazil's legislative elections. I also thank J. Laurent Rosenthal, R. Michael Alvarez, Jonathan Katz and Robert Sherman for their invaluable comments.

1. Introduction

The desire to provide a political system with popular legitimacy and to increase the representativeness of elected public officials have often been asserted as major arguments justifying the imposition of compulsory voting provisions (Verba et al., 1978; Hill, 2002). Twenty-four countries, comprising approximately 20% of the world's democracies, employ mandatory voting to some extent (Australian Joint Standing Committee on Electoral Matters, 2000). Although compulsory voting has been found to be an effective mechanism for increasing turnout (Hirczy, 1994; Lijphart 1997), compelling voters to go to the polls does not automatically mean that they will cast a ballot for one of the candidates. Citizens can cast invalid votes, i.e., blank or null ballots, and thus their right not to vote remains intact (Lijphart, 1997). In fact, a long-standing feature of compulsory voting systems is a higher rate of invalid ballots (Hirczy, 1994). In addition, since mandatory voting does not generate universal compliance (Hirczy, 1994; Power and Roberts, 1995), illegal abstention constitutes a second form of non-voting.

Previous research on compulsory voting systems has focused either on the determinants of electoral absenteeism (e.g., Hirczy, 1994) or on the determinants of invalid voting (Power and Garand, 2007). The common approach of studies in this area has been to treat either the proportion of invalid votes or electoral absenteeism as the dependent variable and regress each of them on a set of explanatory variables. This standard procedure exhibits two main shortcomings. First, it does not take into account the connection between both sources of non-voting and the relationship between their determinants. Since, under compulsory voting, invalid voting and electoral absenteeism can be seen as "functional equivalents" of abstention, jointly modeling them may contribute to better understand abstention and its causes. Moreover, without a model for exploring the interrelation between these two sources of abstention, helpful information from an inferential standpoint maybe discarded because their correlation is assumed to be zero. As shown by Zellner (1971) and Thum (1997), changes in the standard error estimates that might result from a bivariate model could substantially modify the conclusions drawn from separate univariate analyses. Second, the prevailing modeling strategy ignores the "compositional" nature of the data (Aitchison, 1986), i.e., the fact that the proportions of invalid ballots, electoral absenteeism and votes for candidates or parties cannot be negative and that their sum must equal one. Ignoring these non-negativity and unit-sum constraints might lead to unfeasible results, such as negative percentages of invalid ballots or sums of proportions greater or less than one (Katz and King, 1999).

This paper develops a statistical model to address these problems, jointly analyzing the determinants of invalid voting and electoral absenteeism in district-level elections. While national-level studies have the advantage of allowing more countries in the analysis, they are generally based on a small number of observations and may fail to capture the contextual and "neighborhood" effects that typically have considerable influence in local (e.g., legislative) elections (King, 1997; Katz and King, 1999). In addition, given the absence of survey data covering large historical periods in many of the countries with compulsory voting, most of which are recently democratized Latin American nations (International Institute for Democracy and Electoral Assistance, IDEA, 2007), district-level elections allow studying both sources of abstention at the lowest possible level of aggregation.

However, analyzing district-level elections introduces an additional methodological challenge. The proportion of invalid votes and absenteeism may be influenced not only by local variables but also by country-level factors affecting all districts in a given election (Power and Roberts, 1995), violating the standard assumption of independent and identically distributed errors. Hence, ignoring the hierarchical structure of the data and simply pooling national- and district-level variables may result in inefficient parameter estimates and negatively biased standard errors, potentially leading to "spuriously significant" statistical effects (Antweiler, 2001; Maas and Hox, 2004).

Drawing on the literature on compositional data (Aitchison, 1986; Katz and King, 1999), and on multi-level modeling (Goldstein, 1995; Bryk and Raudenbush, 2002; Gelman and Hill, 2007), the model presented here relates both sources of abstention in compulsory-voting systems, accounting for the compositional and hierarchical structure of the data and addressing robustness concerns raised by the use of small samples that are typical in the literature. I illustrate the use of the model analyzing data on invalid voting and electoral absenteeism in Brazil's lower house elections at the state level. Brazil has the largest electorate in the world subject to compulsory voting and has experienced considerable variations in institutional, political and socioeconomic conditions across history and between states, therefore providing an illuminating case to examine rival explanations of invalid voting and absenteeism. The percentage of blank and null ballots in the country has been historically larger and more volatile than in most other democracies with compulsory voting (Instituto Universitario de Pesquisas de Rio de Janeiro, IUPERJ, 2006; IDEA, 2007), and absenteeism has remained relatively high despite mandatory voting.

Power and Roberts (1995) used ordinary least square pooled time-series regressions to separately analyze the determinants of the two sources of abstention in legislative elections between 1945 and 1990, combining country-level and state-level predictors by assigning the national variables to each state. I extend the period of analysis to include all the elections held up to 2006 and compare the results of the model developed in this paper with those obtained from other modeling strategies that that fail to account for the compositional and/or the hierarchical structure of the data. Based on posterior simulations, I show that the compositional-hierarchical model leads to different substantive conclusions and fits the data better than these alternative approaches.

The remainder of the paper is organized as follows. Section 2 briefly reviews alternative theories for explaining invalid voting and absenteeism under compulsory voting systems. Section 3 presents the compositional-hierarchical model developed in this paper to analyze the determinants of invalid voting and absenteeism at the district level. Section 4 applies the model to analyze 16 lower house elections in Brazil and compares the performance of the compositional-hierarchical model with competing approaches. Finally, Section 5 concludes.

2. Alternative explanations of invalid voting and absenteeism

Drawing on the literature on voter turnout in industrialized democracies, three basic explanations, focusing on socioeconomic factors, on institutional variables, and on “protest voting”, have been proposed to account for invalid voting and absenteeism in compulsory voting systems (Power and Roberts, 1995; Fornos et al., 2004; Power and Garand, 2007).

Some scholars have argued that the high rate of blank and null ballots in polities with mandatory voting reflects the alienation of citizens from the political system and is the consequence of mobilizing disinterested and poorly informed citizens who would otherwise abstain (Jackman, 2001). Previous analyses (Power and Roberts, 1995; Power and Garand, 2007) found that socioeconomic variables such as urbanization, literacy and education levels substantially affect the percentage of blank and null ballots cast through their effect on voters’ perceived efficacy, access to information and development of political. Although the literature on electoral behavior has also found a strong correlation between these variables and political participation in voluntary voting settings (Verba et al., 1978;

Rosenstone and Hansen, 1993), empirical evidence from countries with mandatory voting (Power and Roberts, 1995; Fornos et al., 2004) suggests that the impact of socioeconomic factors on electoral absenteeism in these democracies is quite moderate.

Other authors have underscored the role of the institutional context, party system and electoral design in explaining invalid voting and absenteeism. For instance, Blais and Dobrzynska (1998) concluded that a higher number of political parties depresses turnout by increasing the unpredictability of election and policy outcomes, and the same would apply for highly disproportional systems that punish minor parties and reduce voters' perceived efficacy (Jackman and Miller, 1995). In the same direction, Power and Roberts (1995) provide evidence that variables like district magnitude and ballot structures have a considerable impact on invalid voting in mandatory voting settings.

Finally, an alternative explanation can be traced to the literature on protest voting. A protest vote can be defined as a vote primarily cast to express discontent with politics, rather than to affect public policies (Kitschelt, 1995). In a system of compulsory voting, citizens' discontent with the political establishment would translate into higher percentages of null and blank ballots and illegal abstention. This interpretation has often been quoted in Brazil and Latin America to explain temporary increases in invalid voting and absenteeism (Moisés, 1993; Escobar et al., 2002).

Although the socioeconomic, institutional and protest approaches are usually presented as competing rather than as complementary explanations, previous research (Power and Roberts, 1995; Fornos et al., 2004) has shown that fusing them in a combined model helps to better understand the phenomena under study. However, these approaches are grounded in the literature on political participation in developed democracies, where invalid voting has received little academic attention (Power and Garand, 2007). Thus, past work has made no theoretical distinctions regarding the effect of the different sets of variables on invalid voting and electoral absenteeism. The underlying assumption in previous analyses has been that the same basic causal mechanisms account for both forms of non-voting (Power and Roberts, 1995). Furthermore, from a methodological perspective, previous studies failed to examine the potential interactions between the determinants of these two sources of abstention, implicitly assuming that the effect of the relevant predictors on invalid voting is independent of their impact on absenteeism. The statistical model presented in the next section allows me to test these assumptions.

3. A statistical model of abstention under compulsory voting

The model used to analyze the determinants of invalid voting and absenteeism at the district level is grounded in the literature on "compositional data" (Aitchison, 1986; Katz and King, 1999) and on Bayesian hierarchical modeling (e.g., Gelman and Hill, 2007), although it is modified and adapted to the problem under study.

Let $P_{i,t}^{I}, P_{i,t}^{A}$ and $P_{i,t}^{V}$ denote the proportion of invalid votes, electoral absenteeism and valid votes (i.e., votes for candidates or parties) among the electorate in district i at election t, $i = 1,2 \ldots. n$, , $t = 1,2 \ldots T$. For all i and t, $P_{i,t}^{I}$, $P_{i,t}^{A}$ and $P_{i,t}^{V}$ must satisfy the non-negativity and unit-sum constraint $P_{i,t}^{s} \in [0,1]$, $s = I, A, V$ and $P_{i,t}^{I} + P_{i,t}^{A} + P_{i,t}^{V} = 1$. These constraints determine that $P_{i,t}^{I}, P_{i,t}^{A}$ and $P_{i,t}^{V}$ fall in the simplex space, as illustrated by the ternary plot in Figure 1. Each circle in the figure indicates the values of P^{I}, P^{A} and P^{V} in a particular district for all lower house elections held in Brazil between 1945 and 2006.

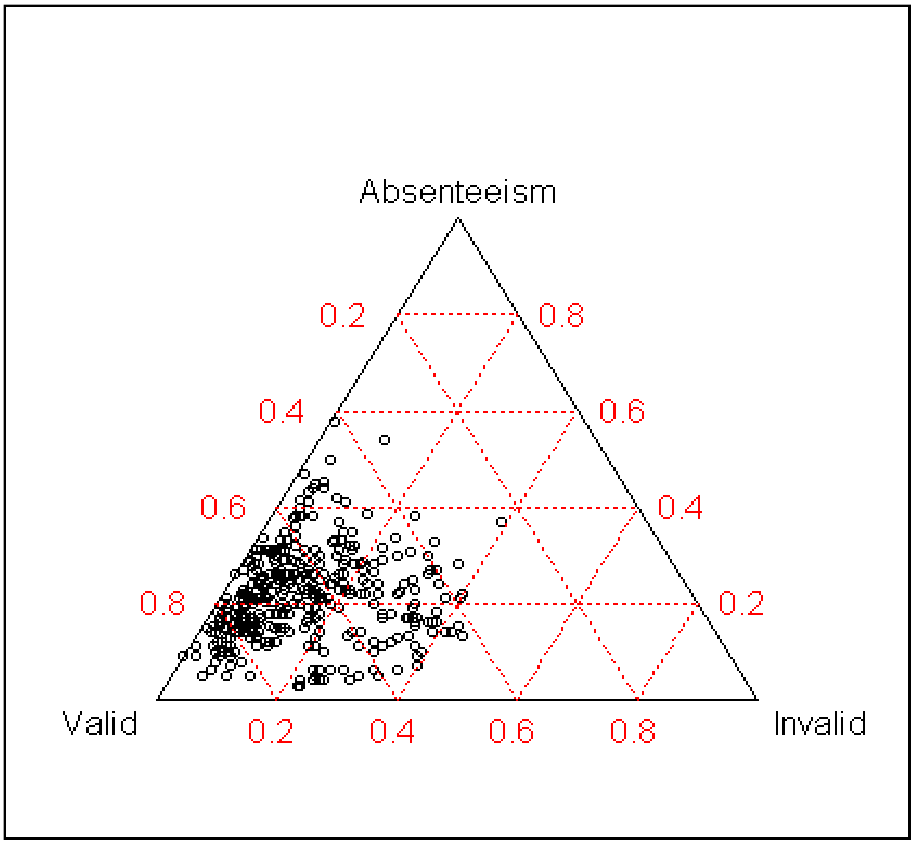

Figure 1: Proportion of electoral absenteeism and valid/invalid votes in Brazil's lower house elections between 1945 and 2006. The diagonal lines parallel to the triangle's left side indicate the proportion of invalid votes, measured on the scale in the triangle's base. The diagonal lines parallel to the right side mark the proportion of valid votes, measured on the scale in the triangle's left side, and the dashed horizontal lines indicate the proportion of electoral absenteeism, measured on the triangle's right side.

A model aimed at analyzing the determinants of abstention in compulsory voting systems must take these constraints into account. Neither the standard approach of regressing invalid voting and absenteeism independently on a set of predictors nor estimating a system of seemingly unrelated equations satisfies both constraints, even if the point predictions obtained from these models happen to fall within the boundaries of the simplex (Katz and King, 1999). In order to address this problem, I adapt Aitchison's (1986) and Katz and King's (1999) models for compositional data, implementing a bivariate mixed model for invalid voting and electoral absenteeism.

Let $Y^I_{i,t} = \ln\left(P^I_{i,t} / P^V_{i,t}\right)$ and $Y^A_{i,t} = \ln\left(P^A_{i,t} / P^V_{i,t}\right)$ denote the log-ratios of the proportion of invalid votes and absenteeism relative to valid votes, respectively.[1] Note that, unlike the baseline composites $P^I_{i,t}, P^A_{i,t}$ and $P^V_{i,t}$, $Y^I_{i,t}$ and $Y^A_{i,t}$ are unbounded and unconstrained. The variables of interest for the analysis, $P^I_{i,t}, P^A_{i,t}$, are obtained from $Y_{i,t} = \left[Y^I_{i,t}, Y^A_{i,t}\right]$ through the additive logistic transformations:

$$P^I_{i,t} = \frac{\exp\left[Y^I_{i,t}\right]}{1+\exp\left[Y^I_{i,t}\right]+\exp\left[Y^A_{i,t}\right]} \quad (1),$$

$$P^A_{i,t} = \frac{\exp\left[Y^A_{i,t}\right]}{1+\exp\left[Y^I_{i,t}\right]+\exp\left[Y^A_{i,t}\right]} \quad (2).$$

Since the $Y^s_{i,t}$, $s = I, A$, are defined over the whole real line, it is possible to model $Y_{i,t} = \left[Y^I_{i,t}, Y^A_{i,t}\right]$ using a normal/independent distribution (Seltzer et al., 2002) that assigns weight parameters to each observation in the sample:

$$Y_{i,t} = \mu_{i,t} + \frac{\varepsilon_{i,t}}{\sqrt{w_{i,t}}} \quad (3),$$

[1] Due to the logarithmic transformations involved, the baseline composites are assumed to be strictly positive. This poses no problem for the type of electoral data considered here. Alternative models based on Box-Cox transformations (e.g., Rayens and Srinivasan, 1991) have been proposed to deal with the potential problem of null composites in other settings.

where $\mu_{i,t} = \left[\mu_{i,t}^{I}, \mu_{i,t}^{A}\right]'$, $\varepsilon_{i,t} = \left[\varepsilon_{i,t}^{I}, \varepsilon_{i,t}^{A}\right]' \sim \mathrm{N}(0, \Sigma)$, $w_{i,t}$ is a positive random variable with density $p(w_{i,t} | \upsilon)$, and υ is a scalar or vector-valued parameter. Due to the unconstrained properties of Σ, the model now allows for any pattern of dependency between $P_{i,t}^{I}$ and $P_{i,t}^{A}$. This is one of the advantages of the normal/independent assumption *vis-à-vis* other distributions (e.g., Dirichlet or S^{-}) commonly used in the statistical analysis of compositional (Barndorff-Nielsen and Jørgensen, 1991). It is also worth noting that, besides including the bivariate normal as a particular case (when $w_{i,t} = 1 \; \forall i,t$), the specification in (3) comprises a variety of thick-tailed distributions often useful for robust inference (Rosa et al., 2003).

Since $\mu_{i,t}^{I}$ and $\mu_{i,t}^{A}$ are unbounded, it is possible to model them as linear functions of relevant covariates. As mentioned in the introduction, it seems plausible that the proportion of invalid votes and electoral absenteeism in a district is influenced not only by district-level variables but also by national conditions that vary across elections. Furthermore, the impact of these district-level variables on invalid voting and absenteeism might itself be mediated by the country-level factors. In order to account for these possibilities, I use a hierarchical random-coefficients model for the components of $\mu_{i,t}$. The first-level equations model $\mu_{i,t}^{I}$ and $\mu_{i,t}^{A}$ as functions of district-level variables measured at a particular election. The second-level equations specify the first-level coefficients as functions of country-level variables measured contemporaneously with the district level variables, plus zero-expectation random effects aimed at accounting for election-to-election variability beyond that explained by national-level variables. In addition, I also introduce zero-mean random intercepts in order to accommodate time-constant heterogeneity across districts. This modeling strategy strikes a balance between a completely pooled approach, which ignores the clustered nature of the data and the potential variability between districts and elections, and local regressions that are likely to be highly unstable when the number of districts or elections in the sample is relatively small (Browne and Draper, 2001).

Letting $x_{i,t}$ and z_t represent $(1 \times K)$ and $(1 \times L)$ row vectors of district - level and country-level variables, respectively, the specification adopted is then:

$$\mu_{i,t} = X_{i,t}\beta_t + \lambda_i \qquad (4),$$

$$\beta_t = Z_t\delta + \eta_t \qquad (5),$$

where $X_{i,t} = \left[I_2 \middle| x_{i,t} \otimes I_2 \right]$; $Z_t = I_{2(K+1)} \otimes \left[1 \middle| z_t \right]$; β_t and δ are $2(K+1)$ and $2(K+1)(L+1)$ vectors or random and fixed coefficients, respectively; and $\eta_t \sim N(0, \Omega_\eta)$, $\lambda_i \sim N(0, \Omega_\lambda)$ are election- and district- random effects.[2]

From (3) - (5), the model can be written as:

$$Y_{i,t} = X_{i,t}Z_t\delta + X_{i,t}\eta_t + \lambda_i + \frac{\varepsilon_{i,t}}{\sqrt{w_{i,t}}} \qquad (6),$$

with error terms $\varepsilon_{i,t}$ and random effects η_t and λ_i assumed mutually independent.

Given the small sample sizes typically available for countries with mandatory voting (IDEA, 2007), Bayesian inference provides an attractive alternative to frequentist estimation in this setting. In particular, unlike with Full or Restricted Maximum Likelihood estimation, inference about the fixed effects

[2] $\otimes$ denotes the left Kronecker product.

within the Bayesian framework does not depend on the accuracy of the point estimates of the variance-covariance parameters. Instead, it is based on the posterior distributions given only the data, averaging over the uncertainty for all the parameters in the model (Goldstein, 1995; Bryk and Raudenbush, 2002). Taking into account the uncertainty in the estimation of the random parameters is especially important in small datasets, where the variance components are usually very imprecisely estimated (Bryk and Raudenbush, 2002).[3] In addition, as shown by Browne and Draper (2001), Maximum Likelihood methods generally exhibit convergence problems in two-level random-coefficients regression models with few higher-level units.

The joint posterior density of the model's parameters is intractable analytically, but inference can be performed by Markov chain Monte Carlo (MCMC) simulations. Adopting conjugate priors for the fixed effects and precision matrices and assuming Normal level-1 residuals - i.e., if all the $w_{i,t},\ i=1,\dots,n,\ t=1,\dots,T$, have degenerate distributions at 1 - leads to closed form conditional distributions for each of the parameters, and thus it is straightforward to approximate the corresponding marginals using Gibbs sampling.[4] However, the assumption of Normal level-1 residuals makes inferences vulnerable to the presence of outliers (Gelman, et al., 2004). In contrast, adopting a bivariate Student-t prior for $Y_{i,t}$ allows attenuating the influence of extreme observations and helps assessing the sensitivity of inferences to prior distributional assumptions (Carlin and Louis, 1996).

[3] In the context of frequentist estimation, this uncertainty can be taken into account through bootstrapping (Goldstein, 1995) or post-estimation simulation (King, Tomz and Wittenberg, 2000). However, the fact that the Bayesian approach allows directly accounting for the uncertainty in the estimates of the variance coefficients without the need for further steps makes it particularly appropriate for this kind of analysis.

[4] The full conditional posterior distributions are derived in the Appendix.

A bivariate Student t prior for $Y_{i,t}$ can be obtained from the normal/independent distribution by assuming $w_{i,t} \mid \upsilon \sim Gamma(\upsilon/2, \upsilon/2)$, $w_{i,t} > 0$, $\upsilon > 0$.[5] While it might be argued that working directly with a bivariate Student t density for $\left[\varepsilon_{i,t}^{I}, \varepsilon_{i,t}^{A}\right]'$ would be preferable to adding nT parameters to the model, the conditioning feature of the Gibbs sampler makes the augmentation of the parameter space quite natural (Carlin and Louis, 1996). In addition, this specification allows obtaining estimates of the weight parameters $w_{i,t}$, which can be useful to identify possible outliers (Rosa et al., 2003). The weight assigned to each observation in calculating posterior distributions of fixed-effects and level-1 regression parameters will depend on the posterior probabilities of the possible values of υ. Specifically, the posterior mean of $w_{i,t}$, $i = 1,...,n$, $t = 1,...,T$ is given by:

$$E\left(w_{i,t} \mid \cdots\right) = \frac{\upsilon + 2}{\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)' \Sigma^{-1} \left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right) + \upsilon} \qquad (7),$$

where $\mid \cdots$ stands for conditioning on the data and on the remaining parameters.[6] Hence, for a large enough υ, the posterior mean of $w_{i,t}$ approaches 1, leading to roughly normal tails are obtained for the level-1 errors. However, for low values of υ, the posterior mean of $w_{i,t}$ decreases as $\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)' \Sigma^{-1} \left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)$ increases. Although the conditional posterior of υ does not have a closed form, it can be approximated by discretizing the density along a grid of values and then sampling from the resulting discrete distributions.[7] When the points in the grid are

[5] I use the parametrization of the gamma distribution found in Rosa et al. (2003).
[6] See equation (25') in the Appendix.
[7] See equation (26') in the Appendix.

spaced closely together, the discrete distribution of υ provides an accurate approximation to the full conditional distribution (Seltzer et al., 2002).[8]

The two variants of the model (with bivariate normal or bivariate Student-t level-1 errors) can be compared using standard Bayesian criteria for model selection such as the Deviance Information Criterion (DIC) or Bayes factors (Spiegelhalter et al. 2002; Gelman et al., 2004). The means and standard deviation of the convergent samples obtained under each of these variants can be used to summarize the posterior distributions of the parameters. These marginal posteriors, however, are of no direct interest for the analysis. Rather, my focus lies on the effect of the explanatory variables on the proportion of invalid voting and electoral absenteeism. I thus compute the impact of each of the district- and country-level regressors on $P_{i,t}^{I}$ and $P_{i,t}^{A}$ using average predictive comparisons (Gelman and Hill, 2007).

Some aspects of the model deserve further comment. First, while in the presentation above it has been assumed that $T_i = T \;\; \forall i, i = 1,...n$ in order to simplify the notation, the model can accommodate unbalanced data sets, with different number of elections per district. In fact, the ability and flexibility to deal with nested unbalanced data sets is another advantage of Bayesian multilevel models versus frequentist approaches (Bryk and Raudenbush, 2002). Second, a more complex specification for the components of Σ could be adopted - e.g., allowing for serial correlation of the level-1 errors. Nonetheless, given the relatively small number of observations available in the application of Section 4 (with very few elections per state in some cases) and the inclusion of district random-effects, an i.i.d. assumption for the components of Σ seems appropriate (Carlin and Louis, 1996; Bryk and Raudenbush, 2002). Finally, although I focus on two particular

[8] Alternatively, a strategy based on Metropolis-Hastings algorithm sampling can be incorporated into the Gibbs sampling scheme to obtain draws from υ (Seltzer et al., 2002).

variants of the bivariate mixed model – i.e., with Normal and Student-t level-1 errors – assuming alternative densities for $w_{i,t}$ would allow obtaining several other thick-tailed distributions – e.g., slash and contaminated Normals, as in Rosa et al. (2003) - that might be appropriate to account for the presence of outliers in other applications.

4. Analyzing invalid voting and electoral absenteeism in Brazil's lower house elections

4.1. Data and methodology

Brazil provides an interesting case to analyze the determinants of abstention in countries with mandatory voting. While invalid ballots in advanced democracies under compulsory voting such as Australia and the Netherlands have historically averaged about 2 to 3 percent, the equivalent rates in Brazil have been substantially higher and more volatile over time, reaching almost 42 percent of the votes cast in the 1994 lower house election (Power and Roberts, 1995; IUPERJ, 2006). In addition, despite the fact that voting has been compulsory in the country for over 60 years, electoral absenteeism has averaged 19 percent in elections held over this period, varying from 5 to 34.5 percent (IUPERJ, 2006). Changes in the institutional design and the freeness and fairness of the elections in Brazil's recent history, together with the sharp differences in socio-demographic characteristics among its states, allow examining the impact of variables of interest on the different forms of non-voting.[9] In order to illustrate the use of the model presented in Section 3 and to compare the results with those obtained using alternative empirical strategies, I

[9] A description of the institutional, socioeconomic and political context of Brazilian elections exceeds the purposes of this paper. An overview can be found in Power and Roberts (1995).

analyze all lower house elections held in the country between 1945 and 2006. The dataset has an unbalanced structure, with 388 observations for 27 states across 16 elections.[10]

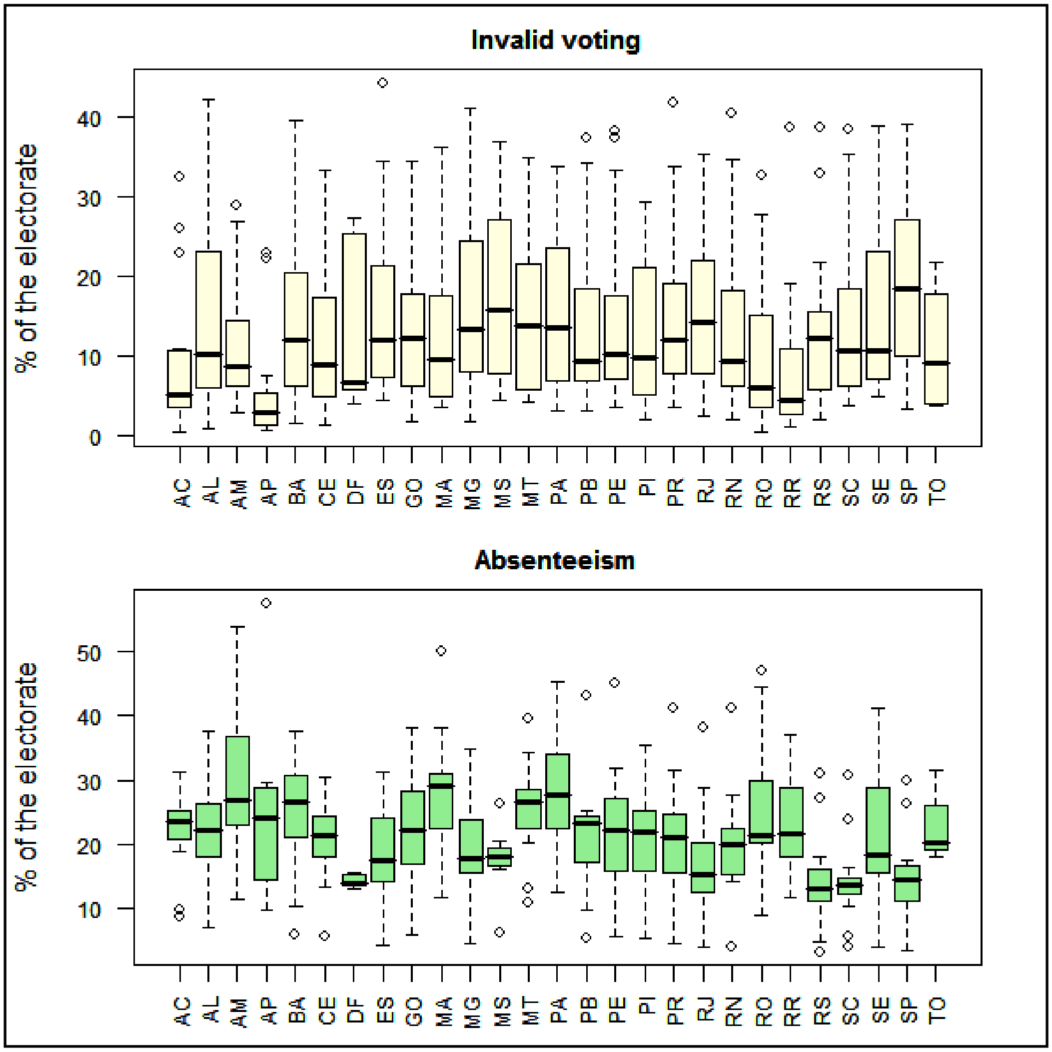

Figure 2: Invalid voting and absenteeism by in Brazilian lower house elections between 1945 and 2006, as a proportion of the electorate in each state.

[10] The number of states in Brazil increased from twenty-two to twenty-seven during this period.

The dependent variables for the analysis are the proportion of invalid votes and electoral absenteeism in lower house elections. The former is computed as the ratio of blank and null votes cast over the population eligible to vote. The latter, in turn, is calculated as the percentage of potential voters failing to comply with their duty. Figure 2 summarizes the evolution of the two variables between 1945 and 2006, discriminated by state. As can be seen, there is considerable variation in the two sources of abstention both between states and within states across elections, as well as possible geographical and temporal correlation between them.[11]

Based on the different hypothesis reviewed in Section 2, socioeconomic, institutional and protest variables are included as explanatory variables in the statistical model. The socioeconomic variables used are: *Illiteracy*, the percentage of the state's voting-age population classified as illiterate; *Urbanization*, the percentage of the state's population living in urban areas; and *FEAP*, the percentage of females in the Economically Active Population, used as a measure of women's status and the state's level of modernization. The institutional variables are: the number of *Candidates per seat*; *Franchising*, a dichotomous variable coded 1 for elections after 1985, when suffrage was extended to the illiterates, and 0 otherwise; *Electorate*, measured as the percentage of the state's total population eligible to vote; and *Ballot*, a dummy variable coded one for elections following the introduction of the single official ballot in 1962, that requires voters to write their candidate's name or registration number on a blank ballot and replaced the previous system of pre-printed ballots.[12] Finally, among the protest variables, *Manipulation* measures the degree of electoral manipulation and "political engineering", coded by Power and Roberts (1995) on a four point-scale ranging from 0 for free elections

[11] The proportions are calculated based on the number of elections held in each state.

[12] Prior to the introduction of the single official ballot ("*cedula unica*"), candidates distributed their own pre-printed ballots, which voters simply had to introduce in the ballot box. While this required considerably less information on the part of voters, it tended to favor wealthier candidates to the detriment of less affluent ones (Power and Roberts, 1995).

held under democratic rule to 3 for elections conducted under authoritarian tutelage; *Growth* is a two-year moving average of the percentage change in the national GDP; and *Inflation* is the natural logarithm of the country's average inflation rate in the two years preceding the election. Table 1 provides summary statistics for the state- and country-level predictors for the period 1945-2006.

Variable	**Mean**	**Std. Dev.**	**Min**	**percentile**	**percentile**	**Max**
State-level predictors						
Illiteracy (%)	40.0	20.8	4.7	24.6	58.9	79.8
Urbanization (%)	55.4	21.0	2.6	38.1	72.1	96.6
Females in the EAP (FEAP) (%)	23.3	16.4	3.0	10.3	40.8	58.1
Candidates per seat	4.4	2.8	1.0	2.3	6.0	15.4
Electorate (%)	40.4	19.2	6.9	24.3	59.0	74.4
Country-level predictors						
Franchising	0.4	0.5	0	0	1	1
Ballot	0.8	0.4	0	0.8	1	1
Electoral Manipulation	0.9	1.1	0	0	1.3	3
Growth (%)	5.3	3.6	-1.7	3.7	7.6	11.1
Inflation	3.7	1.7	1.7	2.7	4.2	7.5
Number of States	27					
Number of Elections	16					
Observations	388					

Table 1: Summary statistics for state- and country-level predictors.

The characterization and measurement of the independent variables closely follows Power and Roberts (1995). The only difference with them lies in the definition of *Illiteracy*. While the authors use the percentage of the state's electorate classified as illiterate (zero until 1985, when illiterates were enfranchised), I use the percentage of illiterates in the state's voting-age population. Although illiterates were not allowed to vote in Brazil until the 1986 election, the fact that more than sixty percent of the population had not finished the fourth grade by 1986, coupled with the difficulty of obtaining alternative reliable indicators covering the period under study led me to use illiteracy as a measure of the electorate's political skills (Power and Garand, 2007). In order to account for the effect of the enfranchisement of illiterates, I include the country-level variable *Franchising* in the model, allowing it to mediate the effect of *Illiteracy* on abstention across elections.

In addition, in line with Power and Roberts' (1995) argument that the country-level predictors *Ballot*, *Manipulation*, *Growth* and *Inflation* affect the proportion of invalid voting and absenteeism in each state-year, I specify the election random-intercepts as functions of these variables. Given the small number of observations in the sample (Table 1), the coefficients of the remaining district-level variables are treated as fixed effects (i.e., their variation across elections is constrained to be 0), although the model can accommodate more general specifications.

The following equations define the hierarchical model for district $i,\ i=1,...,n$ at election $t,\ t=1,...,T$:

$$\begin{aligned} Y_{i,t}^{s} = \beta_{0,t}^{s} + \beta_{1,t}^{s} Illiteracy_{i,t} + \beta_{2,t}^{s} Urbanization_{i,t} + \beta_{3,t}^{s} FEAP_{i,t} + \\ \beta_{4,t}^{s} Candidates\ per\ Seat_{i,t} + \beta_{5,t}^{s} Electorate_{i,t} + \lambda_{i}^{s} + \frac{\varepsilon_{i,t}^{s}}{\sqrt{w_{i,t}}} \end{aligned} \qquad (8),$$

$$\beta_{0,t}^{s} = \delta_{0,0}^{s} + \delta_{0,1}^{s} Ballot_{t} + \delta_{0,2}^{s} Manipulation_{t} + \delta_{0,3}^{s} Growth_{t} + \delta_{0,4}^{s} Inflation_{t} + \eta_{0,t}^{s} \qquad (9),$$

$$\beta_{1,t}^{s} = \delta_{1,0}^{s} + \delta_{1,1}^{s} Franchising_{t} + \eta_{1,t}^{s} \qquad (10),$$

$$\beta_{k,t}^{s} = \delta_{k,0}^{s}, \quad k = 2,\ldots,5 \qquad (11),$$

with $s = I, A,\ \varepsilon_{i,t} \sim N(0,\Sigma),\ \eta_t \sim N(0,\Omega_\eta), \lambda_i \sim N(0,\Omega_\lambda),$ and

$$p(w_{i,t} \mid \upsilon) = \begin{cases} 1\ \forall i,t\ \left(\text{bivariate normal prior for } Y_{i,t}\right) \text{ or} \\ Gamma\left(\dfrac{\upsilon}{2},\dfrac{\upsilon}{2}\right)\ \forall i,t\ \left(\text{bivariate Student t prior for } Y_{i,t}\right) \end{cases} \qquad (12).$$

The model was fit using WinBUGS, as called from R.[13] All the hyperparameters were assigned diffuse priors in order to let the data dominate the form of the posterior densities: the fixed effects were assigned a $N(\mathbf{0},\mathbf{100}I)$ prior, while Wishart priors with identity scale matrix and degrees of freedom equal to $rank(I)+1$ were used for the precision matrices. In order to ensure that inferences are data dependent, several alternative values for the hyperparameters were tried, yielding similar substantive results. Three parallel chains with dispersed initial values were run for 25,000 cycles, with a burn-in of 5,000 iterations. Convergence was assessed based on Gelman and Rubin's (1992) estimated potential scale reduction factor.

[13] The code is available from the author on request.

4.2. Results of the compositional-hierarchical model

Table 2 below reports the posterior means and 90% credible intervals for the fixed effects under the two variants of the model presented in Section 3: assuming bivariate Normal (Model 1-a) and bivariate Student-t (Model 1-b) level-1 priors.[14] The table shows considerable disparity in the posterior means and credible intervals of the fixed effects under both models, particularly regarding the effect of state-level predictors on the log-ratios Y^I and Y^A.[15] Comparisons between the two models based on both the DIC and Bayes Factor favor Model 1-b, indicating that the model with Student-t level-1 errors fits the data better.

Parameters	Model 1-a Gaussian level-1 errors		Model 1-b Student-t level-1 errors	
	Y^I	Y^A	Y^I	Y^A
Illiteracy	-0.03 (-0.94, 0.90)	0.76 (0.19, 1.35)	0.24 (-0.52, 1.01)	0.74 (0.18, 1.24)
Urbanization	-0.88 (-1.65, -0.13)	-0.14 (-0.58, 0.34)	-0.15 (-0.79, 0.49)	-0.17 (-0.62, 0.27)
FEAP	2.30 (0.64, 3.98)	-0.18 (-1.27, 0.87)	1.00 (-0.24, 2.24)	0.48 (-0.42, 1.40)
Candidates per seat	0.03	0.01	0.02	0.01

[14] In addition, I also estimated the model under the assumption of multivariate Student-t priors for the random coefficients. The main results, however, are virtually unchanged when assuming heavy tails at the higher-level of the model. Thus, I retain the assumption of multivariate normality at level-2 and focus on the effect of adopting alternative priors for the data model.

[15] It is worth noting that, when treating υ as unknown, the uncertainty regarding υ is propagated into the posterior distribution of the fixed-effects parameters (Seltzer et al., 2002).

	(0.01, 0.06)	(-0.01, 0.02)	(-0.01, 0.04)	(-0.01, 0.02)
Electorate	1.50 (0.62, 2.55)	0.74 (0.17, 1.30)	1.30 (0.42, 2.17)	0.47 (-0.15, 1.10)
Franchising	1.50 (0.70, 2.30)	0.66 (0.07, 1.27)	1.50 (0.75, 2.32)	0.49 (-0.04, 1.02)
Ballot	-0.04 (-0.82, 0.68)	-0.19 (-0.97, 0.60)	0.22 (-0.45, 0.93)	-0.27 (-0.98, 0.50)
Manipulation	0.52 (0.24, 0.85)	0.33 (0.03, 0.64)	0.41 (0.15, 0.67)	0.36 (0.07, 0.66)
Growth	4.30 (-2.40, 11.20)	-5.40 (-14.1, 2.90)	5.00 (-1.70, 11.60)	-5.50 (-13.70, 2.90)
Inflation	0.38 (0.23, 0.52)	-0.01 (-0.17, 0.17)	0.34 (0.20, 0.48)	-0.01 (-0.16, 0.18)
Intercept	-5.40 (-6.50, -4.30)	-1.90 (-3.02, -0.86)	-5.30 (-6.40, -4.20)	-2.0 (-3.10, -0.90)
N (first level)	388		388	
DIC	557.90		242.30	
Bayes factor	-		9.79 ×	

Table 2: Estimated posterior means and 90% confidence intervals for fixed effects under alternative distributional assumptions for the error terms.

The evidence presented in Figures 3 and 4 further supports Model 1-b. Figure 3 plots the mean posterior values of the standardized univariate and bivariate level-1 residuals from Model 1-a for the 388 observations in the dataset.

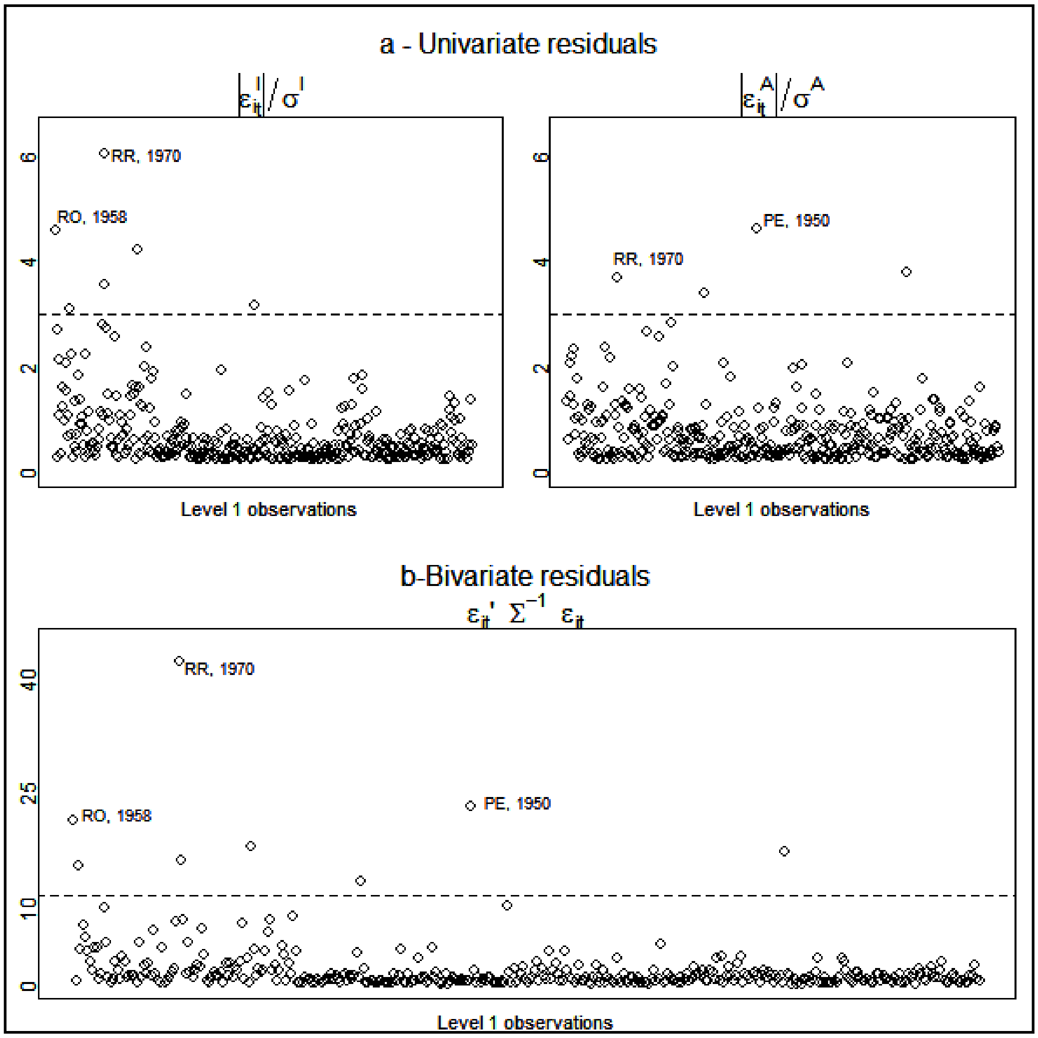

Figure 3: Posterior means of the level-1 residuals from Model 1-a. The circles represent the standardized univariate and bivariate level-1 residuals, computed based on the statistics proposed by Weiss (1994): $\frac{1}{J}\sum_{j=1}^{J}\frac{\left|Y_{i,t}^{s}-X_{i,t}\beta_{t}^{s(j)}-\lambda_{i}^{s(j)}\right|}{\sigma^{s(j)}}$, $s=I,A$ and $\frac{1}{J}\sum_{j=1}^{J}\left(Y_{i,t}-X_{i,t}\beta_{t}^{(j)}-\lambda_{i}^{(j)}\right)^{'}\Sigma^{-1(j)}\left(Y_{i,t}-X_{i,t}\beta_{t}^{(j)}-\lambda_{i}^{(j)}\right)$. The dashed horizontal lines in the upper panel correspond to the threshold of 3, while in the lower panel they are drawn at the cutpoint $k=\chi^{2}_{2(1-\alpha)}$, $\alpha=2\times\Phi(-3)$.

A few data points have standardized univariate residuals with absolute values larger than 5, and more than 2% of the observations are clear bivariate outliers, suggesting that a thick-tailed distribution might be better suited to the data.

In the same direction, the mean posterior estimate of υ under Model 1-b is 3.3, with its marginal posterior density concentrated around small values (Figure 4-a), indicating very strong departure from Normality and pointing to a heavy-tailed error distribution. As noted in Section 3, small values of υ determine that observations are weighted by an inverse function of the Mahalanobis distance $\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)' \Sigma^{-1}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)$, adjusted by the degrees of freedom. Hence, for those observations identified as (bivariate) outliers in the model with Normal level-1 errors, the posterior probability that $w_{i,t}$ is equal or greater than 1 is negligible, as illustrated in Figure 4-b. Overall, the posterior probability that $P\left(w_{i,t}\right) \geq 1$ is less than 1% for roughly 6% of the observations in the sample, providing strong evidence of outliers (Rosa et al, 2003). In addition, given that the "weight parameters" also reduce the influence of extreme observations on the posterior distribution of the election- and state- random coefficients, the number of level-2 bivariate outliers in Model 1-b is also halved with respect to Model 1-a, as shown in Figure 5. Hence, since the different comparison criteria examined above favor the model with Student-t errors, I focus on the results from Model 1-b in the remainder of the paper.

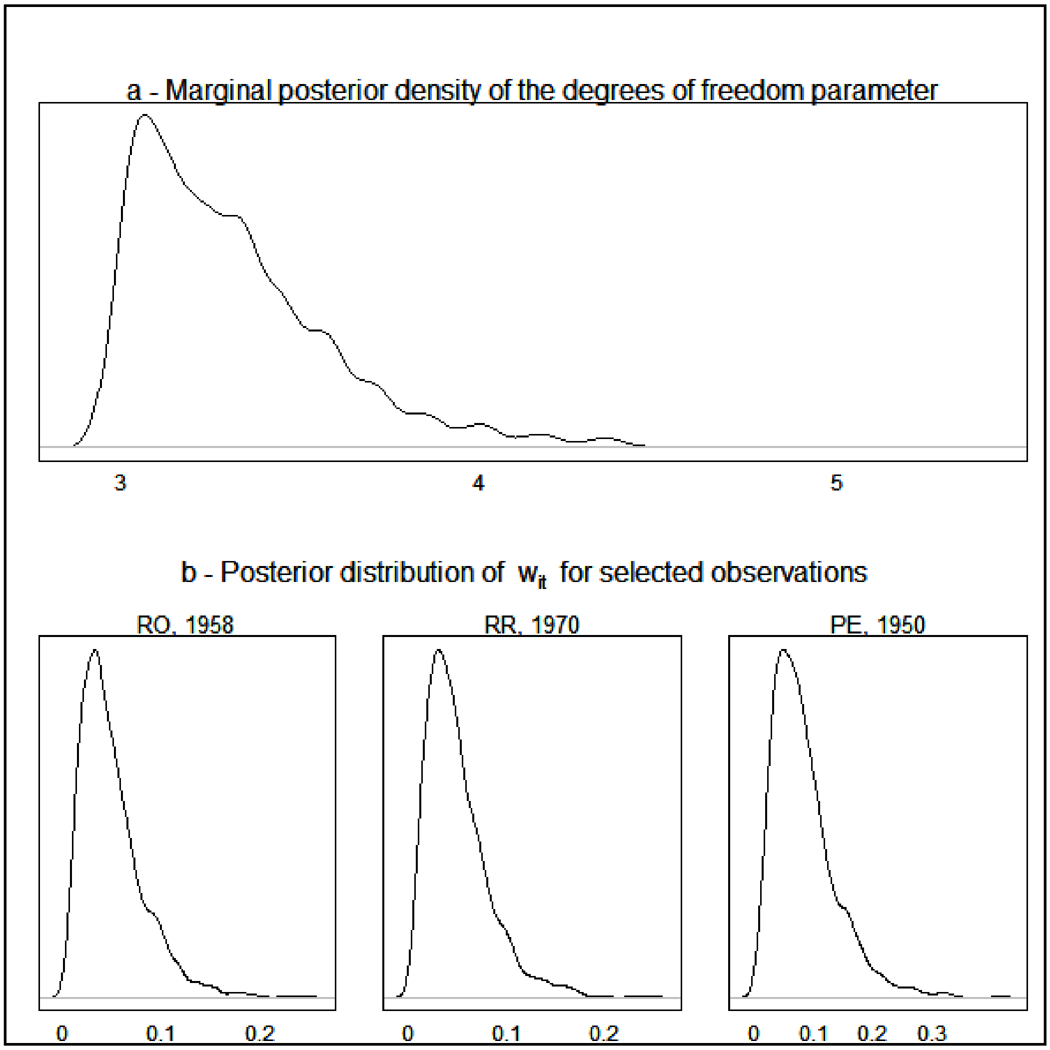

Figure 4: Posterior densities of υ and $w_{i,t}$ under Model 1-b. The upper panel of plots the posterior distribution of the degrees of freedom parameter of the Student-t distribution assumed for the error terms under Model 1-b. The lower panel plots the posterior distribution of the weight parameters for the states of Rondonia (RO), Roraima (RR) and Pernambuco (PE) in the 1958, 1970 and 1950 lower house elections.

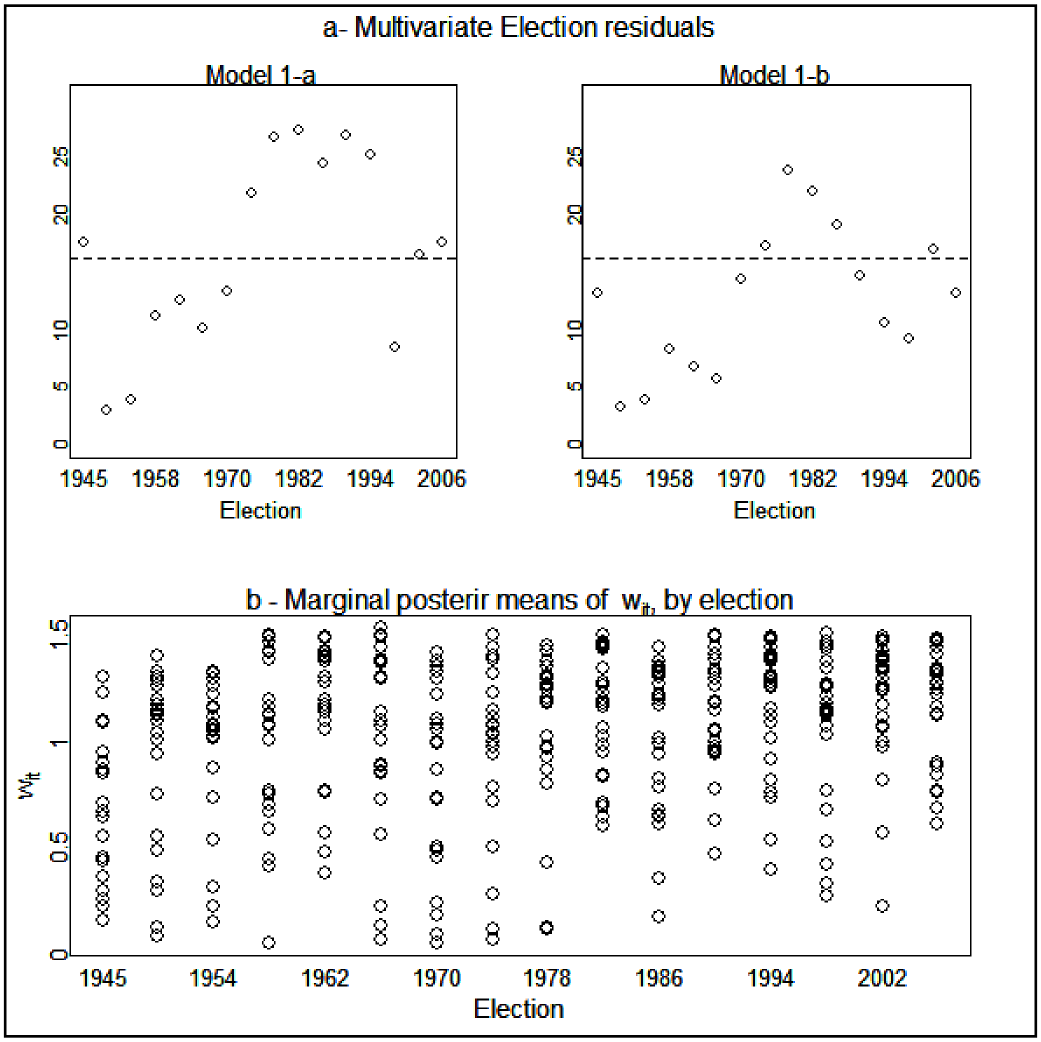

Figure 5: Posterior means of the standardized election residuals and marginal posterior means of the weight parameters. The circles in Figure 5-a represent the posterior means of the standardized election residuals, computed as $\frac{1}{J}\sum_{j=1}^{J}\left(\beta_t^{(j)}-Z_t\delta^{(j)}\right)'\Omega_{\eta}^{-1(j)}\left(\beta_t^{(j)}-Z_t\delta^{(j)}\right)$(Weiss, 1994). The dashed horizontal lines are drawn at $k=\chi^2_{4(1-\alpha)}$, $\alpha=2\times\Phi(-3)$. The circles in Figure 5-b correspond to the marginal posterior means of the weight parameters for each state, by election-year.

Table 3 reports the posterior distribution of the covariance components from the chosen model.

Level 1-errors

	Invalid Voting	**Absenteeism**
Invalid Voting	0.17 (0.03, 0.54)	
Absenteeism	0.03 (0.01, 0.09)	0.08 (0.01, 0.26)

Level 2: State random effects

	Invalid Voting	**Absenteeism**
Invalid Voting	0.16 (0.09, 0.26)	
Absenteeism	0.01 (-0.04, 0.05)	0.11 (0.07, 0.16)

Level 2: Election random effects

	Invalid voting Intercept	**Invalid voting** Illiteracy	**Absenteeism** Intercept	**Absenteeism** Illiteracy
Invalid voting Intercept	0.27 (0.12, 0.51)			
Invalid voting Illiteracy	-0.06 (-0.32, 0.14)	0.62 (0.24, 1.27)		
Absenteeism Intercept	0.05 (-0.14, 0.27)	-0.07 (-0.39, 0.23)	0.50 (0.22, 0.96)	
Absenteeism Illiteracy	-0.01 (-0.18, 0.14)	0.17 (-0.06, 0.52)	-0.12 (-0.40, 0.09)	0.29 (0.12, 0.56)

Table 3: Posterior means of variance-covariance components under Model 1-b.

Numbers in parenthesis are the 90% credible intervals.

The mean posterior correlation between the level-1 errors is moderately positive (0.24) and statistically significant at the 0.01 level. States that experience higher relative proportions of invalid voting in an election than predicted by the model also exhibit higher relative proportions of electoral absenteeism, contradicting the assumption of no correlation underlying separate univariate analyses of invalid voting and absenteeism. In addition, the bottom panel of Table 3 reveals that there is considerable variation in the election effects beyond that explained by the national-level variables included in the model. While the average correlations in Y^I and Y^A within states across elections are 0.28 and 0.24, respectively, the corresponding intra-election correlations between states are as large as 0.57 and 0.75, indicating that election-specific circumstances have a substantial influence on both forms of abstention.

Based on the convergent Gibbs samples of Model 1-b's parameters, I estimate the average effect of a unit change in each of the state- and national-level predictors on the proportion of invalid ballots and electoral absenteeism.[16] The results, reported in Table 4, reveal some interesting discrepancies regarding the determinants of the two sources of abstention. While only *Illiteracy* had a positive and significant effect on electoral absenteeism at the usual confidence levels, invalid voting in Brazil's lower house elections was strongly and positively related both to the average education and political skills levels among the electorate and to political protest. The proportion of blank and spoiled ballots rose by 0.09 percentage points for each percentage-point increase in the share of illiterates in the voting-age population. The extension of suffrage to illiterates in 1985 drove boosted invalid voting further by more than 6 points. In the same direction, each percent increase in the fraction of the states' population eligible to vote was associated to a 0.13 percentage-point rise in blank and null votes.

[16] In the case of the two binary variables, *Ballot* and *Franchising*, the effect is measured as a change from 0 to 1.

Predictor	Effect on Invalid voting	Effect on electoral absenteeism
Illiteracy	0.09* (0.05)	0.13** (0.05)
Urbanization	-0.02 (0.04)	-0.03 (0.04)
FEAP	0.10 (0.08)	0.06 (0.09)
Candidates per seat	0.11 (0.17)	-0.26 (0.17)
Electorate	0.13** (0.06)	0.07 (0.06)
Franchising	6.15*** (2.57)	1.03 (2.35)
Official Ballot	2.73 (4.17)	-5.85 (8.09)
Electoral manipulation	3.37* (2.21)	4.52 (3.44)
Growth	0.67 (0.45)	-1.00 (0.82)
Inflation	0.04*** (0.01)	-0.01 (0.02)

Table 4: Effect of a one-unit change in the predictors on invalid voting and absenteeism under Model 1-b (in percentage points). Standard errors are reported in parenthesis. Significance levels: *** 0.01, ** 0.05, *0.1.

Among the protest variables, higher levels of authoritarian political engineering resulted in an average increase of 3.4 percentage points in invalid voting. Although electoral manipulation also boosted illegal abstention, the impact of this predictor on absenteeism was much more variable across states and elections. The positive and strongly significant effect of *Inflation* on invalid voting suggests that blank and null ballots might reflect not only popular dissatisfaction with inadequate representative institutions, but also discontent with poor macroeconomic performance and economic mismanagement by the political elites. While these results support the "protest hypothesis" of invalid voting, they also suggest that less educated and newly enfranchised voters in Brazil face considerable barriers to voting (Power and Roberts, 1995). The evidence is far less conclusive in the case of electoral absenteeism, underscoring the need to examine additional confounding factors – not considered so far in the literature - that might affect noncompliance with compulsory voting laws.

Remarkably, while all the socio-economic variables tend to affect both sources of abstention in the same direction, many of the institutional and protest variables exhibit opposite average effects on the two forms of non-voting. In particular, two relevant institutional features of the open-list PR system used in Brazil's lower house election, namely, a large number of candidates and the use of a single official ballot, have a positive impact on invalid voting but a negative influence on illegal abstention. The opposite effect of *Ballot* and *Candidates per seat* on the two forms of non-voting suggests that there might be a certain trade-off between attracting voters to the polls and facilitating effective electoral participation. Factors that give voters more opportunities to affect electoral results ex-ante, such as the availability of more electoral options and a ballot design that gives voters more freedom to choose their preferred candidate, tend to increase turnout. However, at the moment of casting a ballot, the proliferation of candidates and the requirement that voters recall their preferred candidate's name or registration number tend to increase invalid voting, probably because they impose substantial informational demands and heavy decision-making costs on the electorate, especially in the context of high illiteracy rates and massive expansion of the franchise experienced in Brazil throughout the century.

4.3. Comparison with alternative modeling approaches

Figure 6 below compares the estimated causal effects of selected predictors under Model 1-b with those obtained under alternative empirical approaches commonly used to analyze abstention in compulsory voting systems. Model 2 uses separate ordinary least squares regressions for invalid voting and absenteeism, assuming independence among observations and simply pooling state- and country-level predictors. Model 3 uses separate hierarchical linear models for invalid voting and absenteeism, accounting for the temporal and geographical clustering of the data but ignoring non-negativity and unit-sum constraints. Finally, Model 4 is a compositional model with random intercepts for each state but no election-random effects, again assuming a deterministic relationship between national- and state-level predictors. [17]

The results summarized in the figure show some noticeable differences between the four models. First, standard errors under Model 1-b tend to be considerably smaller than for Model 3 and much larger than for Models 2 and 4, especially for the country-level predictors. This leads to different conclusions about the statistical significance of the impact of the national-level variables on invalid voting and electoral absenteeism under the different models. For instance, setting the stochastic terms in η_t to zero in Models 2 and 4 leads to significant effects of economic growth and electoral manipulation on absenteeism at the 0.01 level. Neither of these variables have a systematic effect on abstention under Models 1

[17] In the case of Models 3 and 4, a bivariate Student-t distribution was used for the data model, while multivariate Normal distributions were specified for the random coefficients. The substantive results reported below remain unchanged if Gaussian distributions are assumed for the level-1 errors as well. All four models were fit in WinBUGS. The corresponding codes are included in the supplementary materials accompanying this paper.

and 3 even at the 0.05 level. In contrast, *Illiteracy* is significantly and positively correlated with electoral absenteeism under Model 1-b, but not for any of the other 3 models. On the other hand, the large standard errors for Model 3 determine that none of the national-level variables has a significant effect on either source of abstention at the usual confidence levels.

Even the sign of some of the estimated marginal effects differ across the four models. For instance, a higher number of *Candidates per seat* has a positive average effect on invalid voting under Model 1-b, suggesting that a larger number of contestants increases the likelihood of voter error and/or makes it more difficult for voters to choose a single preferred candidate. In contrast, the number of candidates and the proportion of blank and null ballots are negatively correlated under the other 3 models. FEAP, on the other hand, has a negative and significant impact on invalid voting under Model 4, but a positive – though not statistically significant – effect under the other models. Also, as seen in the lower panel of Figure 6, the results from Models 2 and 4 indicate that that the extension of voting rights to illiterates led to significantly lower levels of electoral absenteeism, suggesting that this group of new voters was more likely to show up at the polls even when - unlike for literate citizens - voting is optional for illiterates. The average effect of *Franchising* on electoral absenteeism has the opposite sign under Models 1-b and 3. Finally, and in contrast to the other three models, *Ballot* has a negative and marginally significant effect on invalid voting under Model 2, leading to the rather implausible conclusion that the introduction of a more complex ballot system that requires considerably more information on the part of voters resulted in lower rates of blank and spoiled ballots.

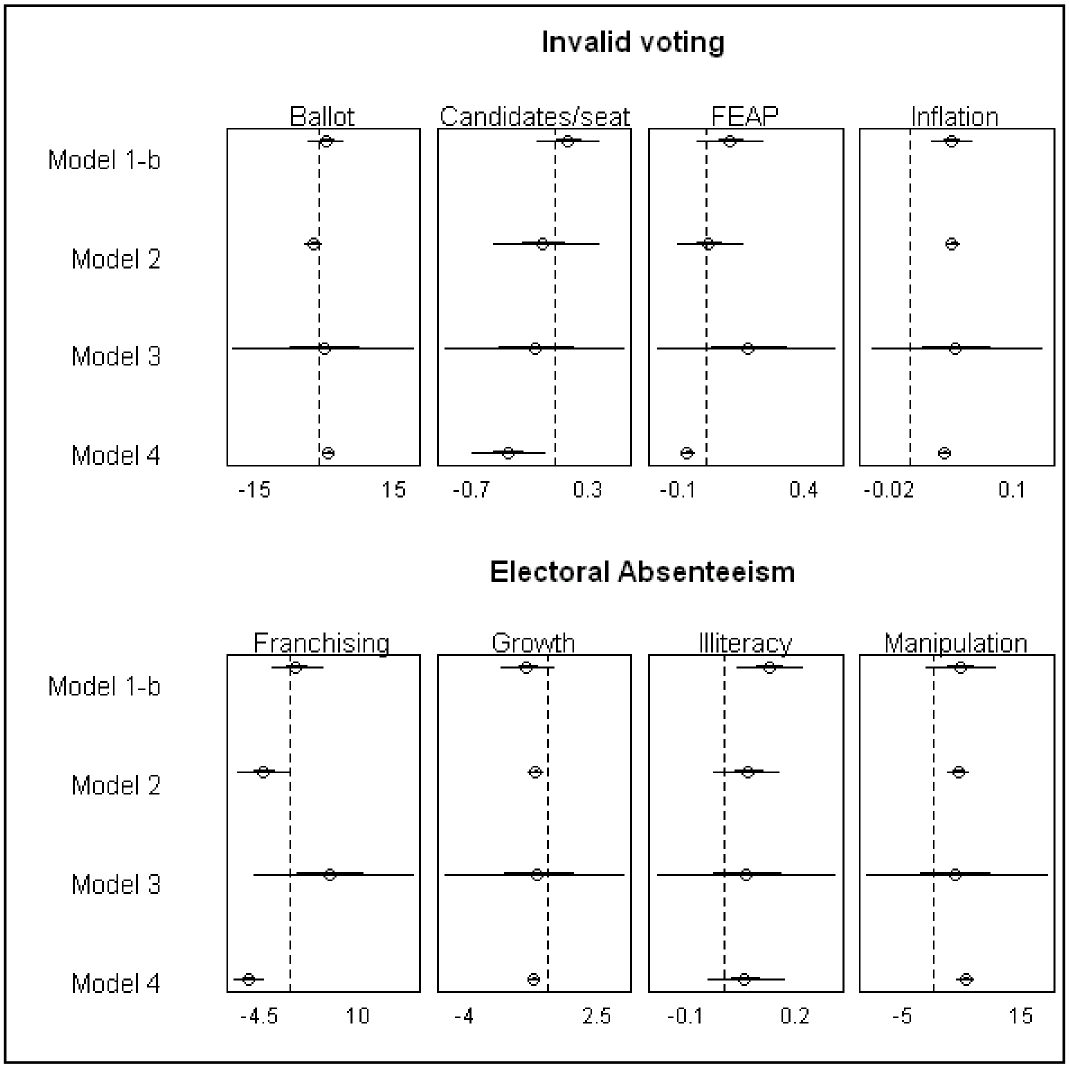

Figure 6: Estimated marginal effects of selected predictors across models, in percentage points. The graph shows the effect of a one-unit change in each of the predictors on invalid voting (upper panel) and electoral absenteeism (lower panel). The center dots correspond to the point estimates, the thicker lines to the 50% credible intervals, and the thinner lines to the 90% credible intervals.

These conflicting results lead to different substantive conclusions about the relative validity of alternative theories proposed to account for abstention under mandatory voting and might entail different implications regarding, for instance, the design of electoral systems and the institutional reforms needed to promote and consolidate political participation in compulsory voting systems (Power and Roberts, 1995). It is therefore important to determine which model fits the data better. Following Iyengar and Dey (2004), a plausible comparison criteria based on the discrepancy between observed and simulated data would favor the model that minimizes the predictive loss $d\left(P^{\mathrm{Rep}}, P^{Obs}\right) = E\left(\left\|P^{\mathrm{Rep}} - P^{Obs}\right\|^2 \middle| P^{Obs}\right)$, which can be estimated as:

$$\hat{d} = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J}\left\|P_{i,t}^{Obs} - P_{i,t}^{\mathrm{Rep}(j)}\right\|^2\right) \qquad (13),$$

where $P_{i,t}^{Rep} = \left(P_{i,t}^{Rep(1)},\ldots,P_{i,t}^{Rep(J)}\right)$ denotes the data sampled from the predictive distribution $p\left(P_{i,t}^{Rep} \mid P_{i,t}^{Obs}\right) = \int p\left(P_{i,t}^{Rep} \mid \theta\right) p\left(\theta \mid P_{i,t}^{Obs}\right) d\theta$ under each model.[18]

Table 5 reports the point estimates (means) and 90% credible intervals for the posterior predictive loss based on $J = 1{,}000$ hypothetical replications of $P_{i,t}^{I}$ and $P_{i,t}^{A}$ for the four models. The compositional-hierarchical model exhibits the lowest discrepancy between the replicated and the actual data (at the 0.01 level). In contrast, the two models that implement separate univariate analyses for each source of abstention have the highest estimated predicted losses. In particular,

[18] In the case of the compositional-hierarchical model, $P_{i,t}^{\mathrm{Rep}}$ are obtained from $Y_{i,t}^{Rep}$ using the logarithmic transformations (1) and (2).

Model 2, which in addition ignores the multilevel nature of the data, exhibits the worst fit.

Model	$\hat{d}$
1-b	2.84 (2.33, 3.51)
2	16.71 (13.72, 19.97)
3	8.33 (7.33, 9.44)
4	6.59 (5.75, 7.55)

Table 5: Estimates of the Posterior Predictive Loss for alternative modeling strategies. Numbers in parenthesis are the 90% credible intervals.

The superior performance of Model 1-b is also illustrated in Figure 7, which plots the actual proportions of invalid voting and absenteeism and the expected proportions under the four models, obtained by averaging $P_{i,t}^{\text{Rep}(j)}$, $j = 1,\ldots,1000$, over the simulations. As seen in the figure, Models 2 and 3 lead to negative expected proportions of invalid votes for 49% and 14% of the state-years in the sample, respectively. While both compositional models avoid this problem, relaxing the assumption of a deterministic relationship between national- and state-level predictors and allowing for additional variability in the election-effects results in a better fit for Model 1-b *vis-à-vis* Model 4.

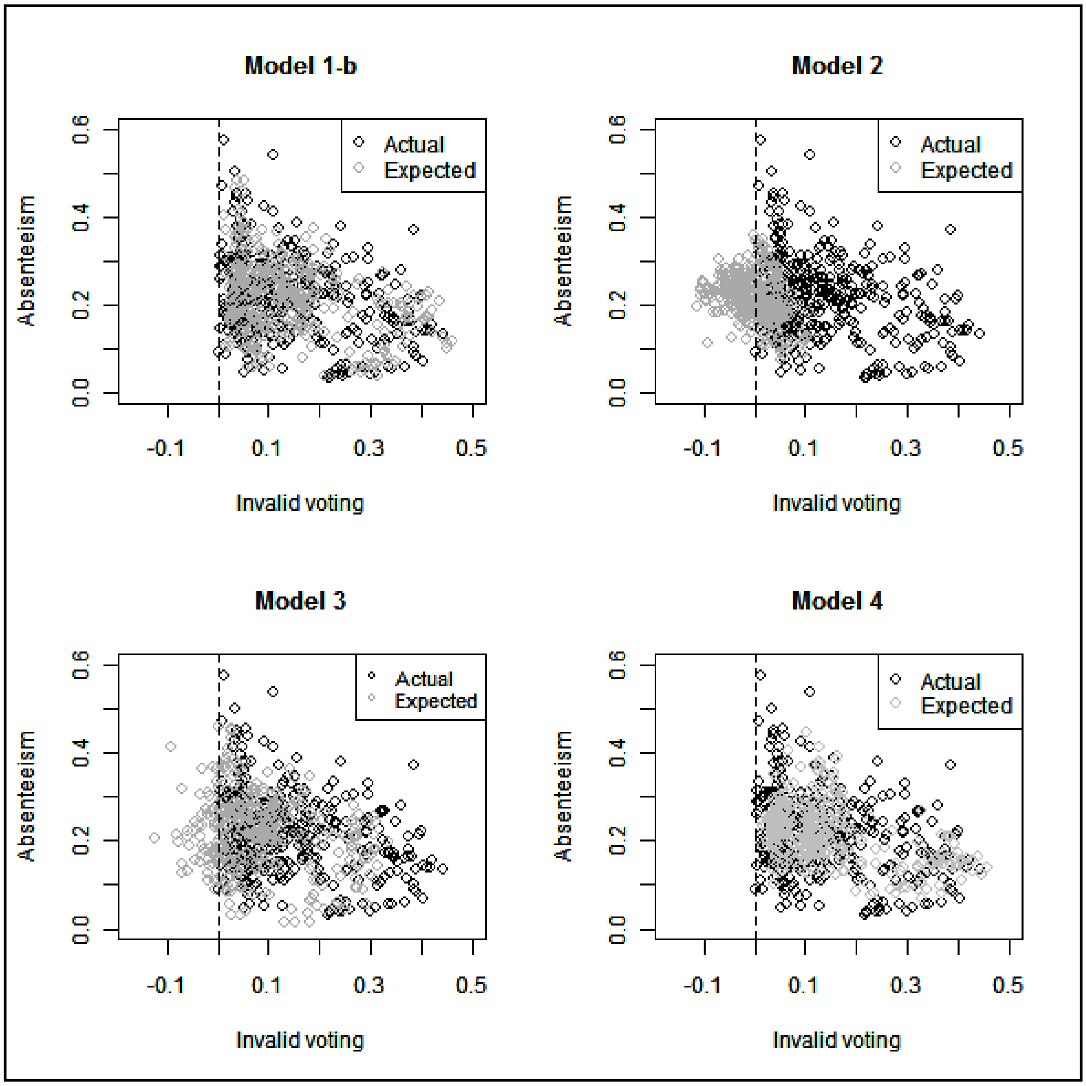

Figure 7: Actual and expected proportions of invalid voting and electoral absenteeism under alternative modeling strategies. The gray circles correspond to the expected proportion of invalid voting and electoral absenteeism for each state-election of the sample for the model under consideration. The black circles correspond to the actual values.

Hence, the evidence presented above indicates that the statistical model developed in this paper provides a much improved fit over the other three modeling approaches considered, and suggests that the methodological differences between these competing empirical strategies may have substantial consequences in terms of the analysis of the determinants of abstention under compulsory voting.

5. Concluding remarks

Different theories, drawing on the literature on voter turnout in industrialized countries, have been proposed to explain the prevalence of invalid voting and electoral absenteeism in compulsory voting systems. This paper integrates the socioeconomic, institutional, and political-protest approaches in a statistical model aimed at analyzing the determinants of both sources of abstention in district-level elections. The model presented here accounts for the compositional and hierarchical structure of district-level electoral data and addresses robustness concerns raised by the use of small sample sizes typically available for democracies with mandatory voting, most of which are recently democratized Latin American nations.

Results obtained from the application of the model to data from Brazil's legislative elections allow drawing interesting substantive and methodological conclusions. The evidence presented above reveals considerable differences in the determinants of both forms of non-voting. In line with Power and Roberts (1995), I find that the proportion of blank and null ballots in Brazil's lower house elections was strongly positively related both to political protest and to the existence of important informational barriers to voting, in particular for less educated and newly enfranchised voters. The influence of these variables on illegal abstention, however, was less evident. In addition, some of the institutional characteristics of the electoral system, such as the proliferation of candidates and the introduction of a complex ballot design, seem to affect the two sources of abstention in opposite directions. Comparisons based on posterior simulations indicate that the model presented here fits the data better than several alternative empirical strategies commonly used in previous studies on this topic. More importantly, some of the substantive conclusions and the policy implications derived from the compositional-hierarchical model differed from those drawn from less appropriate modeling approaches.

Although the model was applied to the particular case of Brazil, it provides a general tool to analyze the determinants of abstention in compulsory systems. Potential avenues for future research would be to examine the determinants of illegal abstention in greater detail – which, as noted above, does not seem to be strongly related to either protest voting or institutional variables – and to analyze the performance of the model and the robustness of the findings reported here from a comparative perspective. From a methodological standpoint, using non-parametric methods to estimate the joint density of invalid voting and absenteeism would allow examining their determinants and interactions without imposing specific parametric distributions.

Appendix - Full Conditional Posterior Distributions

Assuming conditional independence throughout, the model presented in Section 3 can be specified in a Bayesian context as:

$$Y_{i,t} \sim N\left(X_{i,t}\beta_t + \lambda_i,\ \frac{1}{w_{i,t}}\Sigma \right),\ i=1,\dots,n,\ t=1,\dots,T \qquad (14),$$

$$\beta_t \sim N\left(Z_t\delta, \Omega_\eta \right),\ t=1,\dots,T \qquad (15),$$

$$\lambda_i \sim N\left(0, \Omega_\lambda\right),\ i = 1, ..., n \qquad (16),$$

with conjugate priors for the fixed effects and the precision matrices:

$$\begin{aligned}
&\delta \sim N\left(\delta_0, \Omega_\delta\right), \\
&\Sigma^{-1} \sim Wishart(P, \rho_P), \quad \left|P\right| > 0, \rho_P \geq 2 \\
&\Omega_\eta^{-1} \sim Wishart(Q, \rho_Q), \ \left|Q\right| > 0, \rho_Q \geq 2\left(K+1\right) \\
&\Omega_\lambda^{-1} \sim Wishart(R, \rho_R), \ \left|R\right| > 0, \rho_R \geq 2
\end{aligned} \qquad (17).$$

Assuming that all the $w_{i,t},\ i = 1, ..., n,\ t = 1, ..., T,$ are mutually independent, the joint posterior density of all the unknown parameters of the model is given by:

$$\begin{aligned}
&f\left(\beta, \lambda, \delta, \Sigma, \Omega_\eta, \Omega_\lambda, w, \upsilon \middle| Y\right) \propto \\
&\quad \left[\prod_{i=1}^{n}\prod_{t=1}^{T} w_{i,t}\right] \left|\Sigma\right|^{-nT/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)'\Sigma^{-1}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)\right\} \\
&\quad \times \prod_{i=1}^{n}\prod_{t=1}^{T} p\left(w_{i,t}\middle|\upsilon\right) \times p\left(\upsilon\right) \times \left|\Omega_\eta\right|^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\beta_t - Z_t\delta\right)'\Omega_\eta^{-1}\left(\beta_t - Z_t\delta\right)\right\} \\
&\quad \times \left|\Omega_\lambda\right|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{t=i}^{n}\lambda_i{}'\Omega_\lambda^{-1}\lambda_i\right\} \times \left|\Omega_\delta\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\delta - \delta_0\right)'\Omega_\delta^{-1}\left(\delta - \delta_0\right)\right\} \\
&\quad \times \left|\Sigma^{-1}\right|^{\frac{\rho_P - 2 - 1}{2}} \exp\left\{-\frac{1}{2}tr\left(P^{-1}\Sigma^{-1}\right)\right\} \times \left|\Omega_\eta{}^{-1}\right|^{\frac{\rho_Q - 2\left(K+1\right) - 1}{2}} \exp\left\{-\frac{1}{2}tr\left(Q^{-1}\Omega_\eta{}^{-1}\right)\right\} \\
&\quad \times \left|\Omega_\lambda{}^{-1}\right|^{\frac{\rho_R - 2 - 1}{2}} \exp\left\{-\frac{1}{2}tr\left(R^{-1}\Omega_\lambda{}^{-1}\right)\right\}
\end{aligned} \qquad (18).$$

Given $w = \left(w_{i,1}, ..., w_{n,T}\right)'$ the full conditional posterior densities of $\{\beta_t\}, \{\lambda_i\}, \delta, \Sigma, \Omega_\eta$ and Ω_λ are:[19]

$$
\begin{aligned}
&\beta_t \,|\, \cdots \sim N\left(b_t, B_t\right), \quad t = 1, \ldots T, \text{ with} \\
&b_t = \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} X_{i,t} + \Omega_\eta^{-1}\right]^{-1} \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} \left(Y_{i,t} - \lambda_i\right) + \Omega_\eta^{-1} Z_t \delta\right] \qquad (19); \\
&B_t = \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} X_{i,t} + \Omega_\eta^{-1}\right]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&\lambda_i \,|\, \cdots \sim N\left(d_i, D_i\right), \quad i = 1, \ldots n, \text{ with} \\
&d_i = \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} + \Omega_\lambda^{-1}\right]^{-1} \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} \left(Y_{i,t} - X_{i,t} \beta_t\right)\right] \qquad (20); \\
&D_i = \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} + \Omega_\lambda^{-1}\right]^{-1}
\end{aligned}
$$

$$
\delta \,|\, \cdots \sim N\left(\left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} Z_t + \Omega_\delta^{-1}\right]^{-1} \left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} \beta_t + \Omega_\delta^{-1} \delta_0\right], \left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} Z_t + \Omega_\delta^{-1}\right]^{-1}\right) \qquad (21);
$$

$$
\Sigma^{-1} \,|\, \cdots \sim Wishart\left(\left[\sum_{t=1}^{T} \sum_{i=1}^{n} w_{i,t} \left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i\right)\left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i\right)' + P^{-1}\right]^{-1}, \rho_P + nT\right) \qquad (22);
$$

[19] $|\cdots$ stands for conditioning on the data and on the remaining parameters.

$$\Omega_{\eta}^{-1}\Big|\cdots \sim Wishart\left(\left[\sum_{t=1}^{T}\left(\beta_t - Z_t\delta\right)\left(\beta_t - Z_t\delta\right)' + Q^{-1}\right]^{-1}, \rho_Q + T\right) \qquad (23);$$

$$\Omega_{\lambda}^{-1}\Big|\cdots \sim Wishart\left(\left[\sum_{i=1}^{n}\lambda_i\lambda_i' + R^{-1}\right]^{-1}, \rho_R + n\right) \qquad (24).$$

To complete the specification for a Gibbs sampling scheme, the full conditional posterior distributions of w and υ are required:

$$w_{i,t}\Big|\cdot\cdot \propto w_{i,t}\exp\left\{-\frac{\omega_{i,t}}{2}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)'\ \Sigma^{-1}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)\right\}\times p\left(w_{i,t}\middle|\upsilon\right) \qquad (25),$$

$$\upsilon\Big|\cdots \propto p\left(\upsilon\right)\prod_{i=1}^{n}\prod_{t=1}^{T}p\left(w_{i,t}\middle|\upsilon\right) \qquad (26),$$

with $p\left(w_{i,t}\middle|\upsilon\right)$ depending on the particular normal/independent distribution adopted for the level-1 errors. In the case of Model 1-b, which assumes $w_{i,t}\big|\upsilon \sim Gamma\left(\upsilon/2, \upsilon/2\right)$, $w_{i,t} > 0, \upsilon > 0$, (25) and (26) become:

$$w_{i,t}\Big|\cdot\cdot \sim Gamma\left(\frac{\upsilon}{2}+1,\ \frac{1}{2}\left[\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right)'\ \Sigma^{-1}\left(Y_{i,t} - X_{i,t}\beta_t - \lambda_i\right) + \upsilon\right]\right) \qquad (25'),$$

$$\upsilon\big|\cdots \sim \left[2^{\upsilon/2}\Gamma\left(\frac{\upsilon}{2}\right)\right]^{-nT} \upsilon^{\frac{\upsilon nT}{2}} \exp\left\{-\frac{\upsilon}{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t} - \log\left(w_{i,t}\right)\right)\right\} \quad (26').$$

References

Aitchison, John. 1986. *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Antweiler, Werner. 2001. Nested random effects estimation in unbalanced panel data. *Journal of Econometrics* 101: 295 – 313.

Australian Joint Standing Committee on Electoral Matters. 2000. *The 1998 Election: Report of the Inquiry into the conduct of the 1998 Federal Election and matters related thereto.* Canberra: Parliament of the Commonwealth of Australia.

Barndorff-Nielsen, Ole E., and Jørgensen, Bent. 1991. Some Parametric Models on the Simplex. *Journal of Multivariate Analysis* 39: 106 – 116.

Blais, André, and Dobrzynska, Agnieszka. 1998. Turnout in Electoral Democracies. *European Journal of Political Research* 33: 239-261.

Browne, William J., and Draper, David. 2001. Implementation and Performance Issues in the Bayesian and Likelihood Fitting of Multilevel Models. *Computational Statistics* 15(3): 391-420.

Bryk, Anthony S., and Raudenbush, Stephen W. 2002. *Hierarchical linear models: Applications and data analysis methods*. 2nd ed. Newbury Park, CA: Sage Publications.

Carlin, Bradley P., and Louis, Thomas A. 1996. *Bayesian and Empirical Bayes Methods for Data Analysis.* London: Chapman & Hall.

Escobar, Marcelo, Calvo, Ernesto, Calcagno, Natalia, and Minvielle, Sandra. 2002.

Ultimas Imágenes Antes del Naufragio: Las Elecciones del 2001 en Argentina. *Desarrollo Economico* 42: 25-44.

Fornos, Carolina, Power, Timothy, and Garand, James. 2004. Explaining Voter Turnout in Latin America, 1980 to 2000. *Comparative Political Studies* 37: 909 – 940.

Gelman, Andrew, Carlin, John B., Stern, Hal S., and Rubin, Donald B. 2004. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall.

Gelman, Andrew, and Hill, Jennifer. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Goldstein, Harvey. 1995. *Multilevel Statistical Models*. London: Arnold.

Hill, Lisa. 2002. On the Reasonableness of Compelling Citizens to 'Vote': The Australian Case. *Political Studies* 50: 80 -101.

Hirczy de Mino, Wolfgang. 1994. The impact of mandatory voting laws on turnout: A quasi-experimental approach. *Electoral Studies* 13: 64-76.

Instituto Universitario de Pesquisas de Rio de Janeiro (IUPERJ). 2006. Banco de Dados Eleitorais Do Brasil (accessed December 2008).

International Institute for Democracy and Electoral Assistance, IDEA. 2007. Voter Turnout Database, Compulsory Voting: ://www.idea.int/vt/index. (accessed December 2009).

Iyengar, Malini, and Dey, Dipak K. 2004. Bayesian analysis of compositional data. In D. Dey, S. Ghosh and B. Mallick (eds.), *Generalized linear Models: A Bayesian Perspective*, pp. 349-364. New York: Marcel Dekker.

Jackman, Simon. 2001. Voting: Compulsory. In N. Smelser and P. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier Science.

Jackman, Robert W., and Miller, Ross A. 1995. Voter Turnout in the Industrial Democracies during the 1980s. *Comparative Political Studies* 27: 467-492.

Katz, Jonathan, and King, Gary. 1999. A Statistical Model for Multiparty Electoral

Data. *American Political Science Review* 93: 15 – 32.

King, Gary, Tomz, Michael, and Wittenberg, Jason. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *Journal of Political Science 44:* 341-355.

Kitschelt, Herbert. 1995. *The Radical Right in Western Europe. A Comparative Analysis*. Ann Arbor: The University of Michigan Press.

Lijphart, Arend. 1997. Unequal Participation: Democracy's Unresolved Dilemma. *American Political Science Review* 91: 1-14.

Maas, Cora J., and Hox, Joop J. 2004. Robustness issues in multilevel regression analysis. *Statistica Neerlandica* 58(2): 127–137.

Moisés, Alvaro. 1993: Elections, Political Parties and Political Culture in Brazil: Changes and Continuities. *Journal of Latin American Studies* 25: 575-611.

Power, Timothy, and Garand, James. 2007. Determinants of Invalid Voting in Latin America. *Electoral Studies* 26(2): 432-444.

Power, Timothy, and Roberts, Timmons. 1995. Compulsory Voting, Invalid Ballots, and Abstention in Brazil. *Political Research Quarterly* 48: 795-826.

Rayens, W. S., and Srinivasan, C. 1991. Box Cox Transformations in the Analysis of Compositional Data. *Journal of Chemometrics* 5: 227 – 239.

Rosa, Guilherme J., Padovani, Carlos R., and Gianola, Daniel. 2003. Robust Linear Mixed Models with Normal/Independent Distributions and Bayesian MCMC Implementation. *Biometrical Journal* 45(5): 573-590.

Rosenstone, Steve, and Hansen, John. 1993. *Mobilization, participation and democracy in America*. New York: Macmillan.

Seltzer, Michael, Novak, John, Choi, Kilchan, and Lim, Nelson. 2002. Sensitivity Analysis for Hierarchical Models Employing t Level-1 Assumptions. *Journal of Educational and Behavioral Statistics* 27(2), 181-222.

Spiegelhalter, David J., Best, Nicola G., Carlin, Bradley P., and van der Linde, Angelika. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, Series B, 64 (4): 583-639.

Thum, Yeow M. 1997. Hierarchical linear Models for Multivariate Outcomes. *Journal of Educational and Behavioral Statistics* 22(1): 77-108.

Verba, Sydney, Norman Nie, and Jae-On Kim. 1978. *Participation and Political Equality: A Seven Nation Comparison*. New York: Cambridge Univ. Press.

Weiss, Robert E. 1994. Residuals and Outliers in Repeated Measures Random Effects Models. Technical Report 161, Department of Biostatistics, University of California at Los Angeles.

Zellner, Arnold. 1917. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.