DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A STATISTICAL MODEL OF ABSTENTION UNDER COMPULSORY VOTING

Gabriel Katz

SOCIAL SCIENCE WORKING PAPER 1290

June 2008

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.

JEL classification numbers: C11,C30

Key words:

A statistical model of abstention under compulsory voting

Gabriel Katz

1. Introduction

The desire to provide a political system with popular legitimacy and to increase the representativeness of elected public officers have often been asserted as major arguments justifying the imposition of compulsory voting provisions (Verba, Nie and Kim, 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 voter turnout (Hirczy, 1994; Lijphart 1997; Fornos, 1996), compelling voters to go to the polls does not automatically mean that they will cast a vote 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 (Hirczy, 1994; Fornos, Power and Garand, 2004) or on the determinants of invalid voting (McAllister and Makkai, 1993; Power and Garand, 2007). The common approach of studies in this area has been to treat the proportion of invalid votes or electoral absenteeism as the dependent variable and regress each on a set of explanatory variables. This standard procedure exhibits two main shortcomings. Fist, 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 the correlation between them is assumed to be zero, and changes in the standard error estimates that might result from a bivariate model could substantially modify the conclusions drawn from separate univariate analyses (Zellner, 1971; Thum, 1997). 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 among the electorate cannot be negative and that must sum 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 might 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. Ignoring the hierarchical structure of the data and simply pooling national- and district-level variables may thus result in inefficient parameter estimates and negatively biased standard errors, potentially leading to “spuriously significant” statistical effects (Antweiler, 2001; Maas and Hox, 2004; Franzese, 2005).

Drawing on the literature on compositional data (Aitchison and Shen, 1980; 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 alternative 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 modeling 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 three 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 (McAllister and Makkai, 1993; 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 (1993; 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 the perceived efficacy, access to information and development of political skills among the electorate. 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; Powell, 1986; Rosenstone and Hansen, 1993), empirical evidence from countries with mandatory voting (Power and Roberts, 1995; Fornos et al., 2004) suggest that the impact of socioeconomic factors on electoral absenteeism in these countries is quite moderate.

Other authors have underscored the role of the institutional context and design in explaining invalid voting and absenteeism. For instance, Blais and Dobrzynska (1998) and Kostadinova (2003) concluded that a higher number of political parties depress turnout by increasing the unpredictability of electoral and policy outcomes, and the same would apply for highly disproportional systems that punish minor parties and reduce voters' perceived efficacy (Jackman, 1987; Jackman and Miller, 1995). In the same direction, McAllister and Makkai (1993) and Power and Roberts (1995) provide evidence that institutional factors such as 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 (Kitschelt, 1995; Lubbers and Scheepers, 2000). A protest vote can be defined as a vote primarily cast to express discontent with politics, rather than to affect public policies (Van der Brug and Fennema, 2003). In a system of compulsory voting, citizens' discontent with the political establishment would translate into higher null and blank ballots and illegal abstention (Derks and Deschouwer, 1998). This interpretation has often been quoted in Brazil and Latin America to explain temporary increases in invalid voting and absenteeism (Moisés, 1993; Jocelyn–Holt, 1998; Escobar, Calvo, Calcagno and Minvielle, 2002).

Although the socioeconomic, institutional and protest approaches are usually presented as competitors rather than as complementary, 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, since these approaches are grounded in the literature on political participation in developed democracies, where invalid voting has not received much academic attention (Power and Garand, 2007), 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), despite considerable variations in their relative incidence both across countries and elections (IIDEA, 2007). Furthermore, from a methodological perspective, they failed to examine the potential interactions between the determinants of these two sources of abstention, implicitly assuming that the effects of the relevant predictors on invalid voting are independent of their impacts 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 and Shen, 1980; Aitchison, 1986; Katz and King, 1999) and on Bayesian hierarchical modeling (Lindley and Smith, 1972; 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 following non-negativity and unit-sum constraints (Katz and King, 1999):

$$P_{i,t}^{s} \in [0,1], \qquad s = I, A, V \tag{1}$$

$$P_{i,t}^{I} + P_{i,t}^{A} + P_{i,t}^{V} = 1 \tag{2}$$

These constraints determine that $P_{i,t}^{I}$, $P_{i,t}^{A}$ and $P_{i,t}^{V}$ fall in the simplex space. Figure 1 illustrates the simplex sample space using a ternary plot for lower house elections in Brazil between 1945 and 2006. As seen in the Figure, while ignoring the non-negativity and unit-sum constraints in the statistical analysis might not be problematic for some of the observations in the sample – e.g., Roraima (RR) in the 1970 election – in most other cases - e.g., Amapá (AP), 1954 –, this could lead to unfeasible results, such as negative predicted proportions of invalid ballots.

[Figure 1 here]

A model aimed at analyzing the determinants of abstention in compulsory voting systems must take the constraints defined in (1) and (2) 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 these constraints, even if eventually the point predictions obtained 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 using a Bayesian implementation of a bivariate mixed model for invalid voting and electoral absenteeism.

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

$$P_{i,t}^{I} = \frac{\exp\left[Y_{i,t}^{I}\right]}{1 + \exp\left[Y_{i,t}^{I}\right] + \exp\left[Y_{i,t}^{A}\right]} \tag{3}$$

$$P_{i,t}^{A} = \frac{\exp\left[Y_{i,t}^{A}\right]}{1 + \exp\left[Y_{i,t}^{I}\right] + \exp\left[Y_{i,t}^{A}\right]} \tag{4}$$

Since the $Y_{i,t}^{s}$, $s = I, A$, are defined over the whole real line, it is possible to model $Y_{i,t} = \left[Y_{i,t}^{I}, Y_{i,t}^{A}\right]$ using a normal/independent distribution (Andrews and Mallows, 1974; Liu, 1996; Seltzer et al., 2002) that assigns weight parameters to each observation in the sample, as in a Weighted Least Squares analysis:

$$Y_{i,t} = \mu_{i,t} + \frac{\varepsilon_{i,t}}{\sqrt{w_{i,t}}} \qquad (5),$$

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 $p\left(w_{i,t} \middle| \upsilon\right)$, and υ a scalar or vector-valued parameter. The main advantage of assuming a normal/independent distribution is that, due to the unconstrained properties of Σ, the model now allows for any pattern of dependency between $P_{i,t}^{I}$ and

[1] Due to the logarithmic transformations involved, the baseline composites are assumed to be strictly positive. Although this poses no problem for this type of electoral data, alternative models based on Box-Cox transformations (Rayens and Srinivasan, 1991) have been proposed to deal with the problem of null composites.

$P_{i,t}^{A}$.[2] In addition, besides including the bivariate normal as a particular case (when $w_{i,t} = 1 \; \forall i,t$), the normal/independent distribution also provides a group of thick-tailed distributions often useful for robust inference and identification of outliers (Seltzer et al., 2002; Rosa, Padovani and Gianola, 2003), particularly when the number of districts or elections in the sample is relatively small.

The focus of the model lies in the specification of $\mu_{i,t}$. Since $\mu_{i,t}^{I}$ and $\mu_{i,t}^{A}$ are unbounded, it is possible to reparametrize them as linear functions of regressors. 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. Moreover, the impact of district-level variables on invalid voting and absenteeism might itself be mediated by these 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 assumed to be constant across all districts in a given election, 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 account for 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 would be highly unstable given the paucity of the data typically available for analyzing countries with compulsory voting, most of them recently democratized Latin American nations (Browne and Draper, 2001; Gelman and Hill, 2007;).

[2] This is, in fact, the key advantage of assuming a scale mixture of multivariate normals *vis a vis* alternative statistical models for compositional data, such as the Dirichlet distribution (Johnson and Kotz, 1972) and the S^{-} distribution (Barndorff-Nielsen and Jørgensen, 1991).

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:

$$\mu_{i,t} = X_{i,t}\beta_t + \lambda_i \tag{6}$$

$$\beta_t = Z_t\delta + \eta_t \tag{7}$$

where

$X_{i,t}$ is a $2\times 2(K+1)$ matrix, $X_{i,t} = \left[I_2 \middle| x_{i,t}\otimes I_2\right]$,

β_t is a $2(K+1)\times 1$ vector, $\beta_t = \left[\beta^I_{0,t} \quad \beta^A_{0,t} \quad \beta^I_{1,t} \quad \beta^A_{1,t} \quad \dots \quad \beta^I_{K,t} \quad \beta^A_{K,t}\right]'$,

Z_t is a $2(K+1)\times 2(L+1)(K+1)$ block diagonal matrix: $Z_t = I_{2(K+1)}\otimes\left[1\middle|z_t\right]$,

δ is a $2(K+1)(L+1)\times 1$ vector, $\delta = \left[\delta^I_{0,0} \quad \dots \quad \delta^I_{0,L} \quad \delta^A_{0,1} \quad \dots \quad \delta^A_{0,L} \quad \delta^I_{1,0} \quad \dots \quad \delta^A_{K,L}\right]$,

$\eta_t = \left[\eta^I_{0,t}, \eta^A_{0,t}, \eta^I_{1,t}, \eta^A_{1,t}, \dots \eta^I_{K,t}, \eta^A_{K,t}\right]' \sim N(0,\Omega_\eta)$ and $\lambda_i = \left[\lambda^I_i, \lambda^A_i\right]' \sim N(0,\Omega_\lambda)$ are election- and district- random effects.[3]

From (5) - (7), the model can be written as:

$$Y_{i,t} = X_{i,t}Z_t\delta + X_{i,t}\eta_t + \lambda_i + \frac{\varepsilon_{i,t}}{\sqrt{w_{i,t}}} \tag{8}$$

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

In order to estimate the model, I employ a fully Bayesian strategy, treating all unknown quantities as random and specifying prior distributions for all the parameters. The Bayesian approach straightforwardly accommodates problems with small samples typically available for countries with mandatory voting, since it does not rely on asymptotic results for inference (Thum, 2003; Jackman, 2004). In particular, unlike alternative estimation techniques (e.g., Full or Restricted Maximum Likelihood), inference about the fixed effects does not depend on the accuracy of the point estimates of the variance-covariance parameters: they are based on their posterior distribution 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

[3] Throughout this paper, $\otimes$ denotes the left Kronecker product.

the variance parameters are usually imprecisely estimated (Bryk and Raudenbush, 2002).[4,5]

Assuming conditional independence throughout, the model can be specified in a Bayesian context as

$$Y_{i,t} \sim N\left(X_{i,t}\beta_t + \lambda_i,\ \frac{1}{w_{i,t}}\Sigma \right),\ i=1,...,n,\ t=1,...,T \tag{9}$$

$$\beta_t \sim N\left(Z_t\delta, \Omega_\eta\right),\ t=1,...,T \tag{10}$$

$$\lambda_i \sim N\left(0, \Omega_\lambda\right),\ i=1,...,n \tag{11}$$

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

$$\begin{aligned}
&\delta \sim N\left(\delta_0, \Omega_\delta\right), \\
&\Sigma^{-1} \sim Wishart(P, \rho_P), \quad |P| > 0, \rho_P \geq 2 \\
&\Omega_\eta^{-1} \sim Wishart(Q, \rho_Q), \quad |Q| > 0, \rho_Q \geq 2\left(K+1\right) \\
&\Omega_\lambda^{-1} \sim Wishart(R, \rho_R), \quad |R| > 0, \rho_R \geq 2
\end{aligned} \tag{12}$$

and $p\left(w_{i,t} \middle| \upsilon\right)$ depending on the particular normal/independent distribution adopted for the level-1 errors. Routine sensitivity analyses can be performed in order to examine the effect of the hyperparameters on the model fit.

From (9) – (12), and 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

[4] In the context of frequentist estimation techniques, this uncertainty can be taken into account through bootstrapping (Goldstein, 1995) or simulation (King, Tomz and Wittenberg, 2000). However, the fact that the Bayesian approach directly takes into account the uncertainty in variance components makes it particularly appropriate for this kind of analysis.

[5] In addition, as shown by Browne and Draper (2001), Maximum Likelihood methods are susceptible to convergence problems in two-level random-coefficients regression models with few higher-level units.

$$
\begin{aligned}
f\left(\beta,\lambda,\delta,\Sigma,\Omega_\eta,\Omega_\lambda,w,\upsilon \middle| Y\right) \propto & \left[\prod_{i=1}^{n}\prod_{t=1}^{T} w_{i,t}\right] \left|\Sigma\right|^{-nT/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t}\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_i\right)'\Sigma^{-1}\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_i\right)\right\} \\
& \times \prod_{i=1}^{n}\prod_{t=1}^{T} p\left(w_{i,t}\middle|\upsilon\right)\times p(\upsilon)\times\left|\Omega_\eta\right|^{-T/2}\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\beta_t-Z_t\delta\right)'\Omega_\eta^{-1}\left(\beta_t-Z_t\delta\right)\right\} \times \left|\Omega_\lambda\right|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=i}^{n}\lambda_i'\Omega_\lambda^{-1}\lambda_i\right\} \\
& \times \left|\Omega_\delta\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\delta-\delta_0\right)'\Omega_\delta^{-1}\left(\delta-\delta_0\right)\right\} \times \left|\Sigma^{-1}\right|^{\frac{\rho_P-2-1}{2}}\exp\left\{-\frac{1}{2}tr\left(P^{-1}\Sigma^{-1}\right)\right\} \\
& \times \left|\Omega_\eta^{-1}\right|^{\frac{\rho_Q-2(K+1)-1}{2}}\exp\left\{-\frac{1}{2}tr\left(Q^{-1}\Omega_\eta^{-1}\right)\right\} \times \left|\Omega_\lambda^{-1}\right|^{\frac{\rho_R-2-1}{2}}\exp\left\{-\frac{1}{2}tr\left(R^{-1}\Omega_\lambda^{-1}\right)\right\}
\end{aligned}
\tag{13}
$$

Distribution (13) is intractable analytically, but inference on the parameters of interest can be performed by Markov chain Monte Carlo (MCMC) simulations, using Gibbs sampling to repeatedly draw samples from each unknown parameter’s full conditional posterior distribution in order to form the marginal distributions used for Bayesian inference (Gelfland and Smith, 1990; Casella and George, 1992). In order to implement the Gibbs sampler, I subdivide the entire set of unknowns in (13) in such a way that it is possible to sample from the conditional posterior of each subset of unknowns given the other subsets and the data. This leads to an iterative scheme whereby, given an arbitrary set of starting values, samples are drawn from each full conditional posterior given the data and the most recently sampled values for the other unknowns (Gelfland, Hills, Racine-Poon, and Smith, 1990; Seltzer et al., 2002). Under mild regularity conditions (Geman and Geman, 1984), samples from these complete conditionals approach samples from the marginals for a sufficiently large number of iterations. The power and simplicity of the Gibbs sampler in handling complex hierarchical models involving covariates makes it an attractive option against alternative Bayesian/empirical Bayesian methodologies that must often rely on “…a number of approximations whose consequences are often unclear under the multiparameter likelihoods induced by the modeling “ (Gelfland et al., 1990, p. 978).

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:

$$\beta_t \left| Y, \lambda, \delta, \Sigma, \Omega_\eta, \Omega_\lambda, w, \upsilon \sim N\left(b_t, B_t\right), \quad t = 1, \ldots T, \right.$$

$$b_t = \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} X_{i,t} + \Omega_\eta^{-1} \right]^{-1} \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} \left(Y_{i,t} - \lambda_i \right) + \Omega_\eta^{-1} Z_t \delta \right], \tag{14}$$

$$B_t = \left[\sum_{i=1}^{n} w_{i,t} X_{i,t}{}' \Sigma^{-1} X_{i,t} + \Omega_\eta^{-1} \right]^{-1}$$

$$\lambda_i \left| Y, \beta, \delta, \Sigma, \Omega_\eta, \Omega_\lambda, w, \upsilon \sim N\left(d_i, D_i\right), \quad i = 1, \ldots n, \right.$$

$$d_i = \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} + \Omega_\lambda^{-1} \right]^{-1} \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} \left(Y_{i,t} - X_{i,t} \beta_t \right) \right], \tag{15}$$

$$D_i = \left[\Sigma^{-1} \sum_{t=1}^{T} w_{i,t} + \Omega_\lambda^{-1} \right]^{-1}$$

$$\delta \left| Y, \beta, \lambda, \Sigma, \Omega_\eta, \Omega_\lambda, w, \upsilon \sim N\left(\left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} Z_t + \Omega_\delta^{-1} \right]^{-1} \left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} \beta_t + \Omega_\delta^{-1} \delta_0 \right], \left[\sum_{t=1}^{T} Z_t{}' \Omega_\eta^{-1} Z_t + \Omega_\delta^{-1} \right]^{-1} \right) \right. \tag{16}$$

$$\Sigma^{-1} \left| Y, \beta, \lambda, \delta, \Omega_\eta, \Omega_\lambda, w, \upsilon \sim Wishart\left(\left[\sum_{t=1}^{T} \sum_{i=1}^{n} w_{i,t} \left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i \right) \left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i \right)' + P^{-1} \right]^{-1}, \rho_P + nT \right) \right. \tag{17}$$

$$\Omega_\eta^{-1} \left| Y, \beta, \lambda, \delta, \Sigma, \Omega_\lambda, w, \upsilon \sim Wishart\left(\left[\sum_{t=1}^{T} \left(\beta_t - Z_t \delta \right) \left(\beta_t - Z_t \delta \right)' + Q^{-1} \right]^{-1}, \rho_Q + T \right) \right. \tag{18}$$

$$\Omega_\lambda^{-1} \left| Y, \beta, \lambda, \delta, \Sigma, \Omega_\eta, w, \upsilon \sim Wishart\left(\left[\sum_{i=1}^{n} \lambda_i \lambda_i{}' + R^{-1} \right]^{-1}, \rho_R + n \right) \right. \tag{19}.$$

To complete the specification for a Gibbs sampling scheme, the full conditional posterior distributions of w and υ are required. For each element of w the fully conditional posterior density is:

$$w_{i,t} \left| Y, \beta, \lambda, \delta, \Sigma, \Omega_\eta, \Omega_\lambda, \upsilon \propto w_{i,t} \exp\left\{ -\frac{\omega_{i,t}}{2} \left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i \right)' \Sigma^{-1} \left(Y_{i,t} - X_{i,t} \beta_t - \lambda_i \right) \right\} \times p\left(w_{i,t} \middle| \upsilon \right) \right. \tag{20}.$$

For υ, the density is:

$$\upsilon | Y,\beta,\delta,\gamma,\Sigma,\Omega_\eta,\Omega_\lambda,w,\tau \propto p(\upsilon)\prod_{i=1}^{n}\prod_{t=1}^{T} p\left(w_{i,t} | \upsilon\right) \quad (21).$$

From (14) – (21), it is clear that, assuming Normal level-1 residuals (i.e., if all the $w_{i,t}$, $i=1,...,n$, $t=1,...,T$, have degenerate distributions at 1), the conjugacy of the prior distributions at each stage of the hierarchy leads to closed-form full conditional distributions for each parameter of the model, and it is thus straightforward to sample from them in order to obtain the marginal distributions. However, the assumption of Normal level-1 residuals makes inferences vulnerable to the presence of outliers (Andrews and Mallows, 1974; Pinheiro, Liu and Wu, 2001). Assuming a bivariate Student-t prior for $Y_{i,t}$ allows for the possibility of extreme observations, attenuating the influence of outliers (Berger, 1985; Gelman, Carlin, Stern and Rubin, 2004) and providing a valuable tool with which to assess the sensitivity of inferences to prior distributional assumptions (Carlin and Louis, 1996; Thum, 1997).

A bivariate Student t prior for $Y_{i,t}$ can be obtained from the normal/independent distribution by assuming $w_{i,t} | \upsilon \sim Gamma(\upsilon/2, \upsilon/2)$, $w_{i,t} > 0, \upsilon > 0$.[6] The fully conditional posterior densities (20) and (21) then become:

$$w_{i,t} | Y,\beta,\lambda,\delta,\Sigma,\Omega_\eta,\Omega_\lambda,\upsilon \sim Gamma\left(\frac{\upsilon}{2}+1,\ \frac{1}{2}\left[\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_t\right)'\,\Sigma^{-1}\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_t\right)+\upsilon\right]\right) \quad (20')$$

$$\upsilon | Y,\beta,\lambda,\delta,\Sigma,\Omega_\eta,\Omega_\lambda,w \sim \left[2^{\upsilon/2}\Gamma\left(\frac{\upsilon}{2}\right)\right]^{-nT} \upsilon^{\frac{\upsilon nT}{2}} \exp\left\{-\frac{\upsilon}{2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t} - \log\left(w_{i,t}\right)\right)\right\} \quad (21').$$

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

[6] I use the parametrization of the gamma distribution found in Rosa et al., 2003.

that might be masked in standard residual plots (Box, 1979; Seltzer et al., 2002; Congdon, 2003). Note that, from (20'),

$$E\left(w_{i,t}\middle|Y,\beta,\lambda,\gamma,\Sigma,\Omega_\eta,\Omega_\lambda,\upsilon\right)=\frac{\upsilon+2}{\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_i\right)'\,\Sigma^{-1}\left(Y_{i,t}-X_{i,t}\beta_t-\lambda_i\right)+\upsilon} \tag{22},$$

so that for a large enough υ, $E\left(w_{i,t}\middle|Y,\beta,\lambda,\gamma,\Sigma,\Omega_\eta,\Omega_\lambda,\upsilon\right)\to 1$, and approximately normal tails are obtained for the level-1 errors. However, for low values of υ, the expected value 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. Therefore, 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.[7] Although (21') does not have a closed form, this conditional posterior distribution can be approximated by discretizing the density along a grid of values and then sampling from the resulting discrete distributions. When the points in the grid are spaced closely together, the discrete distribution of υ provides an accurate approximation to the full conditional distribution (Draper, 2001; 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, Best, Carlin and van der Linde, 2002; Gelman et al., 2004). The means and standard deviation of the convergent Gibbs samples generated from (14)-(21) under each variant of the model can be used to summarize the posterior distributions of the parameters. These marginal posterior distributions, however, are of no direct interest for the analysis. Rather, interest lies in the effect of the explanatory variables on the proportion of invalid voting and electoral absenteeism. I compute the impact of each of the district-level and country-level regressors on $P^I_{i,t}$ and $P^A_{i,t}$ using average predictive comparisons (Katz and King,

[7] A detailed discussion of this point is provided in Seltzer et al., 2002.
[8] Alternatively, a strategy based on Metropolis-Hastings sampling can be incorporated into the MCMC scheme to obtain draws from υ (Seltzer et al., 2002; Gelman et al., 2004).

1999; King, Tomz and Wittenberg, 2000; Gelman and Hill, 2007). The algorithm implemented to estimate these causal effects is detailed in Appendix I.

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 capacity and flexibility to deal with nested unbalanced data sets is one additional advantage of Bayesian multilevel models versus more traditional frequentist approaches (Bryk and Raudenbush, 2002; Shor et al., 2007). Also, a more complex specification for the components of Σ could be adopted (e.g., allowing for serial correlation of the level-1 errors - Allenby and Lenk, 1994). 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, as mentioned above, although I focus on two particular variants of the mixed model – i.e., with Normal and Student-t level-1 errors – assuming alternative densities for $w_{i,t}$ would allow obtaining 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.

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 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 experienced by Brazil in its recent history and the sharp differences in socio-demographic characteristics among its states allow examining the impact of different factors on invalid voting and absenteeism.[9] In order to illustrate the use of the model presented in Section 3 and to compare the results with those obtained using alternative modeling strategies, I analyze all lower house elections held in the country between 1945 and 2006. The dataset has an unbalance structure, with 388 observations for 27 states across 16 elections.[10]

The dependent variables of interest for the analysis are the proportion of invalid votes and electoral absenteeism in lower house elections. The proportion of invalid votes among the electorate is computed as the ratio of blank and null votes cast over the population eligible to vote. Electoral absenteeism is calculated as the percentage of potential voters failing to comply with their duty. Figure 2 presents the proportion of invalid voting and absenteeism by state for the elections held between 1945 and 2006. As can be seen, there is considerable variation in the two sources of abstention both between states and within states across elections.[11]

[Figure 2]

In line with the different theories under consideration, socioeconomic, institutional and protest variables are included as explanatory variables in the 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

[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).
[10] The number of states in Brazil increased from twenty-two to twenty-seven during this period.
[11] The proportions are calculated based on the number of elections held in each state.

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, *Electoral Manipulation* measures the degree of electoral and political "engineering", coded by Power and Roberts (1995) on a four point-scale ranging from 0 for free elections 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-level and country-level predictors for the period 1945-2006.

[Table 1 here]

The characterization and measurement of the independent variables closely follows Power and Roberts (1995); their data is complemented with information from IUPERJ (2006) for the 1994-2006 elections. The only difference with the authors lies in the definition of *Illiteracy*: while they 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 (Brazilian Institute of Geography and Statistics, 2003) and 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* and model the random-coefficients of *Illiteracy* as functions of it, allowing the effect of *Illiteracy* to vary across elections.

In addition, in line with Power and Roberts' (1995) argument that the country-level predictors *Ballot*, *Electoral Manipulation*, *Growth* and *Inflation* affect the proportion of

[12] Prior to the introduction of the single official ballot ("cedula unica") in 1962, candidates distributed their own pre-printed ballots, which voters just had to place 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).

invalid voting and absenteeism in each state-year, I specify the election random-intercepts $\beta_{0,t} = \left[\beta_{0,t}^{I}, \beta_{0,t}^{A}\right]'$ 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 specified as fixed effects (i.e., their variation across elections is constrained to be 0), although the model could be written more generally to accommodate various plausible design alternatives for parametrizing these coefficients.

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

$$\begin{aligned} Y_{i,t}^{s} &= \beta_{0,t}^{s} + \beta_{1,t}^{s} Illiteracy_{i,t} + \beta_{2,t}^{s} Urbanization_{i,t} + \beta_{3,t}^{s} FEAP_{i,t} + \\ &\quad \beta_{4,t}^{s} Candidates\ per\ Seat_{i,t} + \beta_{5,t}^{s} Electorate_{i,t} + \lambda_{i}^{s} + \frac{\varepsilon_{i,t}^{s}}{\sqrt{w_{i,t}}}, \qquad s = I, A \end{aligned} \tag{23}$$

$$\beta_{0,t}^{s} = \delta_{0,0}^{s} + \delta_{0,1}^{s} Ballot_{t} + \delta_{0,2}^{s} E.\ Manipulation_{t} + \delta_{0,3}^{s} Growth_{t} + \delta_{0,4}^{s} Inflation_{t} + \eta_{0,t}^{s}, \quad s = I, A \tag{24}$$

$$\beta_{1,t}^{s} = \delta_{1,0}^{s} + \delta_{1,1}^{s} Franchising_{t} + \eta_{1,t}^{s}, \qquad s = I, A \tag{25}$$

$$\beta_{k,t}^{s} = \delta_{k,0}^{s} \qquad s = I, A; k = 2,...,5 \tag{26}$$

with $\left[\varepsilon_{i,t}^{I}, \varepsilon_{i,t}^{A}\right]' \sim N(0,\Sigma)$, $\eta_{t} = \left[\eta_{0,t}^{I}, \eta_{0,t}^{A}, \eta_{1,t}^{I}, \eta_{1,t}^{A}\right]' \sim N\left(0, \Omega_{\eta}\right)$, $\left[\lambda_{i}^{I}, \lambda_{i}^{A}\right] \sim N\left(0, \Omega_{\lambda}\right)$,

and

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

The model was fit using WinBUGS 1.4, as called from R 2.4.1.[13] All the hyperparameters in the model 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 (Congdon, 2003). 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

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

values reached approximate convergence after 25,000 iterations, with a burn-in of 5,000 iterations; the results reported below are based on 1,000 samples of the pooled chains of deviates.[14]

4.2. Results of the compositional-hierarchical model

Table 2 below reports the posterior means and 90% confidence intervals for the fixed effects for 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.[15] The values of the Deviance Information Criterion (DIC) for both models and the Bayes Factor for Model 1-b relative to Model 1-a are also presented.

[Table 2 here]

Table 2 shows considerable disparity in the posterior means and confidence 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.[16] 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. The evidence presented in Figures 3 and 4 further support 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 (Chaloner and Brant, 1988; Weiss, 1994). 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.

[Figure 3 here]

---

[14] Approximate convergence is achieved for values of Gelman and Rubin's (1992) estimated Potential Scale Reduction factor below 1.1.

[15] 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.

[16] 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).

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 (Congdon, 2003; Rosa et al, 2003). In addition, given that the "weight parameters" also reduce the influence of extreme observations on the posterior distribution of the random parameters– equations (14) and (15) - , the number of (multivariate) level-2 election outliers (Weiss, 1994) in Model 1-b is also halved with respect to Model 1-a (Figure 5).

[Figure 4 here]

[Figure 5 here]

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. Table 3 reports the posterior distribution of the covariance components from the chosen model. The mean posterior correlation between the level-1 errors is moderately positive (0.24) and statistically significant at the 0.01 level, contradicting the assumption of no correlation underlying separate univariate analyses of invalid voting and absenteeism. Hence, 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. 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 correlation 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, suggesting that election-specific circumstances have a substantial influence on both forms of abstention.

[Table 3 here]

Based on the convergent Gibbs samples of the parameters of Model 1-b, I estimate the average effect of a one-unit change in each of the state-level and national-level predictors on the proportion of invalid ballots and electoral absenteeism.[17] 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 levels of education and skills 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, and it further rose by more than 6 points on average with the extension of suffrage to illiterates in 1985. The addition of new voters was also positively related to invalid ballots: 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. 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 variable on absenteeism was much more variable across states and elections. The positive and significant effect of *Inflation* on invalid voting suggests that blank and null ballots might reflect not only popular dissatisfaction with inadequate representative institutions (Schwartzman, 1973), but also discontent with poor macroeconomic performance and economic mismanagement by the political elites.[18] While these results provide evidence in support of 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

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

[18] High and persistent inflation rates have characterized the Brazilian economy throughout the 20th century, and price stabilization has been the major macroeconomic concern for Brazilian policy-makers, although usually with disastrous results (Langoni, 1997).

conclusive in the case of electoral absenteeism, underscoring the need to examine additional factors that might affect noncompliance with compulsory voting laws and to distinguish between the determinants of both sources of abstention from a theoretical perspective.

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 running for office and the introduction of the single official ballot, have a positive impact on increase invalid voting but a negative average effect 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 influence 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 vote, the proliferation of candidates and the requirement that voters record their preferred candidate's name or registration number on the paper ballot tend to increase invalid voting, probably because they impose considerable informational requirements 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 20th century.

[Table 4 here]

4.3 Comparison with alternative modeling approaches

In order to illustrate the differences between the model presented here and alternative approaches used to analyze abstention in compulsory voting systems, Figure 6 below contrasts the average causal effects of the predictors on invalid voting and absenteeism under Model 1-b with those obtained under three models that fail to account for the compositional and/or the hierarchical structure of the data. Model 2 uses separate ordinary least squares regressions for invalid voting and absenteeism, assuming independence among observations and simply pooling state-level and country-level predictors by assigning the values of the national variables to all the states in a given election. Model 3 uses separate hierarchical linear models for invalid voting and absenteeism, accounting for the temporal and geographical clustering of the data but ignoring the non-negativity and unit-sum constraints (1) and (2). 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. The specifications of Models 2, 3 and 4 are detailed in Appendix II. [19]

[Figure 6 here]

The results reported in Figure 6 shows some noticeable differences between the four models. As seen in the upper and lower panels, the standard errors of the marginal effects of the covariates on both sources of abstention under Model 1-b tend to be considerably smaller than for Model 3 and much larger than for Models 2 and 4, particularly in the case of the country-level variables. This leads to different conclusions about the relative size and the statistical significance of the impact of the national-level predictors 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 on both sources of abstention at the 0.01 level. In contrast, *Growth* has no systematic effect on either source of abstention under Models 1 and 3. At the other extreme, the large standard errors for the country-level comparisons under Model 3 determine that none of national-level variables has a significant effect on either source of abstention at the usual confidence levels.

More importantly, the four models lead to different substantive conclusions regarding the impact of some of the variables on the two sources of abstention. As seen in the

[19] Models 3 and 4 were fitted by MCMC simulations (Gibbs Sampling), using a normal/independent distribution for the data model, Gaussian priors for the random coefficients and diffuse conjugate priors for the hyperparameters. The substantive results remain unchanged if Gaussian level-1 errors are assumed. Details of the estimation are available from the author upon request.

lower panel of Figure 5, the results from Models 2 and 4 show 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 between 18 and 70 years of age, voting is optional for illiterates. While inferences drawn from these two models tend to support the claim that "the fact that voting is…optional for illiterates seems to have little practical effect on their observance of mandatory voting" (Power and Roberts, 1995, p. 800), the average effect of *Franchising* on electoral absenteeism has the opposite sign under Model 1-b. Also, while 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, this relationship is negative under Models 3 and 4. Finally, under Model 2, *Ballot* has a negative and statistically significant effect on invalid voting, leading to the rather implausible conclusion that the introduction of a more complex ballot system that requires considerable more information on the part of voters resulted in lower rates of blank and spoiled ballots. These examples illustrate the fact that some of the inferences drawn from the model developed in this paper contradict the results both from the separate univariate analyses (Models 2 and 3) and from an analysis that ignores election-to-election variability in both sources of abstention beyond that explained by national-level variables (Model 4). The conflicting results from the different models lead to different conclusions about the relative validity of the alternative theories proposed to account for abstention under mandatory voting and might entail very 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 (McAllister and Makkai, 1993; Power and Roberts, 1995).

In order to compare the fit of the four models, I use posterior predictive simulations (Gelman et al., 2004, Gelman and Hill, 2007). 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 $d\left(P^{\mathrm{Rep}}, P^{Obs}\right) = E\left(\left\|P^{\mathrm{Rep}} - P^{Obs}\right\|^2 \middle| P^{Obs}\right)$, where $P_{i,t}^{Rep} = \left(P_{i,t}^{Rep(1)}, \ldots, P_{i,t}^{Rep(J)}\right)$ denotes the replicate 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.[20] The posterior predictive loss d can then be estimated as:

$$\hat{d} = \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\frac{1}{J} \sum_{j=1}^{J} \left\| P_{i,t}^{Obs} - P_{i,t}^{\mathrm{Rep}(j)} \right\|^2 \right). \tag{27}$$

Table 5 reports the estimates and 90% confidence 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, Model 2, which in addition ignores the multilevel nature of the data, exhibits the worst fit.

[Table 5 here]

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 $P_{i,t}^{\mathrm{Rep}(j)}$, $j = 1,...,1000,$ 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 a vis* Model 4. 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 reveals that the methodological differences between these competing empirical strategies have substantial consequences in terms of the analysis of the determinants of abstention under compulsory voting.

[Figure 7 here]

---

[20] In the case of the compositional-hierarchical model, $P_{i,t}^{\mathrm{Rep}}$ are obtained from $Y_{i,t}^{Rep}$ using the logarithmic transformations (3) and (4).

5. Concluding remarks

Different theories, drawing on the literature on voter turnout in industrialized democracies, have been proposed in order to account for the phenomena of invalid voting and electoral absenteeism under mandatory voting. 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 in this paper accounts for the compositional and hierarchical structure of district-level electoral data and easily accommodates sensitivity analysis, encompassing a family of thick-tailed distributions that can be used for robust inference.

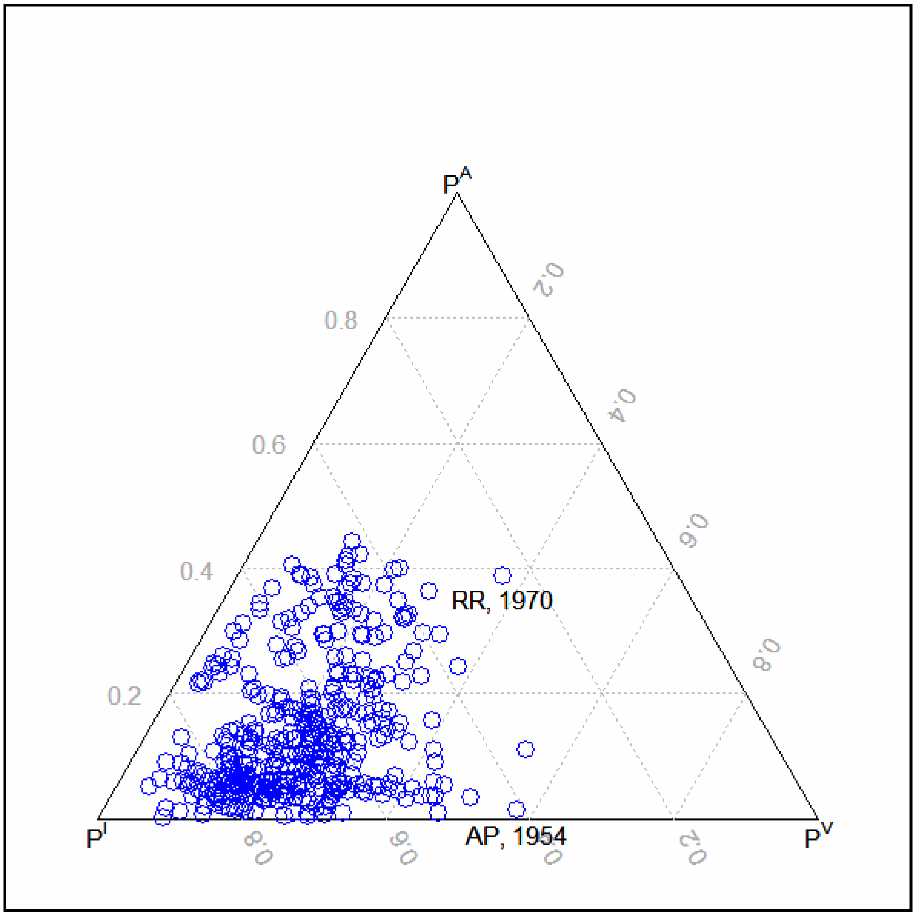
Results obtained from the application of the model to analyze abstention in Brazil's legislative elections allow drawing interesting substantive and methodological conclusions. The evidence presented above reveals substantial differences in the determinants of both forms of non-voting. In line with Power 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 considerably better than several alternative empirical strategies used to analyze abstention under compulsory voting. More importantly, the main conclusions and the policy implications resulting from the compositional-hierarchical model might differ significantly from those drawn using less appropriate modeling approaches prevailing in previous research in this area.

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. Also, the mixed model presented in Section 3 can be modified in order to accommodate other possible distributions of the error terms at each level of the hierarchy (Andrew and Mallows, 1974; West, 1984; Seltzer et al., 2002; Rosa et al. 2003). An immediate extension of the paper would be to include a larger number of countries and additional covariates in order to analyze the performance of the model and the robustness of the results from a comparative politics 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.

Tables and Figures

Figure 1

Proportion of invalid votes, electoral absenteeism and valid votes in Brazil's lower house elections, 1945 – 2006

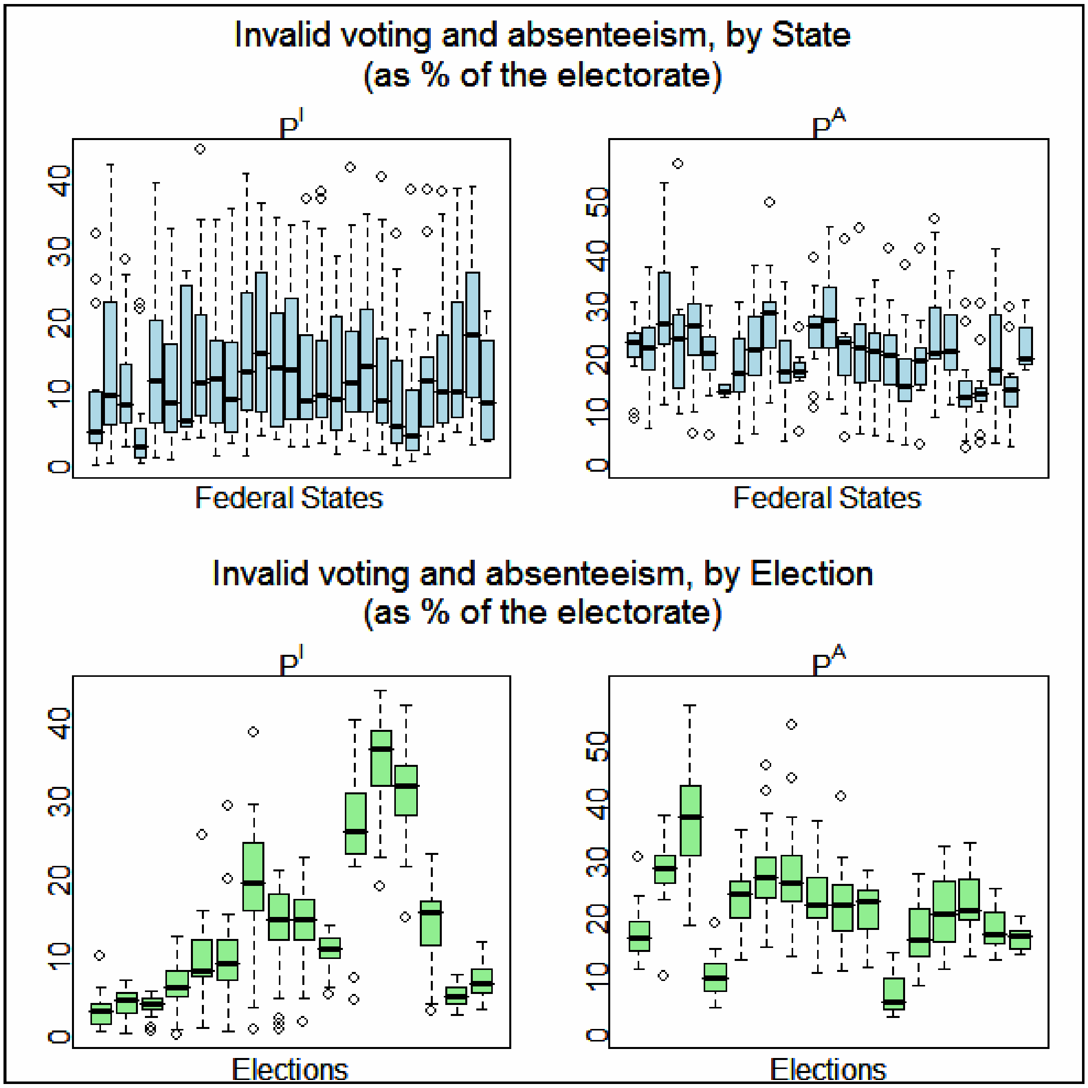


Note: Each circle in the figure indicates the values of P^I , P^A and P^V in a particular district for a given election. P^I is measured on the scale in the triangle's left side, P^A is measured on the right side, and P^V is measured on the scale in the triangle's base.

Figure 2

Invalid voting and absenteeism by state and election, as a % of the electorate

Lower house elections, 1945 - 2006



Sources: Banco de Dados Eleitorais Do Brasil, Instituto Universitario de Pesquisas de Rio de Janeiro (IUPERJ); Power and Roberts (1995).

Table 1

Summary statistics – Independent variables

Variable	Mean	Std. Dev.	Min	25th percentile	75th 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					

Sources: Banco de Dados Eleitorais Do Brasil, Instituto Universitario de Pesquisas de Rio de Janeiro (IUPERJ); Power and Roberts (1995).

Table 2

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

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, 0.06)	0.01 (-0.01, 0.02)	0.02 (-0.01, 0.04)	0.01 (-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)
Electoral 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[1]	557.90		242.30	
Bayes factor[2]	-		9.79×10^6	

[1] The Deviance Information Criterion (DIC) is computed as:

$$2\left(\frac{1}{J}\sum_{j=1}^{J} -2\log p\left(y\middle|\theta^{(j)}\right)\right)+2\log p\left(y\middle|\bar{\theta}\right),$$

with $\bar{\theta} = E(\theta | y)$, the posterior mean of the model's parameters. Lower values of the DIC indicate better fit to the data.

[2] The Bayes factor for model M_j relative to model M_k is given by

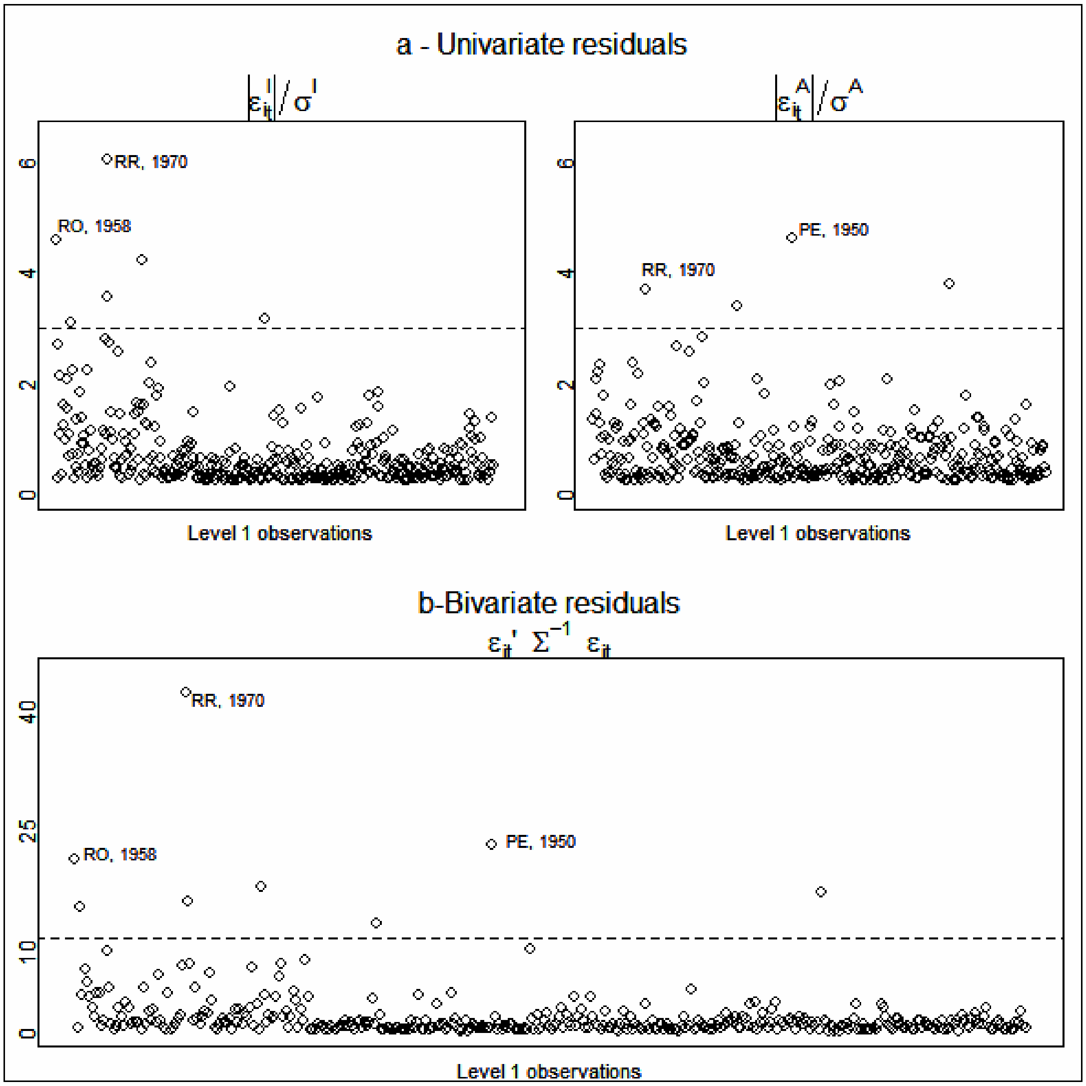
$$B_{j,k} = \frac{p(y|M_j)}{p(y|M_k)} = \frac{\int p(y|M_j, \theta_j)\, p(\theta_j|M_j)\, d\theta_j}{\int p(y|M_k, \theta_k)\, p(\theta_k|M_k)\, d\theta_k}.$$

I use the harmonic mean of the likelihood evaluated at the posterior draws of the parameters (Newton and Raftery, 1994; Rosa et al., 2003) as an estimate for $p(y|M_x)$, $x = j,k$:

$$\hat{p}(y \mid M_x) = \left(R^{-1} \sum_{r=1}^{R} p\left(y \mid \theta_x^{(r)}\right)^{-1} \right)^{-1}$$

Figure 3

Posterior means of the level-1 residuals from Model 1-a



Note: The standardized univariate and bivariate level-1 residuals in Figure 3-a are computed based on the Bayesian outlier 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 $\frac{1}{J}\sum_{j=1}^{J}\left(Y_{i,t}-X_{i,t}\beta_{t}^{(j)}-\lambda_{i}^{(j)}\right)'\Sigma^{-1(j)}\left(Y_{i,t}-X_{i,t}\beta_{t}^{(j)}-\lambda_{i}^{(j)}\right)$.

For the univariate residuals, the dashed horizontal lines correspond to the threshold of 3. For the bivariate residuals, the cutoff point is determined as $k=\chi^{2}_{2(1-\alpha)}$, $\alpha=2\times\Phi(-3)$.

Figure 4

Marginal posterior densities of υ and $w_{i,t}$ under Model 1-b

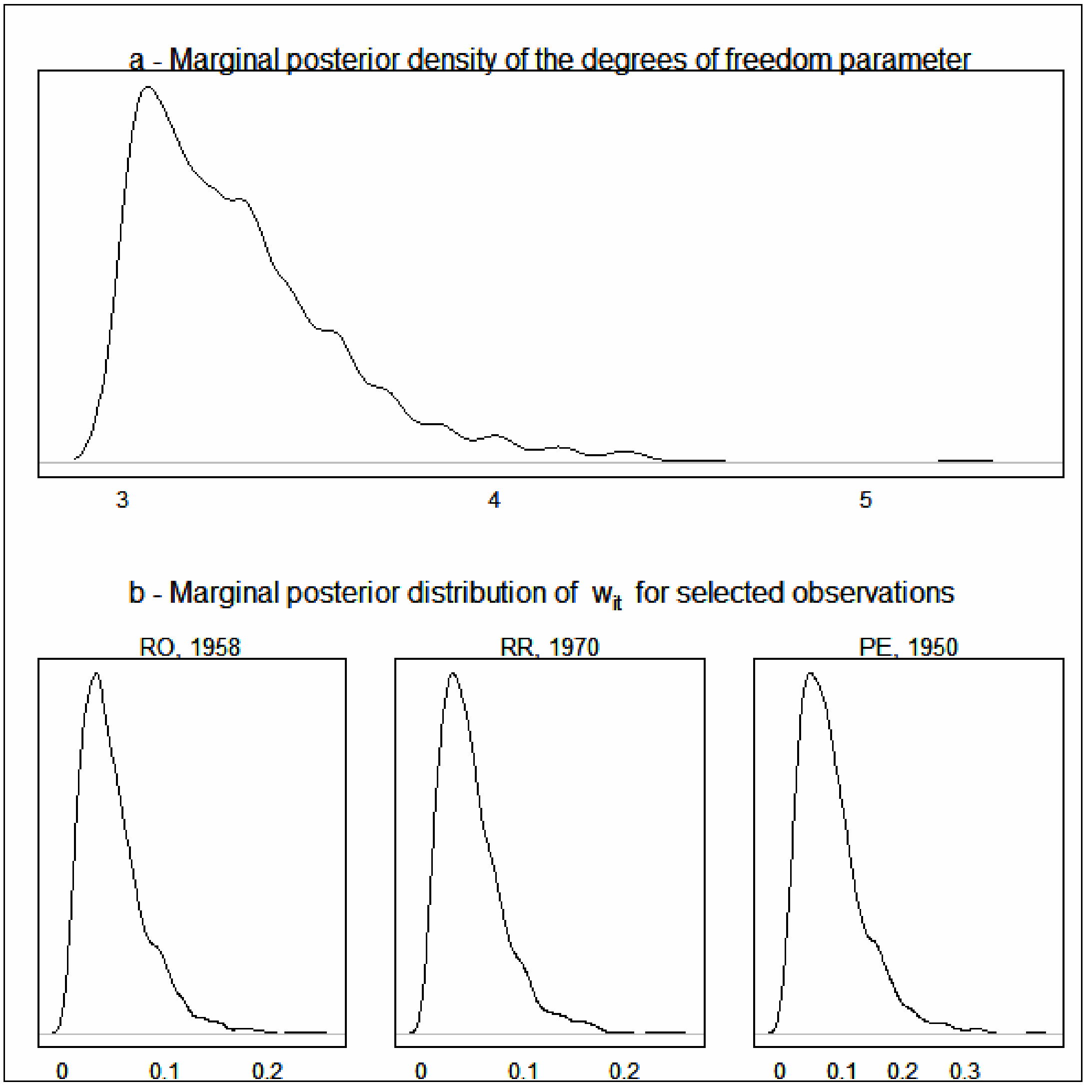
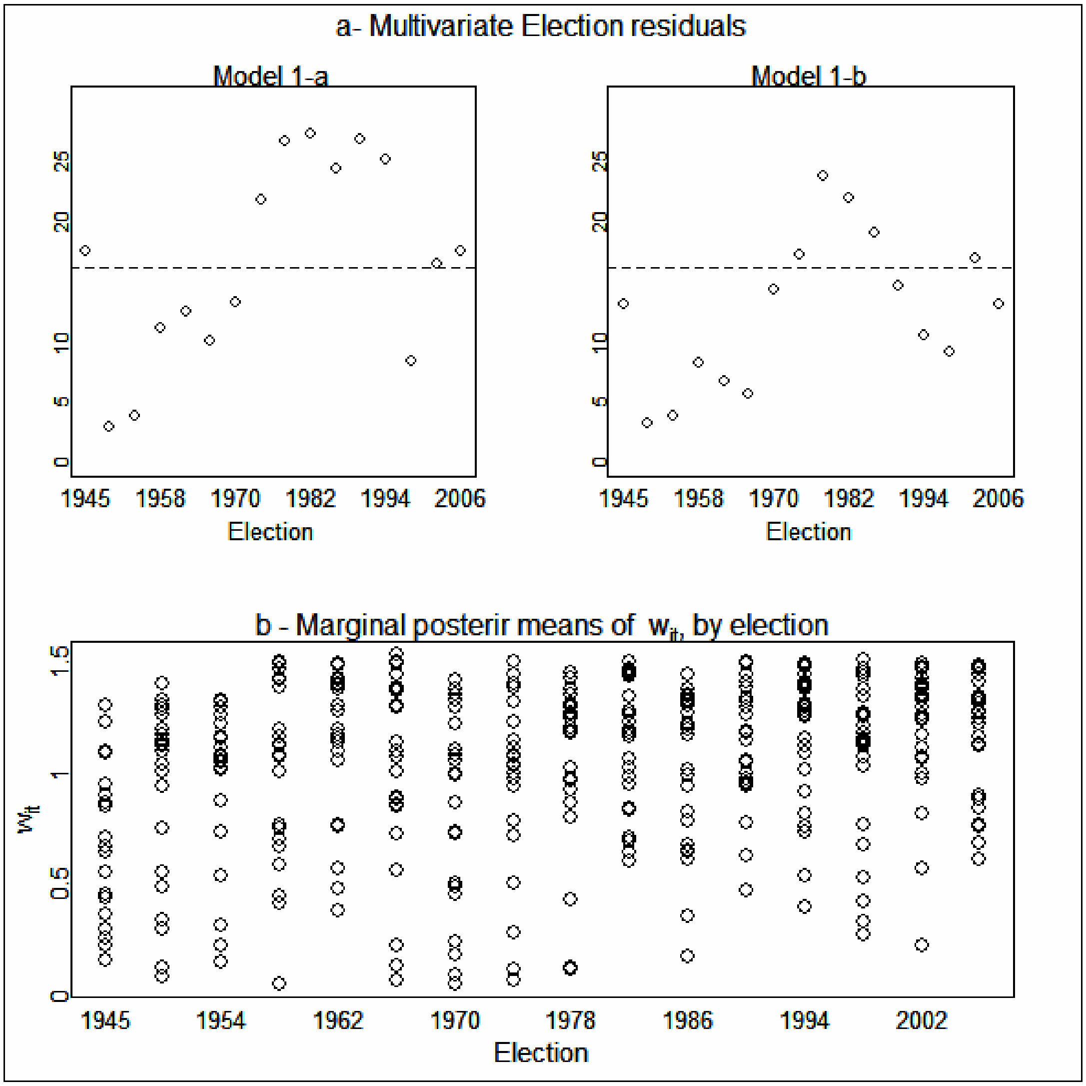


Figure 5



Note: Figure 5-a plots 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 correspond to the cutoff point $k = \chi^2_{4(1-\alpha)}$, $\alpha = 2\times\Phi(-3)$. Figure 5-b plots the marginal posterior means of the weight parameters $w_{i,t}$, by election.

Table 3

Posterior means of variance-covariance components under Model 1-b

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)

Note: 90% confidence intervals reported in parenthesis.

Table 4

Effect of a one-unit change in the predictors on invalid voting and absenteeism under Model 1-b (in percentage points)[1,2]

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)
Females in EAP	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	3.83*** (1.24)	-0.82 (1.72)

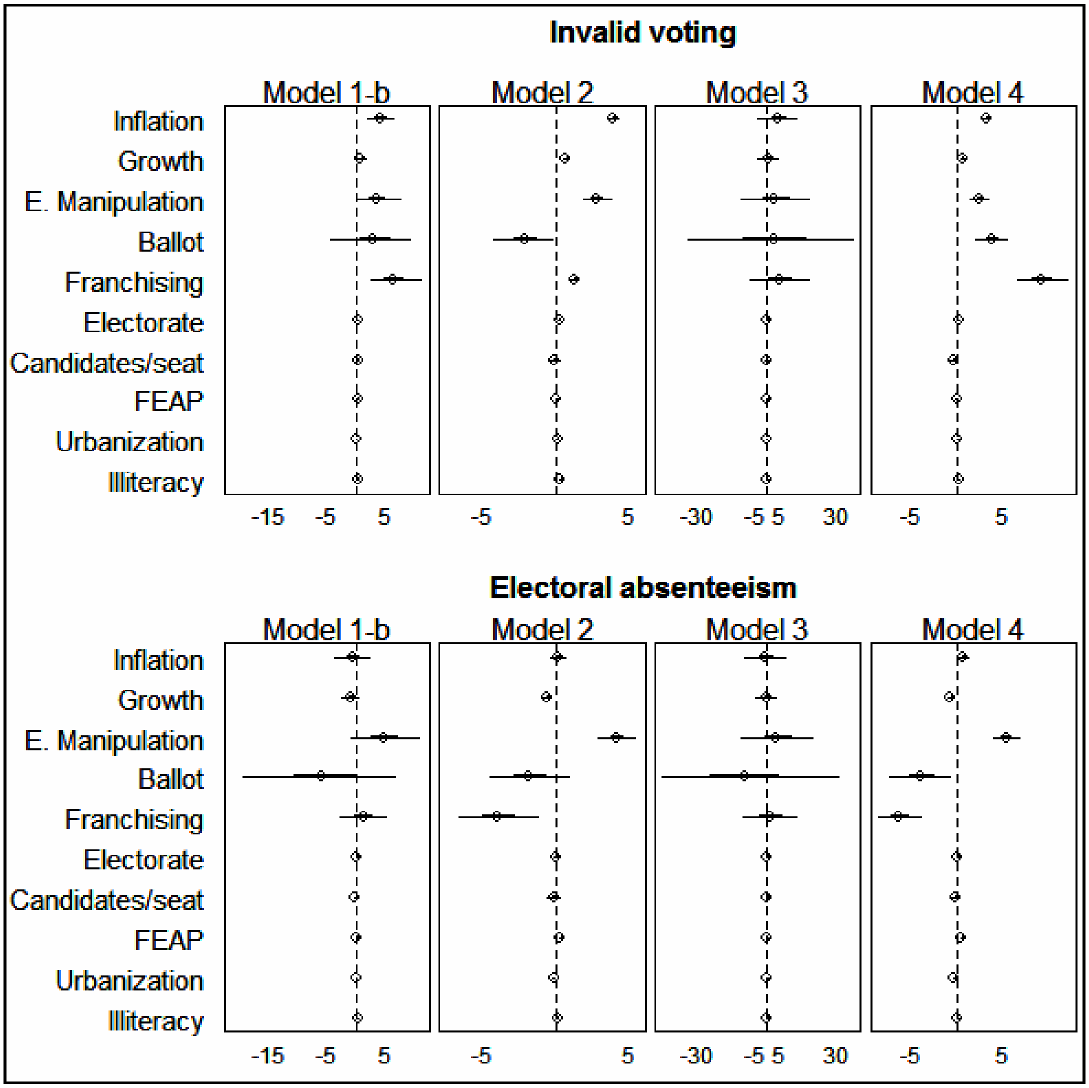
[1] Standard errors are reported in parenthesis.

[2] Significance levels: *** 0.01, ** 0.05, *0.1.

Figure 6

Estimated marginal effects of the predictors across models

(in percentage points)



Note: The graph shows the effect of a one-unit change in each of the predictors on invalid voting and electoral absenteeism. The center dots correspond to the point estimates, the thicker lines to the 50% confidence interval, and the thinner lines to the 90% confidence interval.

Table 5

Estimates of the Posterior Predictive Loss

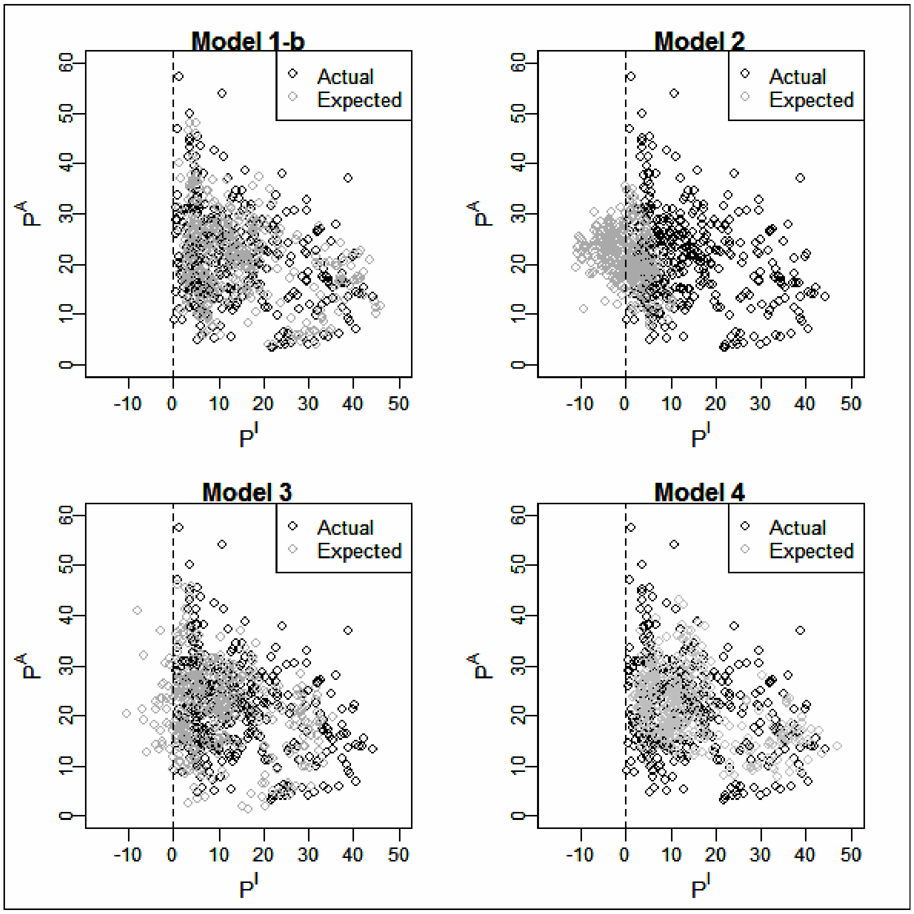
for alternative modeling strategies*

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)

*90% confidence intervals reported in parenthesis.

Figure 7

Actual and expected proportions of invalid voting and electoral absenteeism under alternative modeling strategies



Note: 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.

Appendix I - Algorithm implemented to compute the causal effects

Let $\left(\left\{\eta_t^{(j)}\right\},\left\{\lambda_i^{(j)}\right\},\delta^{(j)},\Sigma^{-1(j)},\Omega_\eta^{-1(j)},\Omega_\xi^{-1(j)},\left\{w_{i,t}^{(j)}\right\},\upsilon^{(j)}\right)$, $j=1,...,J$, denote convergent samples generated from (14)-(21). In order to compute the average effect of each of the independent variables on invalid voting and electoral absenteeism, the following algorithm is implemented (Katz and King, 1999; Bhaumik, Dey and Ravishanker, 2003; Gelman and Hill, 2007):

1. Samples of the estimated expected proportions of invalid voting and absenteeism in each district-year for given covariates are calculated using the additive logistic transformations (3) and (4):

$$\widetilde{P}_{i,t}^{I(j)}=\frac{\exp\left[\widetilde{Y}_{i,t}^{I(j)}\right]}{1+\exp\left[\widetilde{Y}_{i,t}^{I(j)}\right]+\exp\left[\widetilde{Y}_{i,t}^{A(j)}\right]},\quad j=1,...,J$$

$$\widetilde{P}_{i,t}^{A(j)}=\frac{\exp\left[\widetilde{Y}_{i,t}^{A(j)}\right]}{1+\exp\left[\widetilde{Y}_{i,t}^{I(j)}\right]+\exp\left[\widetilde{Y}_{i,t}^{A(j)}\right]},\ j=1,...,J$$

where:

$$\widetilde{Y}_{i,t}^{s(j)}=\delta_{0,0}^{s(j)}+\sum_{k=1}^{K}\delta_{k,0}^{s(j)}x_{i,t,k}+\sum_{l=1}^{L}\delta_{0,l}^{s(j)}z_{t,l}+\left[\sum_{k=1}^{K}\left(\sum_{l=1}^{L}\delta_{k,l}^{s(j)}z_{t,l}\right)x_{i,t,k}\right]+\eta_{0,t}^{s(j)}+\sum_{k=1}^{K}\eta_{k,t}^{s(j)}x_{i,t,k}+\lambda_i^{s(j)},\quad s=I,A,$$

and $x_{i,t},z_t$ are vectors of observed district-level and country-level predictors.

2. Step 1 is repeated after changing the value of the predictor whose effect is analyzed by 1 unit, while keeping all other regressors at their observed levels, obtaining $\widehat{P}_{i,t}^{I(j)}$ and $\widehat{P}_{i,t}^{A(j)}$, $j=1,...J$.
3. The average effect of the predictor on invalid voting and absenteeism for all district-years in the sample can be estimated by averaging $\widehat{P}_{i,t}^{I(j)}-\widetilde{P}_{i,t}^{I(j)}$ and $\widehat{P}_{i,t}^{A(j)}-\widetilde{P}_{i,t}^{A(j)}$ over all $i=1,...,n;\ t=1,...,T;\ j=1,...,J$ (Bhaumik, Dey and Ravishanker, 2003; Gelman and Pardoe, 2007). Confidence intervals summarizing the approximate distribution of the causal effects can also be easily constructed using standard methods from sampling theory (Gelman and Pardoe, 2007).

Appendix II - Alternative strategies to modeling invalid voting and electoral absenteeism

Model 2:

$$P_{i,t}^{s} = \delta_0^s + \left(\delta_1^s + \delta_2^s Franchising_t\right) Illiteracy_{i,t} + \delta_3^s Urbanization_{i,t} + \delta_4^s FEAP_{i,t} + \delta_5^s Candidates\ per\ Seat_{i,t} + \delta_6^s Electorate_{i,t} + \delta_7^s E.\ Manipulation_t + \delta_8^s Ballot_t + \delta_9^s Growth_t + \delta_{10}^s Inflation_t + \varepsilon_{i,t}^s, \quad s=I,A. \tag{II.1}$$

Model 3:

$$P_{i,t}^{s} = \beta_{0,t}^s + \beta_{1,t}^s Illiteracy_{i,t} + \beta_{2,t}^s Urbanization_{i,t} + \beta_{3,t}^s FEAP_{i,t} + \beta_{4,t}^s Candidates\ per\ Seat_{i,t} + \beta_{5,t}^s Electorate_{i,t} + \lambda_i^s + \frac{\varepsilon_{i,t}^s}{\sqrt{w_{i,t}^s}}, \quad s = I, A \tag{II.2}$$

$$\beta_{0,t}^s = \delta_{0,0}^s + \delta_{0,1}^s Ballot_t + \delta_{0,2}^s E.\ Manipulation_t + \delta_{0,3}^s Growth_t + \delta_{0,4}^s Inflation_t + \eta_{0,t}^s, \quad s = I, A \tag{II.3}$$

$$\beta_{1,t}^s = \delta_{1,0}^s + \delta_{1,1}^s Franchising_t + \eta_{1,t}^s \qquad s = I, A \tag{II.4}$$

$$\beta_{k,t}^s = \delta_{k,0}^s \qquad s = I, A; k = 2,...,5 \tag{II.5},$$

with $\varepsilon_{i,t}^s \sim N\left(0,\sigma_{\varepsilon_s}^2\right), \eta_t^s \sim N\left(0,\Omega_{\eta_s}\right), p\left(w_{i,t}^s \middle| \upsilon\right) = Gamma\left(\frac{\upsilon}{2},\frac{\upsilon}{2}\right).$

Model 4:

$$Y_{i,t}^{s} = \delta_0^s + \left(\delta_1^s + \delta_2^s Franchising_t\right) Illiteracy_{i,t} + \delta_3^s Urbanization_{i,t} + \delta_4^s FEAP_{i,t} + \delta_5^s Candidates\ per\ Seat_{i,t} + \delta_6^s Electorate_{i,t} + \delta_7^s E.\ Manipulation_t + \delta_8^s Ballot_t + \delta_9^s Growth_t + \delta_{10}^s Inflation_t + \lambda_i^s + \frac{\varepsilon_{i,t}^s}{\sqrt{w_{i,t}}}, \quad s=I,A \tag{II.6}$$

With $\left[\varepsilon_{i,t}^I, \varepsilon_{i,t}^A\right] \sim N\left(0,\Sigma\right), \left[\lambda_i^I, \lambda_i^A\right] \sim N\left(0,\Omega_\lambda\right), p\left(w_{i,t} \middle| \upsilon\right) = Gamma\left(\frac{\upsilon}{2},\frac{\upsilon}{2}\right),$

and $P_{i,t} = \left[P_{i,t}^I, P_{i,t}^A\right]$ obtained from $Y_{i,t} = \left[Y_{i,t}^I, Y_{i,t}^A\right]'$ using (3) and (4).

References

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

Aitchison, J., and S. Shen. 1980. “Logistic-Normal Distribution: Some Properties and Uses”. *Biometrika*, 67, 261-72.

Aldrich, John. H. 1993. “Rational Choice and Turnout”. *American Journal of Political Science*, 37, 246 – 278.

Allenby, Greg M. and Peter J. Lenk. 1994. “Modeling Household Purchase Behavior with Logistic Normal Regression”. Journal of the American Statistical Association, 89 (428), 1218 – 1231.

Ames, Barry. 1995. “Electoral Strategy under Open-List Proportional Representation”. *American Journal of Political Science*, 39(2), 406-33.

Andrews, D. F., and C. L. Mallows. 1974. “Scale mixtures of normality”. *Journal of the Royal Statistical Society*, Series B, 36, 99-102.

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.

Bandorff-Nielsen, O. E., and B. Jørgensen. 1991. “Some Parametric Models on the Simplex”. *Journal of Multivariate Analysis*, 39, 106 – 116.

Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

Bhaumik, Amitabha, Dipak K. Dey and Nalini Ravishanker. 2003. “A dynamic Linear Model Approach for Compositional Time Series Analysis”. Technical Report, University of Connecticut.

Billheimer, Dean, Peter Guttorp, and William F. Fagan. 2001. “Statistical interpretation of Species Composition”. *Journal of the American Statistical Association*, 96(456), 1205 – 1214.

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

Box, G.E.P. 1979. Robustness in the strategy of scientific model building. In: Robustness in statistics, R. Launer and G. Wilkinson (eds.), 201-236. New York: Academic Press.

Brazilian Institute of Geography and Statistics. 2003. *Estatísticas do Século XX*, Rio de Janeiro: IBGE.

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

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

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

Casella, George and Edward L. George. 1992. "Explaining the Gibbs Sampler". *The American Statistician*, 46(3), 167-174.

Chaloner, Kathryn, and Rollin Brant. 1988. "A Bayesian Approach to Outlier Detection and Residual Analysis." Biometrika, 75, 651-660.

Congdon, Peter. 2003. *Applied Bayesian Modelling*. London: Wiley & Sons.

Derks, Anton, and Kris Deschouwer. 1998. "Vrijzinigen, ongelovigen en protest". In: *Kiezen is verliezen. Onderzoek naar de politieke opvattingen van Vlamingen*, Marc Swyngedouw, Jaak Billiet, Ann Carton & Roeland Beerten (eds.), 85-112. Acco: Leuven.

Draper, David. 2001. *Bayesian hierarchical modeling*. New York: Springer.

Escobar, Marcelo, Ernesto Calvo, Natalia Calcagno, and Sandra Minvielle. 2002. "Ultimas Imágenes Antes del Naufragio: Las Elecciones del 2001 en Argentina". *Desarrollo Economico*, 42, 25-44.

Fornos, Carolina. 1996. *Explaining Voter Turnout in Latin America*. Master's Thesis, Louisiana State University.

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

Francese, Robert J. 2005. "Empirical Strategies for Various Manifestations of Multilevel Data". Political Analysis, 13, 430 – 446.

Franklin, Mark, and Wolfgang Hirczy de Mino. 1998. “Separated powers, divided government, and turnout in U.S. presidential elections.” *American Journal of Political* Science, 42, 316-326.

Gelfland, Alan E., Susan E. Hills, Amy Racine-Poon, and Adrian F. Smith. 1990. “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling”. *Journal of the American Statistical Association*, 85(412), 972 – 985.

Gelfland, Alan E., and Adrian F. Smith. 1990. “Sampling-Based Approaches to Calculating Marginal Densities”. *Journal of the American Statistical Association*, 85(410), 398 – 409.

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

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

Gelman, Andrew, and Iain Pardoe. 2007. “Average predictive comparisons for models with nonlinearity, interactions and variance components”. *Sociological Methodology*, 37(1), 23-51.

Gelman, Andrew and Donald B. Rubin. 1992. “Inference for iterative simulation using multiple sequences”. *Statistical Science*, 7, 457-472.

Geman, S., and D. Geman. 1984. “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

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 2006).

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

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

Jackman, Robert W. 1987. “Political institutions and voter turnout in the industrial democracies.” *American Political Science Review*, 81, 405 – 423.

Jackman, Simon. 2001. ‘Voting: Compulsory’. In *International Encyclopedia of the Social & Behavioral Sciences*, Neil Smelser and Paul Baltes, eds. Oxford: Elsevier Science.

Jackman, Simon. 2004. “Bayesian Analysis for Political Research”. *Annual Review of Political Science*, 7, 483-505.

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

Jocelyn-Holt, Alfredo. 1998. *El Chile Perplejo*. Santiago: Planeta/Ariel.

Johnson, N.L. and S. Kotz. 1972. *Distributions in Statistics: Continuous Multivariate distributions.* New York: Wiley.

Katz, Jonathan, and Gary King. 1999. “A Statistical Model for Multiparty Electoral Data”. *American Political Science Review*, 93, 15 – 32.

King, Gary. 1997. *A Solution to the Ecological Inference Problem.* Princeton: Princeton University Press.

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

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

Kostadinova, Tatiana. 2003. “Voter Turnout Dynamics in Post-Communist Europe.” *European Journal of Political Research*, 42, 741-759.

Langoni, Patricia 1997. “Real Money Holdings, Money Growth and Inflation: Brazil.” Working Paper, Central Bank of Chile.

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

Lindley, D.V., and A.F.M. Smith. 1972. "Bayes Estimates for the Linear Model." J*ournal of the Royal Statistical Society,* Series B, 34: 1-41.

Liu, Chuanhai. 1996. “Bayesian Robust Multivariate Linear Regression With Incomplete Data”. *Journal of the American Statistical Association*, 91(435), 1219-1227.

Lubbers, Marcel, and Peer Scheepers. 2000: “Individual and contextual characteristics of the vlaams blok vote”. *Acta Politica*, 35, 363 – 398.

Mainwaring, Scott. 1991. “Politicians, Parties and Electoral Systems: Brazil in Comparative Perspective.” *Comparative Politics*, 24, 21-43.

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

McAllister, Ian, and Toni Makkai. 1993. “Institutions, Society of Protest? Explaining Invalid Votes in Australian Elections.” *Electoral Studies*, 12, 23-40.

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

Newton, Michael A., and Adrian E. Raftery. 1994. “Approximate Bayesian Inference by the Weighted Likelihood Bootstrap”. *Journal of the Royal Statistical Society*, Series B, 56, 3-48.

Pinheiro, J.C., C. H. Liu and Y. N. Wu. 2001. “Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution”. Journal of Computational and Graphical Statistics, 10, 249 – 276.

Powell, Bingham G. 1986. “American voter turnout in comparative perspective”. *American Political Science Review*, 80, 17-43.

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

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

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

Robinson, William S. 1950. “Ecological Correlation and the Behavior of Individuals.” *American Sociological Review*, 15, 351-357.

Rosa, G. J., C.R. Padovani and D. Gianola. 2003. “Robust Linear Mixed Models with Normal/Independent Distributions and Bayesian MCMC Implementation”. *Biometrical Journal*, 45(5), 573-590.

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

Schwartzman, Simon. 1973. “Twenty Years of Representative Democracy in Brazil”. In: *Mathematical Approaches to Politics*, Hayward Alker, Karl Deutsch and Antoine Stoetzel (eds.), 137-162. New York: Elsevier.

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

Shor, Boris, Joseph Bafumi, Luke Keele and David Park. 2007. “A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data”. *Political Analysis*, 15, 165–181.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. “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.

Thum, Yeow M. 2003. “Measuring Progress Toward a Goal. Estimating Teacher Productivity Using a Multivariate Multilevel Model for Value-Added Analysis”. Sociological Methods & Research, 32(2), 153-2007.

Van der Brug, Wouter, and Meindert Fenemma. 2003. “Protest or mainstream? How the European anti-immigrant parties have developed into two separate groups by 1999”. European Journal of Political Research, 42, 55 – 76.

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.

West, Mike. 1984. “Outlier Models and Prior Distributions in Bayesian Linear Regression”. *Journal of the Royal Statistical Society*, Series B, 46, 431-439.

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

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