Does Ethnic Solidarity Facilitate Electoral Support for Nation-Building Policies?: Evidence from a Political Experiment*

YVES F. ATCHADE† AND LEONARD WANTCHEKON‡

September 18, 2007

Abstract

Would voters support or reject a co-ethnic candidate if she were to adopt a platform that appeals equally to all ethnic groups? We address this counterfactual question using experimental data collected in the context of the 2001 elections in Benin. A hierarchical probit model with a structural equation is used to analyze the data. We adopt a Bayesian approach together with Markov Chain Monte Carlo to handle the computations. We find that ethnic ties strengthen electoral support for national public goods platforms. The effect is stronger among those who are culturally less distant from most other voters; that is, those who speak several languages, watch TV regularly, and travel frequently. We argue that the positive effect of ethnic ties would have remained, even if the experiment has taken place in more urban and ethnically diverse districts. We conclude that ethnic solidarity can help secure electoral support for nation-building policies as long as such policies are adopted by political leaders.

Key words: *Causal inference, Intent-to-treat, Noncompliance, Randomized experiments*

JEL Classification: *C14, C21, C52*

*We are grateful to seminar and conference participants for comments and to Ben B. Hansen for helpful discussions and constructive comments on an earlier version of this paper. Special thanks to the ISPS at Yale University for financial support. We are solely responsible for any remaining errors.

†Assistant Professor of Statistics, University of Michigan. Email: yvesa@umich.edu

‡Professor of Politics and Economics, New York University. Email: leonard.wantchekon@nyu.edu

1 INTRODUCTION

There is a growing consensus among economists and political scientists that excessive and inefficient redistribution leading to underprovision of public goods is one of the prime causes of underdevelopment.[1] More recently, the literature has focused almost exclusively on ethnic divisions as the main determinant of low demand for national public goods. [2] The standard assumption in the literature is that, in the presence of ethnic divisions, voters have stronger taste for targeted redistribution and clientelism. [3] Summarizing this view developed in Easterly (2001), Besley and Ghatak [2003] wrote: "if externalities are limited to within ethnic groups, then the total demand for public goods that benefits all groups such as roads and education will be less. For instance, if ethnic groups are separated geographically, there will be little demand for interregional travel. If different ethnic groups speak different languages and have different cultures, they will be less willing to support investment in public education." p.7. Since ethnic voters have weak preferences for public goods, electoral incentives drive politicians to target specific groups and divert public resources to private patronage.

However, the taste explanations for the adverse ethnic diversity effects is not quite satisfactory because it provides no explanations on what causes ethnic taste differences and how they can be shaped by policies (Miguel [2005], p. 329). In line with the political socialization literature, he finds that "nation-building policies" can promote ethnic cooperation and public goods provision. The question then becomes: how can one credibly communicate nation-building policies in the presence of ethnic diversity? In Miguel's study, national-policies policies were advocated and implemented in Tanzania only because the country was led by a charismatic socialist leader, Julius Nyerere.[4] In this paper, we use data from a randomized political experiment to show that voters are more responsive to platforms based on nation building policies when these policies are advocated by one of their own. The effect is stronger among coethnics of the candidate who speak more than one language, travel more frequently and watch television regularly. Thus, paradoxically, ethnic solidarity can facilitate electoral support for nation-building policies and therefore further ethnic

[1] See Alesina and Rodrik (1994), Alesina, Baqir and Easterly (1994), Alesina, Danninger and Rostagno (1999).

[2] See Robinson and Verdier (2000) for arguments relating income inequality and low productivity to clientelist redistribution. See Easterly and Levine (2000) on ethnic divisions, and Milesi-Ferretti, Perotti and Rostagno (2002) on proportional electoral systems.

[3] See Alesina, Baqir and Easterly (1999).

[4] But suppose Nyerere were a regular ethnic leader running in national election who decided to appeal equally to all voters, regardless their ethnic affiliation. Would he be punished by his co-ethnics and rewarded by those outside his ethnic group? Or would it be the way around?

cooperation. This is particularly evident when cultural distance between voters is reduced through better communication and better access to information. In other words, there might be no opposing effect between group solidarity and nation-building.

Our estimates of the effect of ethnic solidarity on electoral support for nation-building policies are based on voting outcomes in selected districts that were randomly assigned to "purified" national public goods and redistributive platforms by candidates competing in the 2001 presidential elections in Benin. The effect of National public goods platforms or redistributive platforms are measured by the difference in voting behavior between those who were exposed to the "purified" public goods or "purified" redistributive platforms and those who were exposed to "regular" platforms. Results from previous study using experimental data collected a week after the election, (Wantchekon, 2003), indicated a positive effect of redistributive treatments in all regions and for all types of candidates. They also indicated that national public goods messages had a negative effects in Northern districts but positive and significant effect in Southern districts. Finally, female voters were found to have stronger preference for national public goods platforms than male voters.

The current paper uses a much richer data set collected ten (10) months after the elections and concentrates on the way in which ethnic ties between voters and candidates affect voters' responsiveness to nation-building or national public goods and redistributive platforms. In contrast with previous studies that focus on the effect of ethnic diversity on public goods provision, we use a micropolitical approach by (1) focusing on shared ethnicity between an individual voter and a given candidate and (2) introducing an individualized concept of ethnic diversity (i.e. perceived ethno-linguistic fractionalization). The perceived ethno-linguistic fractionalization variable captures the ease of communication and cultural distance between an individual voter and other voters in the country and depends on factors such as number of language spoken, access to the media, and travel frequency.

More specifically, the paper addresses the following counterfactual questions: Would a given voter punish a candidate from his or her own ethnic group if that candidate were to adopt a national public goods platform that appeals equally to voters from all ethnic groups? Alternatively, would a voter from a given region punish a candidate not from his or her ethnic group if that candidate were to adopt a redistributive platform, stressing the needs of that region? We find that ethnic ties tend to strengthen voters support for national public goods platforms, with the effect being particularly strong among those who are more cosmopolitan and are better connected to the nation because of their language skills and access to information. Thus, quite surprisingly, the negative effect of public goods platforms on voting behavior is driven by voters who are not from the ethnic group of the candidate. There is however no significant difference across ethnic groups in their response to redistributive platforms.

From a statistical viewpoint, we frame the experiment as a *hierarchical* (or multilevel) *probit model* where individual voters at the bottom of the hierarchy are nested in villages which in turn are included in political strata. This type of model allows us to separate any village/political stratum effect from voter specific effects. Gelman et al. [2004] provides a detailed introduction to Bayesian hierarchical models. The *perceived ethno-linguistic fractionalization* variable discussed above which captures the ease of communication and ethno-linguistic distance between an individual voter and other voters in the country is more of a construct and therefore cannot be observed. We introduce this variable in the model as a voter-specific *random effect* linked to a number of 'proxies' covariables through a *structural equation.* There is a well-developed literature on structural equation models in social sciences (Bollen [1989], Rabe-Hesketh et al. [2004]).

Another feature of this paper is that we adopt a Bayesian approach for statistical inference. The main reason for this choice is the great simplicity and flexibility of the Bayesian approach in handling latent variables.

RELATION TO THE LITERATURE

Our study contributes to the current debate on ethnicity and public goods provision. Easterly and Levine [1997] and Alesina, Baqir and Easterly [1999] present evidence indicating that ethnic divisions increase the demand for redistribution and adversely affect levels of public of goods in Africa and in several US cities.[5] Bates [1983] argues that ethnic ties and spatial concentration of ethnic groups make it easier for local citizens to lobby for local public goods or projects of regional interest. For Fearon [1999], a greater level of interaction may increase trust among co-ethnics and facilitate coalition building along ethnic lines, which make lobbying more effective. The politics of exclusion that goes together with the formation of ethnic coalitions also leads to a higher demand for "pork" or projects of local interest. The evidence presented here indicates that ethnic solidarity can increase the demand for both "pork"and public goods.

The paper also contributes to the literature on conditions for ethnic cooperation. For Fearon and Laitin (1996) cooperation originates from the combined effect of agreement between group leaders on norms of inter-ethnic cooperation and the ability of those leaders to coordinate group response to those norms through within-group policing. Varshney (2002) stresses the role of elite

[5] In a related paper, Erzo Luttmer (2001) shows that the support of a given individual for welfare spending decreases as the number of welfare recipients in his or her community increase. However, the support increases as the number of recipients from his or her own racial group increases. Miguel and Gugerty (2002) also find a negative correlation between ethnic diversity and public school funding in Kenya. They attribute the result to the fact that collection action is hard to sustain in heterogeneous communities.

coordination through interethnic associational bonds. However, as Miguel (2004) pointed out, the problem with these papers is that it is unclear whether elite cooperation is a cause or a consequence of cooperative ethnic relations. In this study, elite cooperation is secured exogenously through their collective adherence to the experimental protocol. Therefore, group support for nation-building policies can be claimed to be causally related to ethnic ties and elite cooperation.

The concept of perceived ethno-linguistic fractionalization bears some similarities with the growing literature on the role of a common language on trade (Melitz [2002] and Helliwell [1999] among others). Melitz finds that a common language facilitates international trade mostly as a result of the number of individuals who can communicate person to person. In this paper, it is the ability of an individual to speak several languages that facilitates his personal connection with others and therefore his responsiveness to nation-building policies.

The paper is organized as follows. Section II presents the experimental design and the data. We then present the statistical analysis in section III and provide concluding remarks in section IV.

2 EXPERIMENTAL DESIGN

The Benin experiment is a randomized evaluation of electoral platforms. The experimental protocol consists of the following three steps (see Wantchekon [2003] for more details as well as background information on Benin)

The first step consists of selecting and contacting political parties that will be involved in the experiment. In order to limit threats to external validity, that is to facilitate the generalizability of the results to the entire country, parties were chosen along three characteristics (incumbent/opposition, national/local and from various regions and with various political characteristics. There are six major parties in the country. They were separated into two groups, the Northern parties and the Southern parties. There are two government parties among the Northern parties. Among them, one is a local or regional party. There are two opposition parties among the southern parties with one being local. We eliminate the local southern candidate from the opposition and the local northern candidate from the incumbent coalition. We were then left with four parties: one local opposition from the North (UDS), one local incumbent party from the South (PSD), one national incumbent party from the South (RB) and one national incumbent party from the North (FARD).

In the presence of representatives of each candidate, eight districts were selected, two per candidates. Then all the villages in each of them were listed and two of them were randomly picked. The two selected villages were assigned to the treatment groups and the rest of the district served as the control group. Among the two villages picked, one was assigned to the distributive policy treatment and the other to public goods treatment. The control villages were exposed to the regular campaign, which is a combination of public goods and distributive policy messages. Furthermore, in order to limit threats to internal validity, and to avoid a mix up of the two types of treatment, we made sure that parties only selected villages that were at least 25 miles apart with 4 to 10 other villages separating them. The aggregate sample of the population under treatment is 6,633 registered voters for distributive policy treatment group, 6,983 voters for "public goods" villages, and about 220,000 for the control group.

More formally, denote by N the number of political units involved in the experiment. We divide N into $S = 4$ strata on the basis of political characteristics (incumbent-dominated or opposition dominated, northern or southern, and "national" or "local"). There are N_s political units in stratum $s \in \{1, 2, 3, 4\}$ so that $N = \sum N_s$. In fact, N_s is the number of political units (electoral districts) controlled by a given candidate, s. Within each political unit (electoral district), there are several subunits (villages). The randomization process consists of the following four steps:

Step 1. Complete randomization among districts, i.e. given the number of districts N_s, candidate s draws randomly 2 districts (say j and k) out of the population to be part of the experiment.

Step 2. Given the number n_j and n_k of villages or subunits in district j and k, candidate s randomly draws one village among the n_j and one among the n_k.

Step 3. Eliminate some villages (say 5 to 10) among the $n_j - 1$ remaining villages in district j and the $n_k - 1$ remaining villages in district k (the villages that are eliminated are contiguous or in the immediate vicinity of the village picked in stage 2). Then draw randomly one village from each population.

Step 4. Randomly assign one of the two villages chosen in step 2 and step 3 to redistributive treatment, and the other village to public goods treatment. The remaining $n_j - 1$ villages in district j and $n_k - 1$ in district k are assigned to control groups.

Thus, the experiment is a *randomized block experiment* with treatments being assigned to subunits (villages) as within some randomly chosen units (electoral districts).

2.1 TREATMENTS

After the selection of the villages was completed, the two types of messages were designed with the active collaboration of the campaign managers of the parties and based on the platforms that the parties have adopted. A public goods message raised issues pertaining to poverty alleviation; public health and education reform; agricultural and industrial development. A distributive policy message, in contrast, took the form of a specific promise to the village. It took the form of promised government patronage jobs or local public goods such as establishing a new local university, financial support for local fishermen or cotton producers. Thus, by and large, the public goods message and the distributive policy message stressed the same issues. However, the former stressed the issue as part of a national program, while the latter stressed the issue as part of a specific project to transfer government resources to the region or the village. In order to facilitate a clear distinction between the two types of messages and enhance the *internal validity* of the experiment, a public goods message never promised patronage jobs and a redistributive policy message never promised nationwide education reform. In addition, while campaign workers stressed the need for ethnic cooperation and harmony when they deliver the public goods messages, they outline (whenever possible) the ethnic ties of the candidate with the local voters when they deliver the redistributive message.[6]

It is worth stressing the fact that a typical platform is a mixture of redistributive and public goods messages. For the purpose of the experiment, the parties offered to "purify" their platforms in the treatment districts into ones which were purely redistributive or purely public goods. In other words, just like in any regular political campaign, the parties involved in the experiment are running on their own platforms. The only difference here is that they slightly adapted the campaigns that they intended to run in some villages to fit the objectives of the experiment. Thus, there is no real risk of Hawthrone and John Henry effects because treatments were imbedded in regular political campaigns.[7] See Wantchekon (2003) for more details on the implementation of the treatments

After the elections, a survey was conducted in all treatment districts. In each district, a representative sample of voters were interviewed in the two treatment villages and from the control villages. The survey collected basic demographic data (age, gender, marital status, number of people in the households and ethnic affiliation), socioeconomic data (educational attainment, economic activities and assets) and data on respondents' social networks and use of media outlets (radio, television and newspapers). The information on social networks includes membership in organizations (cooperatives, NGOs, parties and unions), travel and languages spoken, and participation in political discussions. The survey also collected data on voting behaviors in the 2001, 1996 and 1991 presidential elections.

[6]The experiment would have been more informative if the platforms were focused on one or two policies, say education, and patronage jobs. This was not possible this time because the platform has to reflect the actual electoral strategies of the candidates.

[7]Hawthrone and John Henry occur when the difference between control and treatment groups is essentially due to the fact that the subjects are aware that they are being observed.

3 STATISTICAL ANALYSIS

3.1 Difference-in-means analysis

A straightforward analysis of the experiment can be carried out on a village level. Let $\bar{y}_i(l)$ be the proportion of votes the candidate campaigning in village i would have received if his campaign uses message l ($l = 0, 1, 2$). The message 0 ($t = 0$) is the standard message, message 1 ($t = 1$) is the redistributive platform and message 2 ($t = 2$) is the public good platform. We define the village-level causal effect of message l in village i as $\bar{y}_i(0) - \bar{y}_i(l)$. The experiment has been designed as a block randomized experiment where villages are randomly assigned to each treatment. Therefore unbiased estimates of the (village-level) causal effects of treatments 1 and 2 are readily available by simple difference-in-means for each district. Table 2 reports these estimates. We also give in Table 1 the estimated proportion of votes received by each candidate in its districts. We have reasons to believe that the experiment protocols have not been followed in Districts 3 and 8 (Stratum 4), which as the result, have been discarded from the analysis. An overall village-level effect of the treatments is computed as a simple average of the effects in the seven districts.

	Dist. 1	Dist. 2	Dist. 4	Dist. 5	Dist. 6	Dist. 7
$t = 0$	76.62 (77)	100.00 (82)	99.12 (114)	93.14 (102)	98.86(88)	66.34 (104)
$t = 1$	78.87 (71)	100.00 (94)	72.79 (136)	99.11 (112)	100.00 (79)	82.46 (57)
$t = 2$	73.68 (76)	95.45 (88)	89.20 (74)	66.67 (99)	100.00 (56)	84.80 (125)

Table 1: Proportion (sample size in parenthesis) of votes received by the leading candidates in their districts.

These results show that the public good platform ($t = 2$) has a generally negative effect on the outcome and that the redistributive platform ($t = 1$) has an overall positive effect on the outcome. A more detailed analysis of these causal effects can be found in Wantchekon [2003]. Our main objective here is to understand the effects of the treatments in presence of ethnic ties. In order to do so, we will move away from a village-level analysis to a voter-level analysis and we will introduce an explicit regression model that relates the voter's outcome to its characteristics.

There is another potential advantage to an individual-level analysis. It offers a more reliable approach to estimate the causal effects of the treatments in more competitive villages. Indeed, for reasons explained above, only noncompetitive districts have been included in the experiment. A natural question to ask is how these effects transfer night have been if the experiment took place in more competitive villages and cities. We give an estimate of such effects by considering individual-level causal effect of the treatments conditioned on individuals having "city-like" characteristics.

	Dist. 1	Dist. 2	Dist. 4	Dist. 5	Dist. 6	Dist. 7	Overall
$t = 1$	-2.25	0.00	26.33	-5.97	-1.14	-16.11	0.14
$t = 2$	2.94	4.55	9.93	26.47	-1.14	-18.46	4.04

Table 2: Village-level causal effects of the treatments (in percentage) estimated by difference-in-means.

3.2 A hierarchical Probit model with a latent class variable

We analyze the data using a hierarchical probit model where the perceived ethno-linguistic fractionalization variable is introduced as a latent variable with a sub-level structural equation. The model belongs to the class of hierarchical structural equation models (Rabe-Hesketh et al. [2004]). We adopt a Bayesian approach for model estimation using Markov Chain Monte Carlo (MCMC). More detail on the MCMC is given in the appendix.

3.2.1 Defining the causal effects

We have divided the country in $D = 4$ political strata (see the experimental design above for more details). Let V_i be the number of villages in stratum i and N_{ij} the number of people in village j of stratum i. We adopt a counterfactual approach to causal inference in a Bayesian framework. We assume the existence of random variables $y_{ijk}(0), y_{ijk}(1), y_{ijk}(2)$, where $y_{ijk}(l)$ is the voting outcome we would observe on individual k in village j in stratum i if he/she **and everybody** in his/her village were exposed to treatment l. $y_{ijk}(l) = 1$ if unit i, j, k votes for the political party campaigning in his/her village (and 0 otherwise). We make the assumption that for (i, l) fixed, all the variables $y_{ijk}(l)$ have the same distribution. This point is discussed further in the remarks below. Clearly, only one of those outcomes is observed for each individual in the experiment. The causal effect of treatment l (with respect to treatment 0) is defined as:

$$\tau^{(l)} = \frac{1}{\sum_{i=1}^{D}\sum_{j=1}^{V_i} N_{ij}} \sum_{i=1}^{D}\sum_{j=1}^{V_i}\sum_{k=1}^{N_{ij}} \left[y_{ijk}(0) - y_{ijk}(l)\right]. \tag{1}$$

Our objective is to make inference about the distribution of the random variables $\tau^{(l)}$, $l = 1, 2$. The counterfactual variables $y_{ijk}(l)$ are not observed for two reasons. First, these are counterfactual variables so that for any given voter, only one of the three outcome $y_{ijk}(l)$, $l = 0, 1, 2$

can be observed. Also, since a follow up randomized survey is performed to collect the data, even if voter i, j, k's village has received treatment l, we will observe $y_{ijk}(l)$ only if that voter is selected in the follow up survey.

Therefore, and for notational clarity, we introduce the variables $\tilde{y}_{ijk}(l)$ to denotes the voting outcome (under treatment l) of the k-th selected individual (in the follow-up survey) from the villages selected to receive treatment j in the ith stratum. When $l = j$, $\tilde{y}_{ijk}(l)$ is observed; and is missing otherwise. For $i = 1, \ldots, D$, let n_{il} be the number of selected units (in the follow-up survey) from the villages that received treatment l in the i-th stratum. Since the experiment and the follow up survey are randomized and also because of the fact that the distribution of $y_{ijk}(l)$ does not depend on (j, k) for (i, l) fixed, it follows that the distribution of the causal effect $\tau^{(l)}$ is the same as the distribution of $\hat{\tau}^{(l)}$ given by:

$$\hat{\tau}^{(l)} = \frac{1}{\sum_{i=1}^{D}(n_{i0} + n_{i1} + n_{i2})} \sum_{i=1}^{D} \sum_{m=0}^{2} \sum_{k=1}^{n_{im}} \left[\tilde{y}_{imk}(0) - \tilde{y}_{imk}(l)\right]. \tag{2}$$

Next, we postulate a statistical model that relates the outcome $\tilde{y}_{ijk}$ to the covariates. Using this model we will be able to sample from the distribution of $\hat{\tau}^{(l)}$ and from the conditional distributions of $\hat{\tau}^{(l)}$ given various covariates of interest.

3.2.2 A statistical model relating voting outcome and covariates

Let x_{ijk} be the p-dimensional vector of covariates of unit k of village j of stratum i. We present these covariates below. As discussed above, the ethno-linguistic fractionalization of the country as perceived by a given voter is an important determinant in her response to the different treatments. We introduce a variable s to denote the individual perception of the ethno-linguistic fractionalization of the country. This variable is really a multidimensional concept that is difficult to observe. As a solution, we introduce s in the model as a latent variable (or a random effect). We elaborate more on the variable s below. We postulate the following hierarchical probit model to relate the outcomes to the covariates and the random effect s:

$$\tilde{y}_{ijk}(l)|\beta_{il}, \alpha_{il}, x_{ijk}, s_{ijk} \overset{i.i.d.}{\sim} \mathcal{B}in\left(\Phi\left(x'_{ijk}\beta_{il} + s_{ijk}\alpha_{il}\right)\right),\ k = 1, \ldots, n_{ij},\ j = 0, 1, 2,\ i = 1, \ldots, D \tag{3}$$

for some random variables (parameters) (β_{il}), (α_{il}). $\mathcal{B}in(p)$ denotes the Bernoulli distribution with parameter p and Φ is the normal standard cdf.

Remark 3.1. 1. Through the coefficient β_{il} and α_{il}, the model allows for interactions between treatments and strata. This can be interpreted in at least two ways. This interaction can account for the existence of different versions of each treatment. For example, message 0, the standard message is not a purified message and very likely will vary from one political group to another. Even the purified messages could have been delivered slightly differently from one stratum to the other. The interaction term can also account for any common characteristic of voters of a given stratum not shared by voters from other strata. But the model also assumes that village-level differences are negligible within each stratum. This is plausible as only one political group is allowed to "treat" any given stratum and all the villages inside.

2. The model also assumes the conditional independence of the voting outcomes inside each village. This latter assumption is more debatable as two related individual might obviously influence each other on how to make sense of the message they collectively received. Our conclusions might be biased to the extent that the data depart from these assumptions.

3. There is one caveat in using the covariates as explanatory variables: these covariates have been measured in a follow-up survey, after the units have been treated. It is then possible that these covariates have been influenced somehow by the treatments received. This would invalidate our analysis. But we believe this possibility to be very unlikely. Indeed, we were only interested in very basic characteristics of the population (age, sex, ethnic ties with the candidate, whether they watch TV etc...) that we believe are not easily influenced by the received treatments.

3.2.3 The perceived ethno-linguistic fractionalization variable

Clearly, we cannot observe the individually perceived ethno-linguistic fractionalization variable s. But we can infer s from a number of related and observed covariates. For instance, how many times a person travel, watch TV, how many languages he/she speaks are good proxies for its perceived ethno-linguistic fractionalization of the country. Therefore, denoting these covariates by z, we assume the following structural model for s:

$$s_{ijk}|z_{ijk}, b \overset{i.i.d}{\sim} \mathcal{B}in\left(\Phi(z'_{ijk}b + \bar{v})\right), \tag{4}$$

for some random variable (parameter) b and a (known) constant $\bar{v} > 0$. This is a Probit model with an offset $\bar{v}$. The model has an alternative representation which is perhaps more intuitive:

$$\begin{aligned} v_{ijk}|z_{ijk}, b &\overset{i.i.d}{\sim} N(z'_{ijk}b, 1), \\ s_{ijk} &= 1 \text{ if } v_{ijk} > -\bar{v} \text{ and } 0 \text{ otherwise, where } \bar{v} = 2.00. \end{aligned} \tag{5}$$

The reason for this model choice is that we find important that s be a dichotomous variable that will separate those who see the country as completely fractionalized ($s = 1$) and those who do not. The choice of the offset $\bar{v}$ is somewhat arbitrary. But the argument behind this choice is that, for an individual for whom all the components of the covariate z are 0, we choose $\bar{v}$ such that with high probability, such individual is classified as $s = 1$. We encode the covariates z appropriately so that the value 0 is consistent with a vision of complete fractionalization of the country. We choose $\bar{v} = 2.0$. We try different values of $\bar{v}$ with similar results.

3.2.4 Prior distribution, Posterior distribution and causal effect estimates

For β_{il} and α_{il}, we assume the following prior.

$$\begin{aligned}
\beta_{il}|\beta_l, \sigma_l^2 &\overset{i.i.d}{\sim} N(\beta_l, \sigma_l^2 I_p), \;\; i = 1, \ldots, D, \; l = 1, 2, 3, && (6)\\
\beta_l &\overset{i.i.d}{\sim} N(0, \tau^2 I_p), \;\; l = 1, 2, 3 && (7)\\
\alpha_{il}|\alpha_l, \sigma_l^2 &\overset{i.i.d}{\sim} N(\alpha_l, \sigma_l^2), \;\; i = 1, \ldots, D, \; l = 1, 2, 3, && (8)\\
\alpha_l &\overset{i.i.d}{\sim} N(0, \tau^2), \;\; l = 1, 2, 3 && (9)\\
\sigma_l^2 &\overset{i.i.d}{\sim} IG(\nu, \delta), \;\; l = 1, 2, 3 && (10)\\
\tau = 100 & \qquad\quad \nu = 1.0, \; \delta = 2.01, && (11)
\end{aligned}$$

where $IG(\nu, \delta)$ is the inverse-gamma distribution with shape ν and scale δ. I_p denotes the p-dimensional identity matrix. We can think of β_l as the mean value of the coefficients β_{il} under treatment l.

For b, we assume that

$$b \sim N(0, \tau^2 I_q), \quad \tau = 100. \tag{12}$$

Write $y_{ilk} = \tilde{y}_{ilk}(l)$ the outcome of the k-th selected individual (in the follow-up survey) from the villages that received treatment l in the i-th stratum. We write $y_{obs} = (y_{ilk})$ for the observed data, v the vector of latent variables v_{ilk} of the selected units as in (5), and denote $\theta = ((\beta_{il})_{il}, (\gamma_l)_l, (\sigma_l^2)_l, b)$ the vector of parameters of the model. Since the experiment is a randomized block experiment, the assignment to the treatment is ignorable and the likelihood of y_{obs} is available as:

$$f(y_{obs}|\theta, v, x) = \prod_{i=1}^{D} \prod_{l=0}^{2} \prod_{k=1}^{n_{il}} p_{ilk}^{y_{ilk}} (1 - p_{ilk})^{1 - y_{ilk}}, \tag{13}$$

where $p_{ilk} = \Phi\left(x'_{ilk}\beta_{il} + s_{ilk}\alpha_{il}\right)$, and $s_{ilk} = 1$ if $v_{ilk} > -\bar{v}$ and 0 otherwise.

Therefore the posterior distribution of (θ, v) given y_{obs}, x and z becomes:

$$\begin{aligned}\pi(\theta, v|y_{obs}, x, z) \quad \propto \quad & f(y_{obs}|\theta, x) \prod_{l=0}^{2} \left(\frac{1}{\sigma_l^2}\right)^{\nu+1} e^{-\delta/\sigma_l^2} e^{-\frac{1}{\tau^2}(\beta_l'\beta_l+\alpha_l^2)} \\ & \times \prod_{i=1}^{D} \frac{1}{\sigma_l^p} e^{-\frac{1}{2\sigma_l^2}\left[(\beta_{il}-\beta_l)'(\beta_{il}-\beta_l)+(\alpha_{il}-\alpha_l)^2\right]} \\ & \times e^{\frac{1}{2\tau^2}b'b} \prod_{i,l,k} e^{-\frac{1}{2}(v_{ilk}-z_{ilk}'b)^2}. \end{aligned} \tag{14}$$

We sample from this posterior distribution using Markov Chain Monte Carlo (MCMC). The MCMC algorithm that we use is an extension of a data-augmentation scheme for probit models developed by Albert & Chib (1993). The details are given in the Appendix.

We sample from the predictive distribution of (2) given θ, v, y_{obs}, x, z by imputation of the missing variables as follows. We first sample θ, v from (14), the posterior distribution of θ, v given y_{obs}, x, z. Then, for i, m, k, l if $l = m$ we set $\tilde{y}_{imk}(l) = y_{ilk}$ otherwise we sample $\tilde{y}_{imk}(l)$ from $\mathcal{B}in\left(\Phi\left(x_{imk}'\beta_{il} + s_{imk}\alpha_{il}\right)\right)$, where $s_{imk} = 1$ if $v_{imk} > -\bar{v}$ and 0 otherwise.

3.2.5 More details on the covariates

The components of the vector x includes the covariates *male* (1 if male, 0 otherwise), *schooling* (1 if ever went to school, 0 otherwise), *age* and *candethn* (1 if from the same ethnic group as the candidate, 0 otherwise).

The variable z which comprises the proxies for the *perceived ethno-linguistic fractionalization* has the following components: *traveled* (1 if travels very frequently, 0 otherwise), *lang* (1 if speaks more than 1 language, 0 otherwise), *tele* (1 if watch tv regularly, 0 otherwise) and *childout* (1 if has a child leaving outside the village, 0 otherwise).

3.2.6 Model validation

To validate the model, we compute the posterior predictive p-values based on the statistics $V_i = \frac{\sum_{\{k:\, stratum_k = i\}} y_k}{\sum_{\{k:\, stratum_k = i\}} 1}$. V_i is the proportion of votes received by the i-th leading candidate in his political stratum. To obtain these p-values, we sample from the posterior predictive distribution of each V_i and compute (empirically) the probability that V_i takes values equal or larger than the value of the statistics computed on the dataset. If the model fits the data well, we expected these p-values to be closer to 0.5 than to the extremes values 0 and 1. The results are given in Table 4. We conclude from this table that our model fits fairly well the data set.

Stratum 1	Stratum 2	Stratum 3
0.479	0.472	0.548

Table 3: Posterior predictive p-values for the proportion of votes received by the leading candidate in his stratum.

3.2.7 Overall effects

In Table 7 we show the posterior predictive means and a 95% posterior interval for the overall causal effects of the two treatments. The 95% posterior interval is formed by the 2.5% empirical quantile and the 97.5% empirical quantile of the posterior distribution. It follows from these results that, on average at an individual level, the redistributive platform is closer to the standard treatment than the public goods platform. The posterior predictive distributions of the two causal effects are given in Figure 1.

In Table 4, 5 and 6, we report the posterior mean, Monte Carlo confidence interval and posterior interval for the component of the coefficient β_0, β_1 and β_2. The Monte Carlo confidence interval is useful to assess the precision of the Monte Carlo simulation. It is a frequentist confidence interval based on the central limit theorem for the output of the simulation. It is important to understand that this Monte Carlo confidence interval does not assess the precision of the estimation as inferred from the model but the precision of the Monte Carlo method in estimating a characteristic of the posterior distribution.

The coefficients $\beta_0, \beta_1, \beta_2$ can be thought as the average (over i, the stratum) of the β_{il}, the coefficients of the components of the covariates x. Although most of these coefficients are not significantly different from zero, the following observations stand out. The sign of the coefficient of *male* indicates that, everything being equal, women adhere more easily to the public good platform than men and the sign of the coefficient of *schooling* indicates that those who have ever been to school are more likely to adhere to the public good platform. In the following section, we discuss the link between ethnicity, ethnic fractionalization and demand for public good as inferred from the model.

Variables	Post. mean	Monte C.I.	Post. interval
Constant	1.79	$(1.78, 1.82)$	$(0.39, 3.39)$
Male	-0.12	$(-0.14, -0.11)$	$(-1.32, 1.08)$
Schooling	-0.49	$(-0.51, -0.49)$	$(-1.69, 0.68)$
Ethnic ties	0.04	$(0.02, 0.06)$	$(-1.51, 1.45)$
Latent variable s	-0.09	$(-0.10, -0.07)$	$(-1.36, 1.19)$

Table 4: Posterior mean 95% Monte Carlo confidence interval and 95% posterior interval for components of β_0.

Variables	Post. mean	Monte Carlo C.I.	Post. interval
Constant	1.32	$(1.29, 1.37)$	$(-0.42, 3.31)$
Male	0.07	$(0.03, 0.10)$	$(-1.42, 1.61)$
Schooling	-0.07	$(-0.10, -0.04)$	$(-1.56, 1.49)$
Ethnic ties	0.61	$(0.58, 0.66)$	$(-1.23, 2.49)$
Latent variable s	0.19	$(0.17, 0.21)$	$(-1.32, 1.73)$

Table 5: Posterior mean 95% Monte Carlo confidence interval and 95% posterior interval for components of β_1.

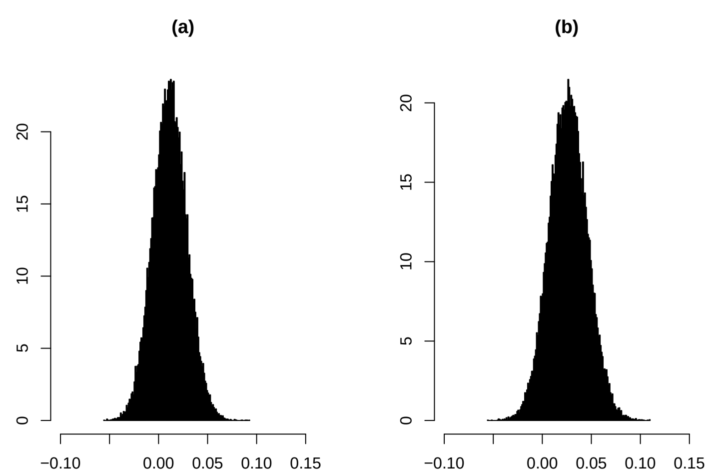

Figure 1: Posterior predictive distributions of the causal effects of the treatments. (a) Redistributive platform. (b) Public goods platform.

Variables	Post. mean	Monte Carlo CI	Post. interval
Constant	0.68	$(0.67, 0.69)$	$(-0.53, 1.89)$
Male	-0.39	$(-0.39, -0.38)$	$(-1.44, 0.68)$
Schooling	0.42	$(0.41, 0.43)$	$(-0.70, 1.64)$
Ethnic ties	0.69	$(0.68, 0.70)$	$(-0.54, 1.90)$
Latent variable s	-0.93	$(-0.95, -0.92)$	$(-2.14, 0.27)$

Table 6: Posterior mean 95% Monte Carlo confidence interval and 95% posterior interval for components of β_2.

Treatment	Post. pred. mean	Post. pred. interval
Message 1	1.11	$(-2.22, 4.62)$
Message 2	2.65	$(-1.09, 6.52)$

Table 7: Posterior predictive mean and posterior predictive interval of the causal effects. In percentage.

3.2.8 Ethnic ties

From the coefficients of the model presented above, it appears that ethnic ties with the candidate seem to strengthen adherence to all types of messages, particularly for the purified messages. Next, we look at the posterior predictive distribution of the causal effects of the treatments conditioned on the existence or not of ethnic ties with the candidate. Table 8 reports the estimated conditional causal effects together with the 95% posterior interval. The results indicate that the candidates loose very few support among voters from its ethnic group, from one message to another. In that case, and in accordance with the overall results, the public good platform has the least support. The interesting finding here is that voters from different ethnic groups as the candidate adhere more easily to the redistributive platform but are more likely to penalize the public good platform (in comparison to the standard or control platform).

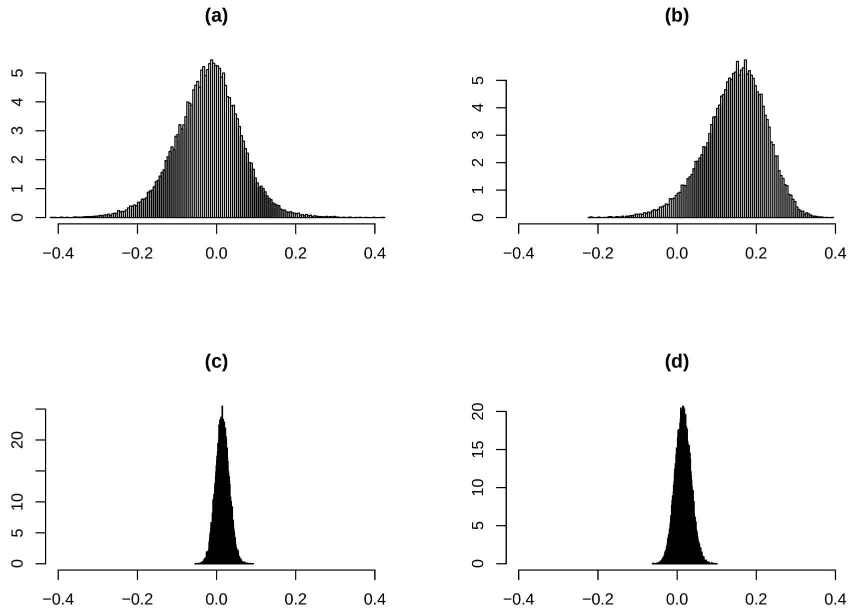

Figure 2: Causal effect distributions conditional on having and not having ethnic ties with the candidate. (a) No ties, Redistributive message. (b) No ties, Public goods message. (c) Ethnic ties, Redistributive message. (d) Ethnic ties, Public goods message.

Treatment & Ethnic ties	Post. pred. mean	Post. interval
Message 1 | ethnic ties	1.47	$(-1.79, 4.87)$
Message 1 | no ethnic ties	-2.17	$(-19.5, 13.8)$
Message 2 | ethnic ties	1.50	$(-2.33, 5.46)$
Message 2 | no ethnic ties	14.7	$(-2.05, 27.92)$

Table 8: Posterior predictive mean and 95% interval for the causal effects with and without ethnic ties. In percentage

3.2.9 Perceived ethnic fractionalization

Table 9 shows the posterior mean of the components of b, the coefficient of z in the model. The covariate z has the following four components *traveled* (1 if travels very frequently, 0 otherwise), *lang* (1 if speaks more than 1 language, 0 otherwise), *tele* (1 if watch tv regularly, 0 otherwise) and *childout* (1 if has a child leaving outside the village, 0 otherwise). These estimates show that when all the components of the covariates z are equal to 1, s is very likely to be 0. As a consequence, and with a certain confidence, we can see this latent variable s as representing the ethno-linguistic fractionalization as perceived by each voter. $s = 1$ means that the voter sees the country as totally fractionalized and $s = 0$ means that the voter does not see the country as not totally fractionalized. The intuition is that how each individual sees the society in that respect is a fundamental variable that drives its response to each treatment. This is shown by looking at the coefficient of s in the model (Table 4, 5 6). These estimates imply that a voter is more likely to support the public good platform when $s = 0$ and the redistributive platform when $s = 1$. More specifically, we compute the posterior causal effect of the two treatment conditioned on $s = 0$ and $s = 1$. These distributions are summarized in Table 10. This seems to indicate that individuals who see the country as completely fractionalized strongly reject the public good platform (in comparison to the standard or control platform). In contrast, the public good platform has more support among voters that do not see the country as totally fractionalized.

Since a voter with $s = 0$ is by definition more cosmopolitan, we claim that we are much more likely to see this type of voter in cities and more competitive electoral districts. In that sense, the conditional causal effects given $s = 0$ indicates that nation-building or national public good platforms would have generated more electoral support in more competitive and in urban districts. In other words, the national public good treatment effect would have been less negative if the experiment took place mostly in cities.

Covariates	post. mean	Monte Carlo CI	post. interval
Traveled	-5.98	$(-6.45, -5.51)$	$(-10.81, -4.17)$
Language	-5.60	$(-5.87, -5.34)$	$(-9.17, -3.74)$
TV	-6.22	$(-6.64, -5.81)$	$(-9.96, -4.48)$
Child outside	-5.92	$(-6.22, -5.63)$	$(-8.98, -4.24)$

Table 9: Posterior mean, 95% Monte Carlo confidence interval and 95% posterior interval for the components of b.

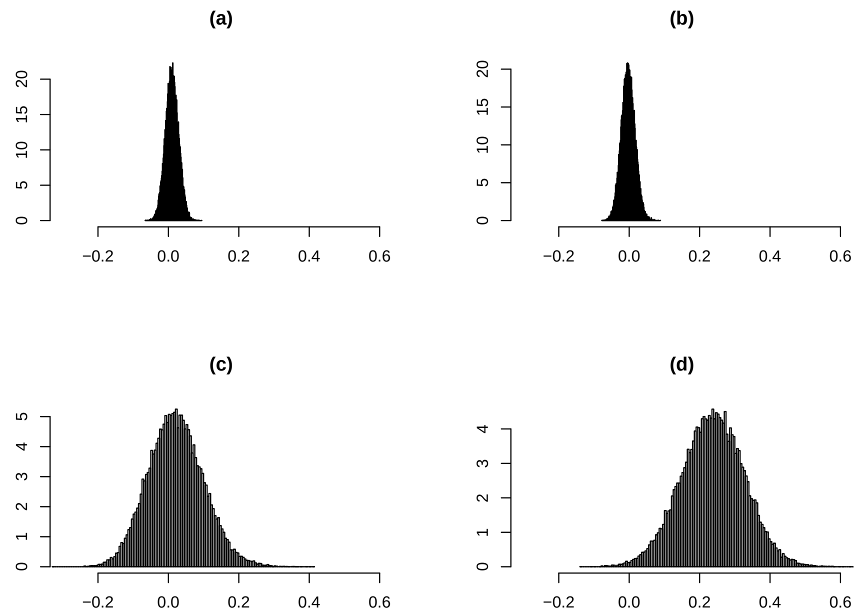

Figure 3: Causal effect distributions conditional on perceived ethno-linguistic fractionalization. (a) $s = 0$, Redistributive message. (b) $s = 0$, Public goods message. (c) $s = 1$, Redistributive message. (d) $s = 1$, Public goods message.

Treatment & Eth.-ling. fract.	Post. pred. mean	Pst. interval
Message 1 | fract.	2.12	$(-12.69, 18.19)$
Message 1 | nonfract.	1.01	$(-2.69, 4.76)$
Message 2 | fract.	23.71	$(5.94, 41.22)$
Message 2 | nonfract.	0.28	$(-4.10, 3.69)$

Table 10: Posterior predictive mean and 95% posterior interval of the causal effects given perceived ethnic fractionalization. In percentage.

4 CONCLUDING REMARKS

In this paper, we provide experimental estimates of the effect of ethnic solidarity on electoral support from nation-building policies. We find that voters from the same ethnic group of the candidate tend to be more supportive of nation-building platforms from that candidate than voters who are not. In addition, voters more connected to the country through their access to the media, family relations and language skills tend to respond more positively to those policies. We believe that these results make important contribution to the current debate on ethnic conflict and ethnic cooperation. However, further investigation is needed to explain why would ethnic ties make support for broad public goods platforms more likely. This could be because ethnic ties contribute to the establishment of trust between candidates and voters or because ethnic voters know "their" candidate will in the end deliver the kind of public good that is most beneficial to their constituency. In future works, we intend to investigate more closely whether the results are mainly driven by trust, credibility or by self-interest.

5 Appendix

5.1 MCMC Sampling

Markov chain Monte Carlo (MCMC) is a popular computational method for generating samples from virtually any distribution π defined on a space $\mathcal{X}$. As in Bayesian computations, these samples are often used to efficiently compute expectations with respect to π by invoking some form of the law of large numbers. The method consists of simulating an ergodic Markov chain $\{X_n, n \geq 0\}$ on $\mathcal{X}$ with transition probability P such that π is a *stationary* distribution for this chain. The two basic algorithms for obtaining such Markov chains are the Gibbs sampler and the Metropolis-Hastings algorithms. In this work we use a Gibbs sampler so we briefly review that algorithm. Then we will briefly discuss the method we use to compute the Monte Carlo confidence interval. For more detail on MCMC, we refer the reader to Robert & Casella (2004).

It is often the case that the sample space $\mathcal{X}$ admits the factorization $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_d$ along which π writes $\pi(x) = \pi(x_1, \ldots, x_d)$. Often $\mathcal{X} = \mathbb{R}^d$ and each $\mathcal{X}_i = \mathbb{R}$. But not always. For $i = 1, \ldots, d$, denote $\pi_i(x|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ the i-th full conditional distribution of π.

$$\pi_i(x|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) = \frac{\pi(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)}{\int \pi(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)dx},$$

with the obvious modifications when $i = 1$ and $i = d$. It is often the case, as in this paper, that direct sampling from π is not feasible while sampling from these conditional distributions is easy. The Gibbs sampling is designed for such situations. The Gibbs sampler generates a Markov chain $\{X^{(n)},\ n \geq 0\}$, $X^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$ with invariant distribution π as follows:
Given $X^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$:

1. Generate Y_1 from the conditional distribution $\pi_1\left(\cdot|X_2^{(n)}, \ldots, X_d^{(n)}\right)$.

2. For $i = 2, \ldots d$, Generate Y_i from the conditional distribution

$\pi_i\left(\cdot|Y_1,\ldots,Y_{i-1},X_{i+1}^{(n)},\ldots,X_d^{(n)}\right).$

Set $X^{(n+1)}=(Y_1,\ldots,Y_d)$.

5.1.1 Markov Chain Monte Carlo confidence interval calculation

Let $\{X_n\}$ be a Markov chain on some state space $(\mathcal{X},\mathcal{B})$ with invariant distribution π and transition kernel P and $f:\ \mathcal{X}\to\mathbb{R}$ a function such that $\int|f(x)|\,\pi(dx)<\infty$. If the Markov chain is ergodic then by the strong law of large numbers for Markov chains, $\hat{\pi}_n(f):=\frac{1}{n}\sum_{k=1}^{n}f(X_k)$ converges to $\pi(f):=\int f(x)\pi(dx)$ as $n\to\infty$. Under various conditions on π, P and f, $\frac{1}{n}\sum_{k=1}^{n}f(X_k)$ satisfies a $\sqrt{n}$-central limit theorem with asymptotic variance:

$$\sigma^2(f)=\mathbb{V}ar_\pi(f(X_0))+2\sum_{k=1}^{\infty}\mathbb{C}ov_\pi\left(f(X_0),f(X_k)\right).$$

In this expression, $\mathbb{V}ar_\pi(f(X_0))$ and $\mathbb{C}ov_\pi\left(f(X_0),f(X_k)\right)$ denote the variance of $f(X_0)$ and the covariance between $f(X_0)$ and $f(X_k)$ respectively, assuming $X_0\sim\pi$. Therefore if $\sigma_n^2(f)$ is a consistent estimate for $\sigma^2(f)$ obtained from the MCMC sample $(X_0,\ldots,X_n)$, a $(1-\alpha)\%$ Monte Carlo confidence interval for $\pi(f)$ is given by:

$$\left(\hat{\pi}_n(f)+z_{\alpha/2}\frac{\sigma_n(f)}{\sqrt{n}},\hat{\pi}_n(f)+z_{1-\alpha/2}\frac{\sigma_n(f)}{\sqrt{n}}\right), \tag{15}$$

where z_p is the p-order quantile of the standard normal distribution.

Let $(X_0,\ldots,X_n)$ be the MCMC sample. Denote $\sigma_{0,n}^2(f)$ the empirical variance of $(f(X_0),\ldots,f(X_n))$ and $\hat{\rho}_k(f)$ the empirical autocorrelation at lag k from $(f(X_0),\ldots,f(X_n))$. We use the following estimate for $\sigma^2(f)$, known to be consistent for $\sigma^2(f)$ under certain conditions:

$$\sigma_n^2(f)=\sigma_{0,n}^2(f)\left[1+\frac{2B}{B-1}\sum_{k=1}^{B}K\left(\frac{k}{B}\right)\hat{\rho}_k(f)\right]. \tag{16}$$

B is the truncation threshold (we use $B=1000$) and K, the Parzen kernel.

5.1.2 Detail of the MCMC sampler used in this paper

Following Albert & Chib (1993), we introduce a latent variable u_{ilk} and rewrite the model as followed:

$$y_{ilk}|u_{ilk}=\begin{cases}1 & \text{if } u_{ilk}>0\\ 0 & \text{otherwise}\end{cases},$$

$$u_{ilk}|x_{ilk},s_{ilk},\beta_{il},\alpha_{il}\sim N\left(x_{ilk}'\beta_{il}+s_{ilk}\alpha_{il},1\right),\quad k=1,\ldots,n_{il},\ i=1,\ldots,D,\ l=1,2,3,$$

$$s_{ilk} = 1 \quad \text{if} \quad v_{ilk} > -\bar{v} \quad \text{and} \quad 0 \quad \text{otherwise, where}$$

$$v_{ilk}|z_{ilk}, b \overset{i.i.d}{\sim} N(z'_{ilk}b, 1), \quad k = 1, \dots, n_{il}, \; l = 1, \dots, 3, \; i = 1, \dots, D;$$

$$b \sim N(0, \tau^2 I_q).$$

$$\beta_{il}|\beta_l, \sigma_l^2 \overset{i.i.d}{\sim} N(\beta_l, \sigma_l^2 I_p), \quad i = 1, \dots, D, \; l = 1, 2, 3,$$

$$\beta_l \overset{i.i.d}{\sim} N(0, \tau^2 I_p), \quad l = 1, 2, 3$$

$$\alpha_{il}|\alpha_l, \sigma_l^2 \overset{i.i.d}{\sim} N(\alpha_l, \sigma_l^2), \quad i = 1, \dots, D, \; l = 1, 2, 3,$$

$$\alpha_l \overset{i.i.d}{\sim} N(0, \tau^2), \quad l = 1, 2, 3$$

$$\sigma_l^2 \overset{i.i.d}{\sim} IG(\nu, \delta), \quad l = 1, 2, 3$$

$$\tau = 100 \;\; \nu = 1.0, \; \delta = 2.01.$$

Writing $u = (u_{ilk})$, $v = (v_{ilk})$ and $\theta = ((\beta_{il})_{il}, (\alpha_{il})_{il}, (\beta_l)_l, (\alpha_l)_l, (\sigma_l^2)_l, b)$, then we sample from the posterior distribution of (θ, u, v)using a Gibbs sampler which samples successively as follows.

Sampling σ_l^2: We sample σ_l^2, $l = 1, 2, 3$ from its full conditional distribution which is an inverse Gamma distribution $IG(\nu_1, \nu_2)$ where: $\nu_1 = \nu + D(p+1)/2$, where p is the dimension of β_{ilk}, and $\nu_2 = \delta + \sum_{i=1}^{D} (\beta_{il} - \beta_l)'(\beta_{il} - \beta_l)/2 + \sum_{i=1}^{D} (\alpha_{il} - \alpha_l)^2$.

Sampling (β_l, α_l): We sample (β_l, α_l) together from its full conditional distribution which is a normal distribution $N(M_l, V_l)$ with mean M_l and variance V_l, where

$$V_l = \frac{1}{\frac{1}{\tau^2} + \frac{D}{\sigma_l^2}} I_{p+1}, \quad M_l = \frac{1}{\sigma_l^2} V_l \sum_{i=1}^{D} \begin{pmatrix} \beta_{il} \\ \alpha_{il} \end{pmatrix}.$$

Sampling $(\beta_{il}, \alpha_{il})$: Denote $U_{il} = (u_{ilk})_k$, $X_{il} = (x'_{ilk}, s_{ilk})_k$ and write:

$$V_{il} = \left(\frac{1}{\sigma_l^2} I_{p+1} + X'_{il} X_{il} \right)^{-1}, \quad M_{il} = V_{il} \left(\frac{1}{\sigma_l^2} \begin{pmatrix} \beta_l \\ \alpha_l \end{pmatrix} + X'_{il} U_{il} \right).$$

For $i = 1, \dots, D$ and $l = 1, 2, 3$, we sample $\begin{pmatrix} \beta_{il} \\ \alpha_{il} \end{pmatrix}$ from its full conditional distribution which is a normal distribution $N(M_{il}, V_{il})$.

Sampling u_{ilk}: We sample u_{ilk} from its full conditional distribution which is a truncated Gaussian distribution with density proportional to $\exp\left(-\frac{1}{2}(u_{ilk} - x'_{ilk}\beta_{il} - s_{ilk}\alpha_{il})^2\right)\mathbf{1}_{(-\infty,0)}(u_{ilk})$ if $y_{ilk} = 0$ and proportional to $\exp\left(-\frac{1}{2}(u_{ilk} - x'_{ilk}\beta_{il} - s_{ilk}\alpha_{il})^2\right)\mathbf{1}_{(0,\infty)}(u_{ilk})$ if $y_{ilk} = 1$.

Sampling b: We sample b from its full conditional distribution $N(M_b, V_b)$, where

$$V_b = \left(\frac{1}{\tau^2} I_q + Z'Z\right)^{-1}, \quad M_b = V_b\left(Z'V\right),$$

where $Z = (z_{ilk})_{ilk}$ and $V = (v_{ilk})_{ilk}$.

Sampling v_{ilk}: We sample each v_{ilk} from its full conditional distribution which is proportional to:

$$\begin{aligned}\phi(v_{ilk}) &= p_{ilk}^{y_{ilk}}(1-p_{ilk})^{1-y_{ilk}} e^{-\frac{1}{2}(v_{ilk}-z'_{ilk}b)^2}\mathbf{1}_{(-\infty,\bar{v})}(v_{ilk}) \\ &\quad + q_{ilk}^{y_{ilk}}(1-q_{ilk})^{1-y_{ilk}} e^{-\frac{1}{2}(v_{ilk}-z'_{ilk}b)^2}\mathbf{1}_{[\bar{v},\infty)}(v_{ilk}),\end{aligned}$$

where $p_{ilk} = \Phi(z'_{ilk}\beta_{il})$ and $q_{ilk} = \Phi(z'_{ilk}\beta_{il} + \alpha_{il})$, and Φ is the cdf of the standard normal distribution.

The actual simulation: We run the actual simulation for $N = 120,000$ iterations. We discard the first $20,000$ iterations. The results reported in the paper are based on the $100,000$ iterations left.

5.2 Notes on the political process in Benin

The Republic of Benin (formerly Dahomey) is a former French colony, located in West Africa between Togo and Nigeria and is considered one of the success stories of democratization in Africa. According to a survey by Reporters without borders, Benin ranked second in Africa in terms of freedom of the press and 25th in the word ahead of the US, Japan and Italy. The president is elected through simple majority rule with run-off elections.[8]

Benin presents a number of advantages for a political experiment. First and foremost, it is considered one of the most successful cases of democratization in Africa. Thus, elections are meaningful and voters' policy preferences can be inferred from their behavior at the polls. Benin is perceived by many political scientists as the "democracy laboratory of Africa" because its political elite has the reputation to be open to political experiments. [9] Finally, the distribution of votes in

[8]That is, if no candidate obtains a majority during the first round, a second round is organized for the top two candidates on the list and the plurality winner is elected.

[9]For instance, the political leaders in Benin were the first to introduce the rotating presidency formula to curb ethnic strife in 1969. This formula was later adopted by leaders of the former Yugoslavia in 1980 following Tito's death. Benin also invented the national conference formula in 1989 as a way of facilitating a peaceful post-authoritarian transition (Boulaga [1993])

previous elections in the country is such that the risk of a field experiment seriously affecting the outcome of the 2001 election was non-existent. This is because (1) nationwide election outcomes have always revealed a significant gap between the top two candidates (Kerekou and Soglo) and the remaining candidates and (2) electoral support for those top two candidates has always been between 27 to 37%.[10] As a result, a second round election posing Kerekou against Soglo in the 2001 presidential elections was a near certainty.

There are twenty nine ethnic groups in the country and they fall into four major Ethno-Linguistic groups (Adja-Fon, Bariba, Otamari, Yoruba). Democratic reforms in the early 1990s led to a proliferation of ethnic parties: there are up to 80 ethnic parties with 16 of them effectively represented in the National Assembly. The main government parties are the Action Front for Renewal and Development (FARD-Alafia) led by Saka Salley, which provides the main grassroots support for the current government in the northern region; the Social Democratic Party (PSD) which is led by Bruno Amoussou and the African Movement for Democracy and Progress (MADEP) led by Sefou Fagbohoun. The opposition coalition is comprised of the Benin Renaissance party (RB) based in the south and central regions and led by the former presidential couple Nicephore and Rosine Soglo; the Union of Democracy and National Solidarity (UDS) led by Saka Lafia based in the north-east region and finally the Party for the Democratic Renewal (PRD) led by the current National Assembly President Adrien Houngbedji based in the south-east region. The main feature of ethnic politics in Benin is that ethnic coalitions in government and opposition are very unstable.

6 REFERENCES

Albert, J. H. and Chib, S. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669-679.

Alesina Alberto, Reza Baqir and William Easterly. 199x. “Public Goods and Ethnic Divisions”, *Quarterly Journal of Economics,* CXIV, 1243-1284.

Alesina, Alberto and Dani Rodrik. 1994. “Distributive Politics and Economic Growth”, *Quarterly Journal of Economics*, 109, 465-490.

[10] In 1991, Soglo obtained 27.2% of the vote, Kerekou 36.30 and the next candidate Tevoedjre 14.21%. In 1996, Soglo received 35.69% of the vote, Kerekou 33.94% and Houngbedji 19.71%.

Banegas, Richard. 1998. "Bouffer l'Argent, Politique du Ventre, Democratie et Clientelisme au Benin" in Jean-Louis Briquet et Frederic Sawicki (eds) *Clientelisme Politique dans les Societes Contemporaines*, Presses Universitaires de France.

Bates, Robert H. 1983. "Modernization, Ethnic Cooperation, and the Rationality of Politics in Contemporary Africa." In *State versus Ethnic Claims: Africa Policy Dilemmas*" Donald Rotshild and Victor Olunsorola, eds. Boulder Colorado, CO: Westview Press.

Besley, Timothy and Maitreesh Ghatak. 2003. Public Goods and Economic Development, Prepared for "Policies for Poverty", A. Banerjee, R. Benabou, and D. Mookherjee, Princeton University Press.

Boulaga, Eboussi, 1993. *Les Conferences en Afrique Noire.* Paris: Editions Karthala.

Bollen, Kenneth A. 1989. Structural Equations With Latent Variables. New York: Wiley.

Degboe Kuassi .1995. *Elections et Realites Sociologiques du Benin.* Cotonou: Entermonde Editions

Dixit Avinash and John Londegran. 1996. "The Determinants of Success of Special Interest in Redistributive Politics." *Journal of Politics*, Vol. 58, pp. 1132-1155.

Easterly William. 2001. The Elusive Quest for Growth: the Economists and Misadventures in the Tropics. Cambridge: MIT Press

Easterly William and Ross Levine. 1997. Africa Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics,* 112, 1203-1250..

Fearon, James D. 1999. "Why Do Ethnic Politics and Pork Tend to Go Together" Working Paper. Stanford University.

Fearon, James D. and David Laitin. 1996. "Explaining Ethnic Cooperation." *American Political Science Review* 4 pp. 715-35.

Gelman, Andrew; Carlin, John B.; Stern, Hal S. and Rubin, Donald B. 2004. Bayesian Data Analysis, 2nd Edition, Chapman & Hall/CRC.

Gerber Alan and Donald Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct mail on Voter Turnout: A Field Experiment." *American Political Science Review*, Vol. 94, No. 3, pp. 653-662

Hellbrunn, John R. 1993. "Social Origins of National Conferences in Benin and Togo". *Journal of Modern African Studies* v 31: 277-99.

Institut de Statistique et d'Analyse Economique (INSAE) and IRC Macro, 2001. Benin 2001: Enquete Demographique et de Sante. Cotonou, Benin.

Lemarchand, René. 1977. "Political Clientelism and Ethnicity in Africa: Competing Solidarities in Nation-Building", in *Friends, Followers, and Factions: A Reader in Political Clientelism*, Edited by Steffen Schmidt, Laura Guasti, Carl Landé and James C. Scott. Berkeley: University of California Press.

Nwajiaku, Kathryn. 1994. "The National Conferences in Benin and Togo Revisited." *Journal of Modern African Studies* v 32: 429-47.

Nielsen, François 1985. "Toward a Theory of Ethnic Solidarity in Modern Societies", *American Sociological Review*, 50, 133-149.

Rabe-Hesketh, Sophia; Skrondal, Anders and Pickles, Andrew 2004. Generalized multilevel structural equation modeling. *Psychometrika*, Vl. 69, No. 2, 167-190.

Robert, Christian and Casella, George 2004. Monte Carlo statistical methods, 2nd Edition. New York: Springer.

Varshney, Ashutosh. 2002. *Civic Life and Ethnic Conflict: Hindus and Muslims in India* (New Haven: Yale University Press)

Wantchekon, Leonard. 2003. Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics,* Vol. 55, No. 3, 399-422.

World Bank. 2002. Engendering Development: Through Gender Equality in Rights, Resource and Voice. *Oxford University Press and World Bank.*