Co-ethnics Co-vote in Africa: Studying Electoral Cleavages with a Co-Voting Regression Model*

Carl Müller-Crepon† Nils-Christian Bormann‡

September 13, 2024

Abstract

Ethnicity is an important political cleavage in Africa, yet the degree of its influence on voting is contested. Selection biases from restricted choice sets complicate micro-level analyses, while bias from ecological inferences and unobserved confounders hamper meso and macro-level approaches. We develop the *Co-Voting Regression* (CVR) model to tackle these challenges. It estimates the effect of co-ethnicity on the probability that pairs of voters co-vote for the same party/candidate while conditioning on other characteristics that connect voters. In doing so, CVR mirrors the micro-foundations of widely-used aggregate indicators, such as the effective number of parties and the Herfindahl-Hirschman index of ethnic homogeneity. Our data consists of dyadic comparisons between respondents from Afrobarometer surveys. Pooling across 28 countries, our results show that co-ethnicity increases co-voting intentions by 16 percentage points. The effect of co-ethnicity is consistent across institutionally diverse countries and at least five times larger than that of other cleavages. Beyond ethnicity, the approach we propose addresses key methodological concerns in studies of the electoral consequences of socio-economic cleavages and bridges gaps between levels of analysis.

Keywords: Cleavages; Voting; Ethnic voting; Ethnic parties; Ethnicity; African politics; Computational Methods

*We are grateful for helpful comments by Paola Galano Toro, Sara B Hobolt, Michael Laver, Emmy Lindstam, and panel participants at the General Conference of the European Political Science Association, 22-24 June 2023, in Glasgow, UK, and the Political Behavior Work in Progress Seminar at the London School of Economics.

†Department of Government, London School of Economics and Political Science. Corresponding author: `c.a.muller-crepon@lse.ac.uk`

‡Department of Philosophy, Politics & Economics, Witten/Herdecke University. `nils.bormann@uni-wh.de`

It used to be a truism among political scientists that African voters would support co-ethnic candidates and African parties would target co-ethnics in election campaigns (e.g., Horowitz 1985; Rabushka and Shepsle 1972; Bayart 2009).[1] Increasingly, however, researchers question this hypothesis. Some micro-level studies propose a broader sociological understanding of voting by identifying other relevant cleavages like religion or urban-rural differences (e.g., McCauley 2014; Nathan 2016); others adopt rationalist perspectives that emphasize individual economic interests and the quality of information available to voters (e.g., Bratton, Bhavnani and Chen 2012; Casey 2015; Ferree, Gibson and Long 2021). Meso- and macro-level comparative analyses identify electoral systems and ethnic inequality as conditioning factors of ethnic voting (Huber 2012; Huber and Suryanarayan 2016). Some studies even suggest that the effect of co-ethnicity on vote choice and party systems in Africa is entirely spurious, and simply reflects underlying geographic clustering (Ferree and Horowitz 2010; Boone et al. 2022).

Yet, micro, meso, and macro-level analyses of the electoral effects of ethnic cleavages suffer from potentially severe methodological problems. At the micro-level, the interdependence between observed voting patterns and the fixed menu of parties and candidates in any one election complicates inference. If researchers conceptualize ethnic voting as the support of voters for candidates from the same ethnic group, they risk *selection bias* if some groups do not field candidates (see also Ferree 2022). Moreover, the idiosyncratic and ever-changing menu of parties or candidates in individual elections prevent comparisons across countries and elections. Meso and macro-level comparative approaches address this challenge by analysing the degree to which ethnic groups vote for the same party (Huber 2012; Huber and Suryanarayan 2016; Houle, Park and Kenny 2019). Yet, inferring individual-level voting motivations from group or country-level analyses constitutes a clear case of *ecological inference*. Since ethnic cleavages frequently correlate with other social divisions, for example, geographic or economic differences, this approach might be biased by omitting such non-ethnic determinants of vote choice (Ferree and Horowitz 2010; Boone et al. 2022).

[1] Our argument focuses on elections of parties as well as individual candidates, which is why we use these labels interchangeably unless otherwise noted.

We introduce a Co-Voting Regression (CVR) model as a new analytical approach that combines the complementary strengths of micro-level voting studies and comparative work at the meso and macro-levels to solve the problems affecting either. Like the aggregate indices employed by comparativists, we assess the likelihood of co-voting intentions among pairs of individuals.[2] Yet, instead of building aggregate country or group-level measures, we model co-voting preferences at the micro-level of individual *pairs of voters* in a linear probability model.[3] This allows us to estimate the effect of co-ethnicity (and any other cleavage) between individuals conditional on other dyadic socio-economic characteristics with a standard linear probability model. Beyond their direct micro-level interpretations, we show that model coefficients capture the elasticity of party system concentration with regard to changes in countries' ethnic homogeneity at the macro-level. This congruence arises because the CVR directly mirrors the micro-foundations of the classic Herfindahl-Hirschman Index. Our method thus bridges the prevailing gap between the differing levels of analysis in studies of electoral behavior and outcomes with possible applications to the study of cleavages beyond ethnicity.

CVR solves the aforementioned methodological problems of micro, meso, and macro-level approaches. By modeling pairwise co-voting, parties and candidates disappear from our formulation. This strongly reduces selection biases and allows for comparative analyses across countries and over time.[4] At the same time, our study of the co-occurrence of co-ethnicity and co-voting intentions at the micro-level circumvents ecological inference problems while drastically increasing statistical power and the ability to account for confounders. Finally, our contribution in this article is primarily methodological and empirical, not theoretical. However, our conceptual focus on dyadic co-voting brings sociological explanations of vote choice into focus (Lazarsfeld, Berelson and Gaudet 1968/1944). It thus complements the individual-based theoretical accounts rooted in psychologocial and

[2]Empirically, we measure co-voting intentions or preferences for parties and candidates. Conceptually, our model applies to actual co-voting, and we use the shorter term when discussing theoretical applications

[3]Our dyadic approach shares similarities with the study of international relations where joint unit-level features of countries influence collective action such as the formation of trade blocs, military alliances, or peace communities.

[4]Selection bias continues to be a problem when respondents have no (stated) voting intentions.

rational choice paradigms that dominant research on voting in Sub-Saharan Africa (Bates 1974; Horowitz 1985).

Empirically, we test seminal theories of ethnic voting and use the CVR to provide comprehensive evidence on the effect of co-ethnicity on co-voting preferences across Sub-Sahara Africa. Drawing on multiple rounds of the Afrobarometer surveys from 28 states across Sub-Sahara Africa, we recast each country sample into pairwise comparisons between respondents. We capture our *outcome variable* by measuring co-voting intentions in presidential elections and shared preferences for political parties. Co-ethnicity, the main *explanatory variable*, is measured as a shared mother tongue among respondents. Shared demographic, economic, and geographic characteristics as well as survey-round-fixed effects constitute our controls. Building on our conceptual derivation of the CVR, we estimate the probability of co-voting preferences in linear probability models, and adjust standard errors to address the repeated inclusion of individuals from a limited number of ethnic groups across many dyads.

Our results show strong support for the dominance of co-ethnicity in determining co-voting intentions in most African states in our sample. Co-ethnicity by mother tongue increases the probability that two respondents share voting preferences by 16 percentage points or 35 percent of the mean rate of co-voting intentions. The result suggests an average elasticity of the concentration of countries' party system to changes in their ethnic homogeneity by .16. In addition to alternative measures of co-ethnicity and linguistic distance, different sampling procedures, and modelling choices, our results are robust to studying co-voting preferences only within administrative regions and survey enumeration areas. This shows that ethnic cleavages are more than just reflections of geographically determined political preferences.

Zooming in on specific countries and elections, we discuss variation in the effect of co-ethnicity on co-voting intentions over time and across cases. In a set of descriptive analyses, we do not find that electoral systems, the level of democracy, or the strength of traditional institutions moderate the effect of co-ethnicity on co-voting preferences. While our results also shows positive effects of socio-economic characteristics discussed in the literature such as religious, educational, occupational, and geographic similarities (e.g., Boone et al. 2022; Bratton, Bhavnani and Chen 2012; Koter 2016; McCauley 2014), these are at least 5 times smaller across our broad sample of African elections than the effect of co-ethnicity.

We conclude by discussing the implications of our new analytical approach and findings for the wider study of electoral cleavages in Sub-Sahara Africa and beyond. In particular, we highlight the utility of the CVR in studying the electoral effect of socio-economic cleavages beyond ethnicity and discuss potential extensions to apply it to local election results which would overcome its reliance on individual-level data. We end with a more general note on the importance of bridging micro, meso, and macro-levels of analysis to achieve inferences of high internal and external validity.

Ethnicity and Voting in Sub-Sahara Africa

Political scientists have come a long way from the once paradigmatic view that elections in Africa constituted an ethnic census (Horowitz 1985, 196). Classic works on vote choice in Sub-Saharan Africa either stress instrumental or psychological motivations for ethnic voting and the corresponding existence of ethnic parties (Mozaffar, Scarritt and Galaich 2003). In short, instrumentalists suggest that African voters support co-ethnic candidates to receive economic benefits through clientelist exchanges during the election period and patronage distribution afterwards, if their co-ethnic candidate joins the ruling coalition. Political elites themselves prefer to build ethnically-based support coalitions in order to limit access to state funds to ethnic insiders (Bates 1974; Laitin and van der Veen 2012). The psychological approach entails that voters reaffirm their identity through voting for co-ethnic candidates and attempt to avoid discrimination by ethnically distinct rulers. Political leaders cannot escape the logic of ethnic outbidding, in which more extreme political demands on behalf of co-ethnics gain more electoral support (Rabushka and Shepsle 1972; Horowitz 1985).[5]

However, over the last two decades the dominant role of ethnicity in shaping

[5] As briefly discussed in the introduction, sociological explanations of voting following Lazarsfeld, Berelson and Gaudet (1968/1944) are curiously absent from both classic and recent analyses of ethnic voting, a theme we return to in the conclusion.

vote choice and party systems has come under scrutiny. More recent micro-level studies question both the mechanism by which ethnic identity explains vote choice, and explore alternative theoretical explanations. Scholars studying the first question frequently use (quasi-)experimental methods to understand the mechanisms that underlie the positive correlation between co-ethnicity and vote choice. Primarily, this research program questions psychological theories of expressive voting. The main alternative suggested by these studies holds that ethnicity simply constitutes an informational shortcut that signals the likelihood of economic benefits voters might receive from co-ethnic rule (Ferree 2006; Carlson 2015). Several experimental and quasi-experimental studies that vary the amount of information voters have about candidates support the instrumental interpretation (Conroy-Krutz 2013; Casey 2015; Carlson 2018; Ferree, Gibson and Long 2021).[6] Overall, these findings support an instrumentalist interpretation of ethnic voting while suggesting that ethnicity would cease to affect vote choice if African voters had more information about their candidates, or if other cleavages could fulfill the informational role of ethnicity (Dunning and Harrison 2010).

Another strand of voting research explores alternative voter motivations on the basis of survey data. Several studies pit co-ethnicity with presidential incumbents against prominent motivations found among voters in the United States and Europe, including economic performance evaluations and education (Bratton and Kimenyi 2008; Bratton, Bhavnani and Chen 2012), partisanship (Ferree and Horowitz 2010; Hoffman and Long 2013), populism (Resnick 2012), and urban-rural differences (Nathan 2016, 2019). Others stress that ethnic voting depends on local factors characteristic for many African countries, such as the presence and importance of traditional authorities (Baldwin 2013; Koter 2016), and the make-up of local ethnic geography (Ichino and Nathan 2013). The relative prominence of ethnic voting varies across these single-country or small-N case studies.

Although each of these studies is innovative in its own way, we note two limitations. First, the focus on one or few countries raises questions about the generalizability of results. Second, most studies measure ethnic voting as stated support for

[6] Though see Adida et al. (2017), who highlight how performance evaluation is inextricably linked to ethnic identity through motivated reasoning.

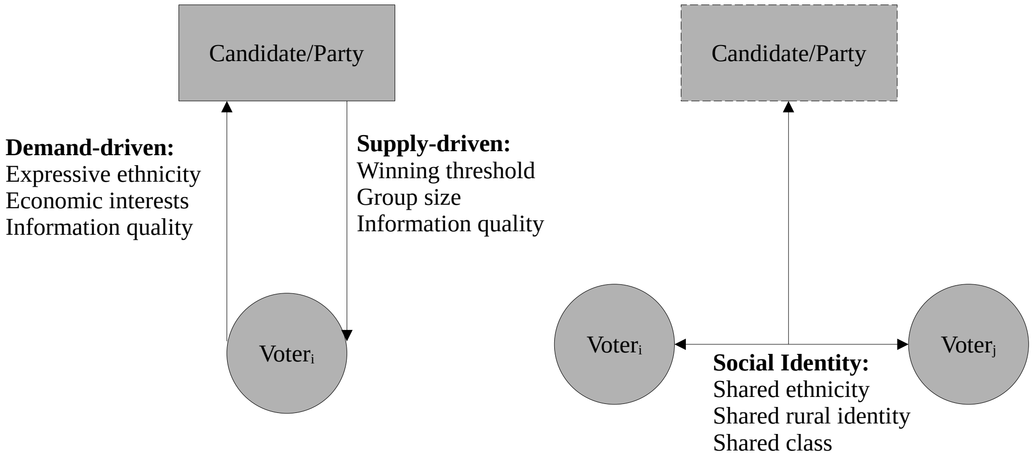
co-ethnic candidates. This analytical choice raises conceptual and methodological issues. In an important recent contribution, Ferree (2022) shows that two thirds of voters in Ghanaian, Ugandan, and Kenyan legislative elections faced a choice between either only co-ethnic candidates or exclusively candidates from different ethnic groups. Methodologically, a limited "choice set" introduces selection bias into existing analyses of voter motivations particularly where the choice set is shaped by expectations about voters' electoral behavior.

To illustrate this bias, we consider a country with three ethnic groups that each constitute one third of the population respectively. We further assume that each individual obtains the same positive utility from voting for a co-ethnic candidate. Yet only parties representing two out of the three groups field candidates for an election. Any statistical analysis will underestimate the strength of ethnic voting because 33.3% of population cannot vote for a co-ethnic candidate. Put differently, a "0" classification for co-ethnicity in vote-choice outcomes mixes the absence and the impossibility of ethnic voting. Nathan (2016), for example, attempts to avoid this problem by dropping all individuals without a co-ethnic on the ballot. Yet, this causes selection bias if the "missing" candidate was not fielded in anticipation of a lack of ethnically structured support from the respective group members. The same problem arises when the choice set is broad but the analyst artificially limits vote choice, for example by only evaluating support for the incumbent (e.g., Bratton and Kimenyi 2008; Bratton, Bhavnani and Chen 2012).

The selection bias we describe raises broader conceptual questions about ethnic voting conceived as co-ethnicity between voters and their preferred candidate. The left panel in Figure 1 illustrates the prevalent approach by summarizing existing vote choice motivations under the demand category. However, voters can only vote for candidates and parties that are on the ballot, and it is elites who decide to run in anticipation of electoral success. Thus, observational voting studies that capture ethnic voting through candidate-voter co-ethnicity would need to adjust their work for the factors that influence candidate supply, as Ferree (2022) convincingly argues.[7]

[7] This concern does not affect experimental studies of voters' attitudes towards synthetic candidates with attributes fully under the control of the researcher.

Figure 1: Different conceptualizations of vote choice.



Comparativists study these supply factors when investigating different types of party systems. Moving to the macro-level, they follow up on Duverger's famous prediction that party system size is a function of social cleavages and electoral rules. Equating ethnic cleavages measured through ethnic fragmentation indices with demand for parties, and the permissiveness of electoral systems as a proxy for party supply, previous research supports the notion that the effective number of ethnic groups in a country correlates with the effective number of parties (Mozaffar, Scarritt and Galaich 2003; Clark and Golder 2006; Lublin 2017). While these studies show that elites consider the ethnic landscape when forming parties, they cannot rule out that ethnic diversity indices capture other underlying cleavages, such as shared regional-economic interests (cf. Boone et al. 2022; Ishiyama 2012).

Indeed, case studies frequently show that party competition goes beyond ethnicity. Elischer (2013), for example, describes catch-all, programmatic, and personalistic parties in three African states. Horowitz (2022) shows how presidential candidates in Kenya pursue swing voters among non-co-ethnics. Both studies demonstrate that group size strongly determines whether or not a group fields a candidate or party of their own (cf. Posner 2004). Bridging the gap from this macro-level size criterion to micro-level findings of the importance of other cleavages, Ferree (2010) shows how divisions within the largest ethnic group of a country induce intra-group competition and high levels of electoral volatility.

While research at the macro-level thus points to the supply-factors that are missing from micro-level research, it frequently starts from an assumption of fixed ethnic preferences among voters by positing that cleavages directly translate into party demand. Small-N studies point to the relevance of group size in determining whether party supply will be viable but potentially overlook an expressive demand for party representation among individuals that belong to smaller groups with little chance of winning elections directly (Mor 2022).

In recent years, some comparativists set out to reconcile variation in individual voter preferences and supply factors at the meso-level. First introduced by Huber (2012), they model ethnic voting as the joint vote of group members for one party (right panel, Figure 1). This group-level approach side-steps concerns about constrained choice sets, because individuals from smaller ethnic groups without their own candidate can vote together for non-co-ethnics. As long as they support the same candidate, their behavior would be classified as ethnic voting. Another advantage of these studies is their explicit recognition of the contextual nature of identity effects. Huber (2012), for example, finds that ethnicity is less predictive of vote choice in countries operating proportional representation systems and in decentralized states. Huber and Suryanarayan (2016) and Houle, Park and Kenny (2019) find that higher levels of between-group economic inequality increase the likelihood of ethnic voting, especially when within-group inequality is low.

Yet as the group-based analytical strategy relies on *ecological inference* to deduce individual voting motivations, it risks omitted variable bias that stems from two distinct sources. First, group-level analyses reify group boundaries and thereby preclude a more nuanced understanding of potentially variable ethnic boundary markers, such as language versus religion, or even endogenous processes of identity change (Posner 2005; McCauley 2014; Green 2021; Müller-Crepon 2023). Second, and more importantly, ethnic group-level analyses start from an assumption that ethnicity is the prime dimension along which voting behavior is structured. Yet individual-level voting studies stress the relevance of several other non-ethnic boundary markers in influencing individual voting decisions, including urban-rural differences (Nathan 2016; Wahman and Boone 2018; Harding 2020), regional coalitions (Ferree and Horowitz 2010; Boone et al. 2022), and economic class interests (Bratton, Bhavnani and Chen 2012; Resnick 2012). The threat of omitted variable bias increases to the extent that ethnicity is endogenous to or correlated with these cleavages.

In sum, research on voting and parties in Africa increasingly uncovers evidence that voting intentions and party programs are as diverse as in other regions of the world. Cleavages beyond ethnicity, such as individual-level economic interests, urban-rural differences, and shared economic-territorial preferences, matter for individual vote choices and party programs. To understand the limits and remaining power of ethnicity for vote choice in Africa, we then need a novel analytical approach that addresses (some of) the limitations of existing work on voting in Africa and beyond. Particularly, such a study should (1) recognize and separate candidate demand and supply to overcome selection bias; (2) account for alternative motivations and identity categories that might serve as a basis for voting to avoid omitted variable bias; and (3) compare a broad number of countries over time to ensure external validity.

Co-Voting Regression: A new way to estimate the effects of (ethnic) cleavages

Our new Co-Voting Regression (CVR) bridges prevailing approaches to study the effect of ethnic cleavages by combining their strengths which together address their respective weaknesses. In short, we follow macro- and meso-level measurement strategies to conceptualize the effect of ethnic cleavages as the effect of co-ethnicity on co-voting intentions. Overcoming reliance on aggregate indicators which leads to problems of ecological inference, we study co-voting preferences in pairs of individuals contained in survey data with standard regression models. In the following, we introduce our approach to bridging the micro, meso, and macro-levels of analysis, explain the transformation of the Afrobarometer data into voter pairs, and discuss the empirical CVR specification.

Bridging micro, meso, and macro-approaches

Macro-approaches often study the structure of party systems as the result of an interplay between institutional determinants and socio-demographic factors such as societal cleavages. Empirically, they employ aggregate measures to operationalize the main variables of interest, such as the effective number of parties (ENP) to measure party system fragmentation, or Herfindahl-Hirschman Indices (HHI) of ethnic homogeneity. Yet, while the empirical focus is on the macro-level, these measures have explicit meso- and micro-foundations in theory and measurement. In particular, the HHI and its inverse, Laakso and Taagepera's (1979) ENP are often computed with meso-level measures of party's vote share as

$$ENP = HHI^{-1} = \left(\sum_{p=1}^{K} \left(\frac{N_p}{N} \right)^2 \right)^{-1}, \tag{1}$$

where s_p is the vote share of party $p \in K$. The effective number of parties thus increases with more and more equally sized parties. As scholars of ethnic fragmentation note, the HHI has clear micro-foundations: it reflects the chance that two randomly drawn individuals belong to the same group or category (e.g., Alesina, Baqir and Easterly 1999). In terms of voting, the relative vote shares of one party simply reflect the probability that an individual voter supported that party. Squaring that probability then yields the chance that two random voters voted for the same party. We can thus reformulate the definition of the ENP as as the inverse of the chance that two randomly chosen voters vote for the same party:

$$ENP = \left(\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{co-voting}_{i,j} \right)^{-1}, \tag{2}$$

where i and j are individual voters drawn from all voters N, and $\text{co-voting}_{i,j}$ is an indicator that returns 1 if i and j co-vote for the same party p and 0 otherwise.

Rarely discussed, the co-voting formulation in Equation (2) includes comparisons within the same individual $i = j$, inducing downward bias when the ENP in Equations (1) and (2) is computed from a finite sample of individuals N. This is because comparisons within the same voter $i = j$ must yield $\text{co-voting}_{i,j} = 1$.[8] Drawing on Simpson (1949), this bias can be corrected by avoiding "within-individual" comparisons when computing the ENP:

$$ENP = \left(\sum_{p=1}^{N} (\frac{N_p}{N} \frac{N_p - 1}{N - 1}) \right)^{-1} \tag{3}$$

$$= \left(\frac{1}{N^2 - N} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \text{co-voting}_{i,j} \right)^{-1} \tag{4}$$

Equation 4 shows how the unbiased expectation of ENP is equivalent to the inverse of the average *co-voting* rate between all voters.[9] In turn, ethnic homogeneity among voters as measured through the HHI is equivalent to the average rate of co-ethnicity between voters. Moving beyond measures of diversity, measures of dispersion such as the Gini coefficient can be similarly reformulated as comparisons between pairs of individuals.[10]

We hone in on these micro-foundations of meso and macro-level approaches and propose to model the effect of ethnic cleavages by estimating the effect of co-ethnicity on co-voting in pairs of individuals i and j. We start deriving our *Co-Voting Regression* (CVR) model as

$$\text{Co-voting}_{i,j} = \beta_0 + \epsilon_{i,j} \tag{5}$$

with $i, j \in N, i > j$.[11] In this formulation, $\hat{\beta}_0$ captures the average rate of co-voting among all pairs of individuals and is thus equivalent to ENP^{-1} in Equation (4).[12] The pairwise regression model in Equation (5) can easily be extended by adding dyadic predictors which measure individuals' similarities or difference on impor-

[8]The bias decreases proportional to the weight of the biasing within-individual comparisons with increasing N.

[9]Equation (4) approaches Equation (2) as N increases towards infinity.

[10]The Gini coefficient can be computed as half the mean absolute (wealth, income, education, etc.) difference among all pairs of individuals, see e.g., Sen (1997, p. 31).

[11]As co-voting is not directional, we can limit ourselves to one comparison between any two voters. Computationally though, the result is the same as in Equation 4.

[12]The estimate of the intercept in an empty linear regression model, $\hat{\beta}_0$, equals $\bar{y} = \frac{1}{n} \sum_i^n y_i$ (Wooldridge 2008, 29). See Appendix Figure A1 for an empirical demonstration of the equivalence. Our estimation method of ENP and HHI has the added benefit of yielding confidence intervals that reflect the uncertainty introduced by the sampling of survey respondents.

tant socio-economic cleavage dimensions as well as other control variables. We thus propose to estimate the effect of co-ethnicity on co-voting as

$$\text{Co-voting}_{i,j} = \beta_0 + \beta_1 \, \text{co-ethnicity}_{i,j} + \gamma \, \mathbf{x}_{i,j} + \epsilon_{i,j} \tag{6}$$

Due to the equivalence of the $\text{co-ethnicity}_{i,j}$ indicator in Equation (6) with the micro-level interpretation of the HHI index, the estimate for β_1 has interpretations at all levels of analysis. At the micro-level, it can be interpreted as marginal effect of co-ethnicity on the probability of co-voting between individuals. At the meso-level of pairs of ethnic groups, β_1 captures the difference in average co-voting within groups as compared to co-voting between groups, with group-pairs weighted by the product of their size. At the macro-level, it is the elasticity of the party-system concentration in response to marginal changes in ethnic homogeneity such that

$$\frac{\delta \, HHI(party)}{\delta \, HHI(ethnic)} = \frac{\delta \, \text{co-voting}_{i,j}}{\delta \, \text{co-ethnicity}_{i,j}} = \beta_1, \tag{7}$$

By mirroring the construction of the HHI at the micro-level, the regression model in Equation (6) thus effectively bridges the micro, meso, and macro-levels. This characteristic extends to other dyadic comparisons between voters that reflect macro-level measures, such as pairwise wealth differences which constitute the building blocks of the Gini coefficient.

Accommodating such additional cleavages, the CVR avoids the challenges induced by ecological inference and allows for inter-temporal and cross-country comparisons without incurring selection bias or requiring any ex ante coding or standardization of parties or candidates. To facilitate understanding of our empirical strategy, we first introduce our data structure and then present our regression model.

Dyadic data on co-voting preferences, co-ethnicity, and other cleavages

To operationalize our analysis of co-voting intentions among individuals, we transform survey data into pairs of individual respondents. For each pair, we encode whether respondents share co-voting preferences and measure ethnic and other cleavages through respondents' pairwise shared ethnicity and similarity in other socio-economic characteristics.

Our main data source consists in the nationally and, in expectation, locally representative Afrobarometer survey series which contains data on political preferences across an increasingly large set of states in Sub-Sahara Africa since 1999. For the most part, we rely on the survey's seventh round fielded between 2015 and 2018 in 28 states.[13] To gauge variation in the effect of ethnic cleavages over time, we draw on rounds 3–7.[14] In addition to surveying preferences for presidential candidates and political parties, the surveys cover a large range of demographic and economic items, and provide geographic information on respondents' place of residence. The resulting information allows us to capture co-voting intentions along with a wide range of non-ethnic cleavage dimensions discussed by existing work.

Unit of analysis: Closely following the logic introduced in Equation (4) above, we transform the data from each survey-round in each country into the set of all undirected dyadic comparisons between respondents $i, j \in N_{c,t}$ with $i \neq j$.[15] This gives rise to a total of $(N_{c,t}(N_{c,t}-1))/2$ observations per country-round. After dropping observations with missing data, our main analysis of preferences for presidential candidates (parties) draws on a median number of 688 (471) respondents and 236'328 (110'685) dyadic comparisons between them per country surveyed.[16]

[13]Benin, Botswana, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Ethiopia, Gabon, Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, South Africa, Tanzania, Togo, Uganda, Zambia, Zimbabwe.

[14]Rounds 1 and 2 do not include an item on preferences over candidates in potential presidential elections.

[15]Equation 4 does not depend on any notion of directionality in the comparison between i and j and can therefore be reformulated from directed to undirected dyads without any loss of information or precision.

[16]Appendix Figure A4 shows that our results are robust to reducing the number of comparisons per respondent down to as few as one. Weighing observations such that each country-round receives equal weights slightly increases our main estimates, see Appendix Table A8.

Measuring co-voting intentions: We encode our main two measures of co-voting intentions by drawing on answers to Afrobarometer's questions on respondents' preferences over presidential candidates and parties:

> Voting intention: If a presidential election were held tomorrow, which party's candidate would you vote for?
>
> Party preference: Do you feel close to any particular political party? Which party is that?[17]

Drawing on these items, we record two dummy variables that take the value of 1 if respondents i and j share a preference for the same candidate or party and 0 otherwise.[18] The result is visualized for a sample of 10 respondents from Ghana in Figure 2. Each dot represents one respondent with its color reflecting their preferred candidate/party. Lines between respondents are drawn in black ('1') where they share a preference and in grey ('0') where they do not. We note that the average value of these outcome variables within a country-round corresponds directly to the HHI of party concentration or the inverse ENP (see Equation 4).

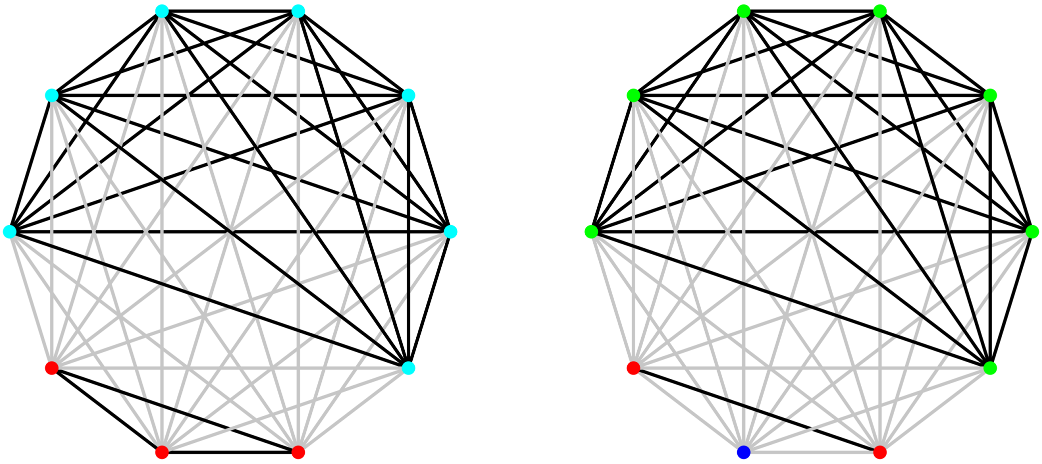
Co-ethnicity: We capture our main explanatory variable of interest – respondents' pairwise co-ethnicity – in a binary variable that records whether they share the same mother tongue (1) or not (0).[19] As visualized for the randomly drawn 10 Ghanaians in Figure 3a, this leads to many co-ethnic dyads among respondents from large language groups (e.g., the Akan in red) and non-co-ethnic ones between groups. We note again that the average pairwise co-ethnicity in a country-round corresponds directly to the Herfindahl-Hirschman Index of ethnic homogeneity.

Respondents' reported mother-tongue is among the least malleable ethnic identity indicators and therefore least likely affected by reverse causality or omitted

[17]Missing values are recorded for respondents who do not feel close to any party.

[18]We drop individuals with missing responses. For the main analysis, we recode answers classified as "other" as missing. Yet, Appendix Table A9 shows that coding such answers as separate parties for each respondent does not change the results.

[19]The respective question reads: "Which [enter nationality] language is your mother tongue or language of origin?" Note that Afrobarometer round 7 is the first to ask specifically about respondents' mother tongue as separate from the language spoken in their home *now*. Hampering comparisons over time, all previous rounds ask about respondents' "home language" which leaves this crucial distinction open.



(a) Shared voting intentions

Note: NPP green, PPP blue, NDC red, co-voting in black

(b) Shared party preferences

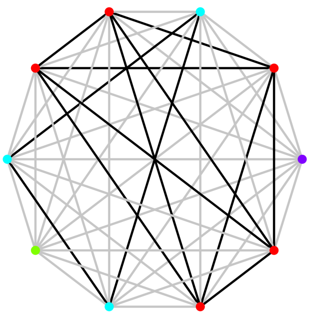
Note: NPP turquois, NDC red, co-voting in black.

Figure 2: Co-voting dyads from 10 respondents in Ghana, Round 7

variable bias. In particular in contrast to the language spoken at respondents' home or their self-proclaimed ethnic identity, mother tongues are unlikely affected by political concerns of respondents (e.g. Green 2021) and assimilation over their life-course (e.g. Müller-Crepon 2023).[20] We employ three different strategies to address the remaining potential for omitted variable bias through, for example, economic factors affecting political preferences as well as ethnic identities, and reverse causation, such as multi-generational assimilation that aligns ethnic to political identities. First, we condition our estimates on several other individual-level covariates which might affect individuals' stated ethnic origin and their political preferences. Second, we analyze variation in effects at short and large linguistic distances, which are harder to overcome through assimilation or misrepresentation. Third, a set of robustness checks zooms in on co-voting intentions among respondents from the same enumeration area, thus holding geographic factors constant.

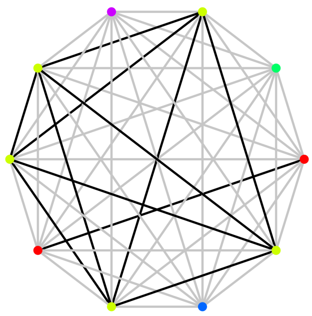
Control variables: With regard to our first strategy of conditioning on observables, we encode a set of pairwise comparisons between respondents that capture prominent political cleavages and might affect individuals' reported language. All are visualized for our exemplary 10 Ghanaian respondents in Figures 3b to 3i. For

[20]We test the alternative measurements of ethnic identity in Appendix B.1.



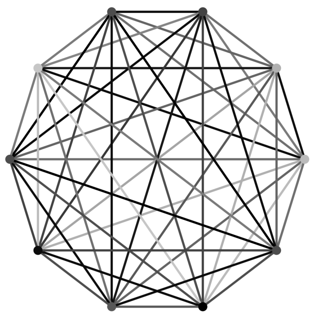
(a) Shared mother-tongues

Note: Colors denote mother-tongues, intra-group edges in black



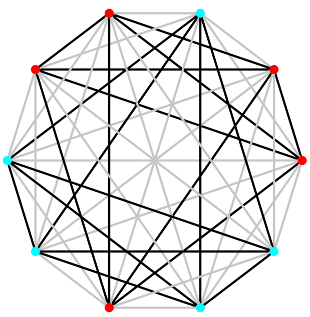
(b) Shared religion

Note: Colors denote religious groups, intra-religion edges in black



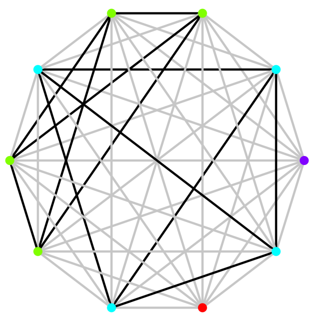
(c) Age similarity (decades)

Note: Grey-scale denotes older age and greater similarity



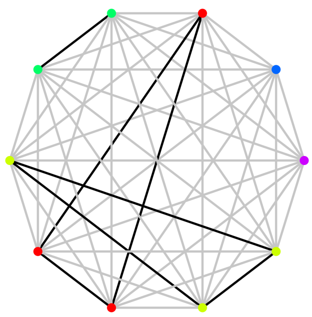
(d) Shared gender

Note: Female in light blue, male in red, intra-gender edges in black



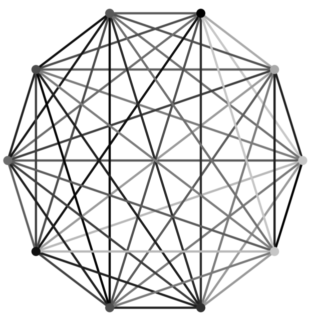
(e) Shared education

Note: Colors denote education levels, intra-education level edges in black



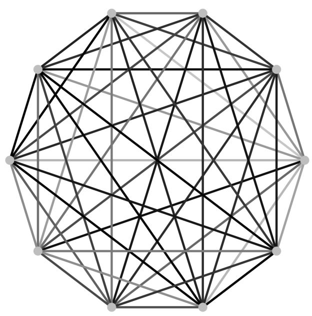
(f) Shared occupation

Note: Colors denote occupations, intra-occupation edges in black



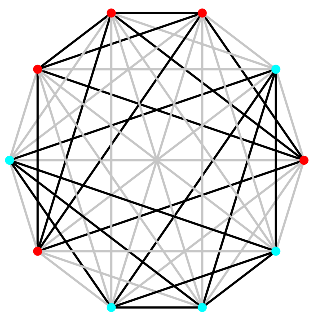
(g) Wealth similarity

Note: Dark shades denote wealthier respondents and greater similarity



(h) Geographic proximity

Note: Darker shades denote greater proximity



(i) Shared urban-rural status

Note: Urban blue, rural red, shared status edges in black

Figure 3: Encoding of main explanatory variables on example graph of 10 respondents from Ghana

reasons of consistency, we construct our measures such that larger positive values denote greater similarity between respondents which should, in expectation, come with higher probabilities of co-voting intentions.

First, we complement our measure of shared mother tongues by accounting for whether respondents share the same *religion*. Second, we capture *demographic similarities* between respondents by recording age and gender similarities. Third, we approximate *economic cleavages* by adding dummy variables for shared educational and occupational background as well as wealth similarity, measured as one minus absolute wealth differences.[21] Lastly, we capture purely *geographic cleavages* by including as-the-crow-flies proximity between respondents (in 1'000km) and a dummy variable capturing whether respondents share their urban or rural status.

Combining data across countries and rounds: Because our measures of co-voting intentions, co-ethnicity, and additional control are measured as binary or continuous indicators of similarity the data can be stacked and analyzed across countries and rounds without any additional processing. This is a substantive advantage over standard approaches of modeling the effect of ethnic (or other) cleavages on party or candidate preferences which require harmonization across context with the selection biases this gives rise to.

Modelling the effect of ethnic cleavages on co-voting preferences

With the undirected dyad of respondents i and j as our main unit of analysis, we employ a linear probability model to estimate co-voting preferences between respondents along the lines of the CVR proposed in Equation 6.[22] We move beyond the confines of one survey sample of individuals and generalize the model across countries and time as

$$\text{Co-voting}_{i,j,c,t} = \alpha_{c,t} + \beta_1 \, \text{co-ethnic}_{i,j} + \gamma \, \mathbf{x}_{j,k} + \epsilon_{i,j} \tag{8}$$

[21]We compute wealth differences from an individual-level wealth index constructed with a principal component analysis of respondents' availability of food, water, healthcare, and income. As noted above, the resulting measure is closely related to the Gini coefficient.

[22]Logistic regression models yield equivalent results. See Table A10.

where dyads are constructed only among individuals observed in the same country c at time t. $\alpha_{c,t}$ is a fixed effect for each country c at each time t, capturing general factors that affect party system fragmentation and the rate of co-ethnicity, such as states' population size, history, or electoral institutions. β_1 now captures the *average* effect of co-ethnicity by mother tongue on co-voting intentions among individuals across all countries and times in the sample, conditional on the covariates $\mathbf{x}$ introduced and visualized in Figure 3 above. Since all socio-economic factors underlying $\mathbf{x}_{i,j}$ can plausibly be causes *and* consequences of respondents' ethnic identity, adding these controls could trigger post-treatment bias. At the risk of omitted variable bias, we therefore first estimate a baseline model without any controls. We then compare the estimates across the baseline and the fully specified model with controls. Although it is possible that the potential biases from omitted variables and post-treatment controls lead us to overestimate the effect of co-ethnicity in both cases, we argue that small differences between the two specifications should reassure us that we approximate the true effect of co-ethnicity.

By construction, the CVR model is estimated on interdependent data. We thus consider various strategies to adjust standard errors. In our main analyses, we rely on the conservative two-way clustering on the ethnicity of individuals i and j that constitute each dyad. These clusters correspond to the level of "treatment assignment" if we consider ethnic groups and their members to be jointly treated as groups. The resulting confidence intervals are as large as clustering on the level of entire countries. They are also significantly more conservative than clustering on the level of individuals or their enumeration area. Lastly, Appendix Figure A5 shows that employing Aronow et al.'s (2015) cluster-robust variance estimator for dyadic data at the level of individuals, their ethnicity, or their locations of residence leads to even smaller uncertainty estimates.

Beyond its effects on uncertainty estimates, unit-interdependence may bias point estimates in our setting. An extensive robustness analysis shows equivalent results when using a *Probabilistic Partition Model* recently developed by Müller-Crepon, Schvitz and Cederman (2023) that explicitly models unit-interdependence. Discussed in more detail below, we adapt the model to estimate the effect of our dyadic cleavage indicators on the partitioning of voters into parties and find results that closely coincide with our main estimates which have the crucial advantage of simplicity in implementation and interpretation.

Results

Our analysis yields strong support of the hypothesis that co-ethnicity increases the rate of shared voting intentions and party preferences among among Afrobarometer respondents. Table 1 presents our main estimates, showing the unconditional and conditional effect of co-ethnicity on respondents' co-voting intentions, measured as shared voting intentions and preferences for parties. We find that pairs of respondents who share their mother tongue are between 16.1 and 17.4 percentage points more likely to have the same voting intention and party preference. The effect is very stable across specifications, does not vary between our two outcomes, and is associated with little uncertainty ($p < .001$).

Substantively, these effects are large. We observe shared co-voting intentions in 46 percent of all survey respondent dyads. The conditional increase in co-voting intentions resulting from shared mother-tongues of 16 percentage points (Models 2 and 4) thus amounts to 35 percent of the mean rate of co-voting preferences. The effect of co-ethnicity also swamps the effect of any other pairwise similarity between respondents, the substantively largest being that of shared occupation with an effect of 3.1 percentage points. We will return to a more thorough comparison of the effect of co-ethnicity with other cleavage dimensions below.

According to our model, a marginal increase in ethnic homogeneity at the macro-level translates to a marginal increase in the concentration of candidates or parties at a proportion of 1 to .16 (see also Equation 7).[23] This positive elasticity stands in drastic contrast to the *negative* bivariate relationship observed when using country-level data (see Appendix A2),[24] highlighting the caveats of ecological inferences drawn from aggregate data.

We observe little systematic change in the aggregate effect of co-ethnicity on

[23]The elasticity of the effective number of parties to changes in ethnic homogeneity depends on the value of other covariates, in particular the country fixed effects and the country-level of ethnic homogeneity.

[24]The bivariate relation contrasts the generally positive relationship estimated by Clark and Golder (2006).

Table 1: Co-voting and shared mother tongue

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.174***	0.165***	0.170***	0.161***
	(0.031)	(0.030)	(0.033)	(0.033)
Shared religion (0/1)		0.023**		0.021**
		(0.008)		(0.008)
Age similarity (decades)		-0.004***		-0.003*
		(0.001)		(0.001)
Shared gender (0/1)		-0.001		-0.003**
		(0.001)		(0.001)
Shared education (0/1)		0.022***		0.019***
		(0.004)		(0.004)
Wealth similarity (sd)		0.003*		0.002
		(0.001)		(0.002)
Shared occupation (0/1)		0.031***		0.031***
		(0.006)		(0.007)
Geographic proximity (100km)		0.046*		0.039*
		(0.018)		(0.018)
Shared urban vs. rural (0/1)		0.021***		0.019***
		(0.005)		(0.005)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.458	0.458	0.458	0.458
Countries	26	26	28	28
Respondents	16,833	16,833	12,989	12,989
Dyads	5,805,059	5,805,059	3,318,087	3,318,087
R^2	0.075	0.078	0.074	0.076
Within R^2	0.019	0.022	0.018	0.021

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

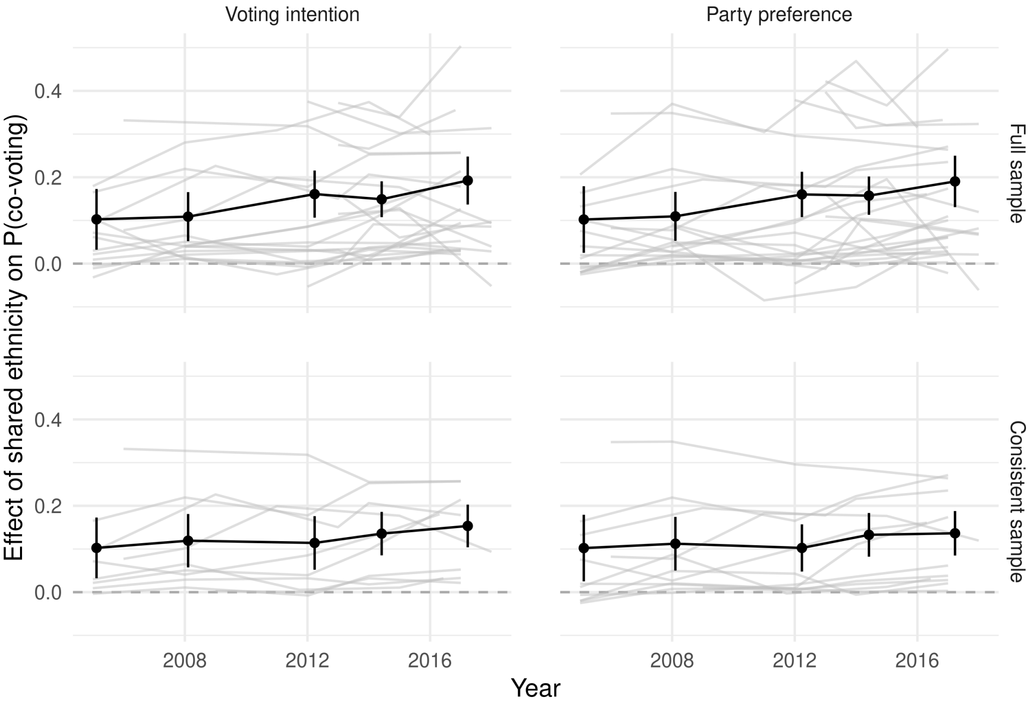


Figure 4: Effect over time, by Afrobarometer survey round

Note: Coefficients result from the fully specified model in Equation 8 estimated separately for each Afrobarometer survey round and using respondents' self-identified ethnicity to construct the co-ethnicity indicator. 'Full sample' refers to all countries included in any one survey round, while 'consistent sample' refers to countries included since Afrobarometer Round 3. Grey lines plot country-by-country estimates over time, see Appendix Figures A6 to A8 for full results.

co-voting preferences over time. When repeating our analysis for Afrobarometer rounds 3 to 7 in Figure 4 we find a slight upwards trend in the full sample, which includes increasingly many countries.[25] Yet, there is no significant increase in the effect of co-ethnicity once we subset the sample to countries that have always been surveyed. In other words, the upwards slope observed in the upper panels in Figure 4 results mostly from the increasing sampling of countries with more extensive ethnic voting.

In contrast, the estimates of ethnic co-voting intentions vary within countries over time. We discuss three cases that feature prominently in previous studies, sometimes as examples that demonstrate the weakness or even absence of ethnic voting preferences. Figure 5 displays the estimated effect of co-ethnicity on co-voting intentions across the Afrobarometer rounds 3-7 for Kenya, Malawi, and

[25]For this inter-temporal analysis, we construct the co-ethnicity indicator based on respondents' self-identified ethnicity, which has been consistently asked since round 3.

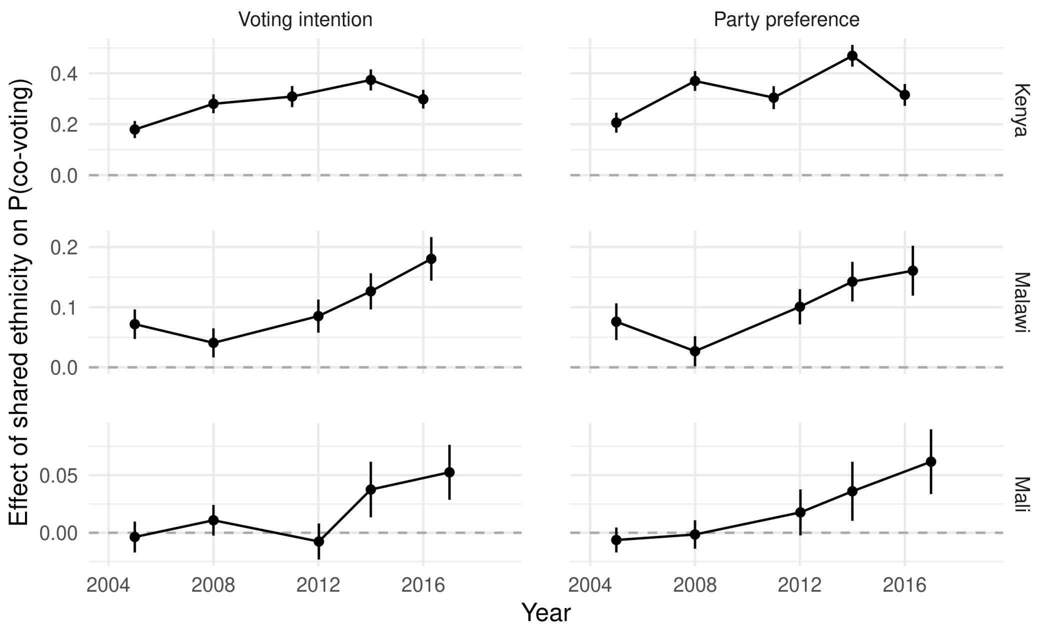


Figure 5: Ethnic voting over time in Kenya, Malawi and Mali

Note: Coefficients result from the fully specified model in Eq. 8 estimated separately for each Afrobarometer survey round since Round 3 with respondents' self-identified ethnicity as ethnicity indicator. See Appendix Figures A6 to A8 for all countries in the sample.

Mali. Political scientists typically describe Kenyan elections as classic cases of ethnic voting (Bratton and Kimenyi 2008). While Ferree (2022) recently showed that this evaluation might arise because many Kenyans have no other choice but to vote for co-ethnics, our analysis confirms the earlier interpretation. The strength of the shared ethnicity coefficient reaches twice the estimated effect from our full sample in six out of ten survey round-question combinations. Over time, ethnic co-voting preferences have increased on average, with a small decline in 2016. Even though parties and candidates in Kenya may target swing voters that have no co-ethnic on the ballot (Horowitz 2022, 6), the vast majority of Kenyans intends to vote with their co-ethnics.

Next, we turn to Malawi and Mali, two countries for which prominent studies diagnosed weak ethnic voting patterns. In line with Ferree and Horowitz' analysis of Malawi, we indeed observe near-zero coefficients for shared ethnicity in the run-up to the 2009 election, in which the ethno-regional voting "pattern broke down in dramatic fashion" (2010, 535). However, since then our estimates indicate a strengthening of co-ethnic voting intentions that reach the estimated average effect for all elections in SSA in the latest Afrobarometer round. This observation challenges recent work that identifies Malawi's persistent regional voting blocs and underlying shared economic interests as the better fitting explanation of co-voting intentions (Boone et al. 2022). Yet Boone et al.'s meso-level analysis imposes geographic blocs as the main cleavage. In contrast, our dyadic approach at the micro-level allows us to estimate the relative influence of multiple cleavages without favoring one over another. Finally, in a widely-cited study, Dunning and Harrison (2010, 21) "help explain why ethnicity has a relatively minor role in Mali ... [a] country in which ethnic identity is a poor predictor of vote choice." Our analysis confirms Dunning and Harrisons's verdict when they wrote their study in the late 2000s. More recently however, shared ethnicity has gained prominence in Malian citizens' voting preferences, a trend that underlines the importance of broad comparative work like ours.

Robustness checks

We systematically test the robustness of our results to the measure of ethnicity used, accounting for the potentially biasing effect of geography, changes in the setup of the data and estimation, as well as to using an alternative statistical network estimator. Our discussion below summarizes the results presented in Appendix B and C.

Accounting for potentially endogenous ethnicity: As discussed above, one threat to inference consists in endogenous ethnic change or identity misreporting among respondents. For example, minority members might be incentivized to report membership in a powerful majority group (Green 2021) or economic incentives might shape political preferences and ethnic groups in parallel (Pengl, Roessler and Rueda 2022). In Appendix B.1, we implement two strategies to gauge in how far such processes can explain our main findings.

First, we leverage differences in the malleability of different ethnic markers. Beyond respondents mother tongue, interviewers in Afrobarometer round 7 also asked respondents about (a) the language spoken in their homes *now* and (b) their "ethnic community, cultural group, or tribe". Both are more malleable than reported mother tongues, with the current language at home being most susceptible to change and strategic reporting. Yet, in particular the ethnicity item is also more precise in reflecting their current ethnic identification than information about individuals' mother tongue, thus reducing measurement error and related downward bias. Re-estimating the CVR using these two variables to construct the indicator of pairwise co-ethnicity, we find the smallest effect (12 percentage point) for shared language spoken at home. The more precise indicator of shared ethnicity has a slightly larger effect (19 percentage points) than our baseline specification.

Our second strategy draws on the assumption that misreporting and assimilation is least likely to affect pairs of respondents with very distinct and linguistically unrelated mother tongues. We thus estimate the effect of the pairwise linguistic proximity between respondents and find that shared co-voting intentions are least likely among respondents who grew up speaking unrelated languages.[26] In combination, these results suggest that strategic misreporting or endogenous ethnic change are unlikely to substantively affect our results.

Accounting for geographic variation: A second threat to inference originates in the geography of ethnic groups. Many ethnic groups' inhabit regionally distinct homelands. Co-ethnic voting intentions might simply emerge from an alignment of political preferences of individuals who live in the same administrative region or even location (Boone et al. 2022; Boone 2024), an argument that dovetails with findings of non-ethnic voting of local minorities in presidential elections (Ichino and Nathan 2013). This risk is further compounded by previous findings that the drawing of subnational borders has partially shaped ethnic geography itself (Posner 2005; Müller-Crepon 2023). Yet, individuals' place of residence is not entirely exogenous either, but shaped through ethnic migration patterns (Müller-Crepon 2023, see also Marbach 2021).

We address this threat by excluding any variation between administrative regions or single localities from our data (see Appendix B.2). We do so by constructing our dyadic comparisons *after* splitting each country-round into disjoint samples

[26] We compute linguistic distance through the ethnic linkages data from Müller-Crepon, Pengl and Bormann (2022).

from (a) administrative regions and (b) enumeration areas (EAs). The resulting data then features no dyads that span across these spatial units, leaving only comparisons among respondents who live in the same region/EA. Doing so increases the rate of shared mother tongues from 20 percent in the full sample to 40 percent within regions and 61 percent within enumeration areas.[27] While decreasing in size, the estimated effects of a shared mother tongue remain sizeable even within regions (10-12ppts) and enumeration areas (6-7ppts). While the decrease speaks to more frequent alignment of electoral preferences across ethnic lines within small geographic radii, the result also shows that geographic sources of ethnic identification and vote choice do not explain our results.

Data construction: We vary a number of choices made in the construction of our dyadic comparisons between survey respondents (see Appendix B.3). We first sequentially reduce the number of comparisons to the point of leaving only one comparison per respondent. This yields stable coefficient and uncertainty estimates. Second, we account for variation in the number of dyadic comparisons per country by weighting each dyad by the inverse number of dyads from its country such that every country receives the same weight.[28] This increases coefficient estimates slightly. Lastly, we recode preferences for "other" candidates and parties such that each such response is coded as its own candidate or party instead of being dropped. Doing so does not materially change the results.

Model specification: We furthermore test the robustness of our results regarding the most important modeling decisions (see Appendix B.4). We first reestimate the main specifiations in Table 1 using logistic regression models, which yields equivalent results (Table A10). Second, we test various ways of clustering our standard errors to account for the interdependence between dyadic comparisons, which yields less conservative estimates with the exception of clustering at the country level which yields marginally more conservative uncertainty estimates. Third, we implement different fixed effect specifications to account for potential sources of bias

[27]Similarly, co-voting intentions increase from 46 percent to 51 and 59 percent within regions and enumeration areas, respectively.

[28]This imbalance results from differing rates of missingness in the data.

at the level of language groups, enumeration areas, and individual respondents on each side of a comparison. Doing so drastically improves the variation in outcomes explained by the model but does not substantively change the estimated effect of a shared mother tongue on co-voting preferences.

Employing a network-based partition model: We lastly test whether our results are consistent when modeling our data using a network-based *Probabilistic Partition Model* (Müller-Crepon, Schvitz and Cederman 2023). The model allows us to estimate the effect of co-ethnicity on the partitioning of voters into parties while accounting for the overall dependency structure in the data as well as co-variates (see Appendix C).[29] Here, the outcome is not whether two individuals share the same candidate preference, but the *set of* individuals with which a respondents shares her preference. As reported in Appendix C, the results closely align with our baseline findings. Co-ethnic ties between voters increase the likelihood that they together vote for the same party (a partition in the notation of the model) in a substantive and statistically significant manner. The effect is also consistently larger than that of the other predictors with a ratio comparable to that found in our main analysis.

(No) Heterogeneous effects

Prior research highlights theoretical reasons to expect substantive variation in the extent to which ethnic cleavages structure the menu of parties and candidates as well as voting (e.g. Huber 2012; Mozaffar, Scarritt and Galaich 2003). We analyze such heterogeneity along electoral systems, countries' level of democracy, and the strength of traditional institutions. We do not find substantive or statistically significant variation in the effect of co-ethnicity on co-voting intentions across these three arguably important institutional dimensions. We consider two reasons for the discrepancy with the existing literature. First, our findings can only be understood descriptively as we do not account for potential endogeneity of the moderating factors. Second, it is possible that existing analyses find empirical support for theoretical mechanisms that only operate at the meso and macro-levels. Ferree (2022)

[29] Since the sampler underlying the parametric bootstrap proposed by Müller-Crepon, Schvitz and Cederman (2023) yields unstable result for our fully connected network data, we cluster standard errors through a non-parametric country-level bootstrap.

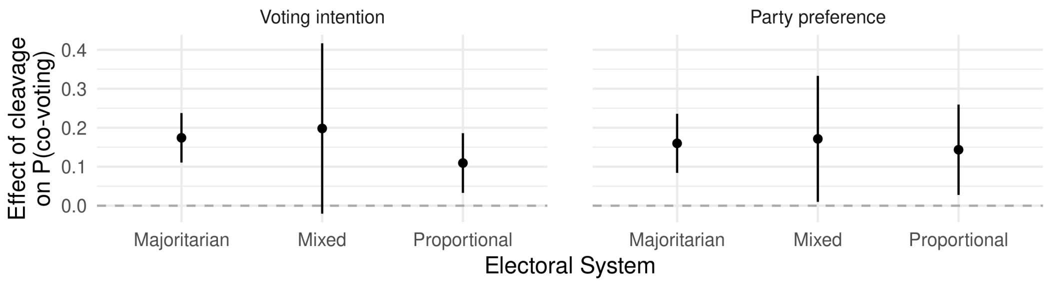


Figure 6: Heterogeneity by countries' electoral system

argues for a separation between demand (voter) and supply-side (candidate, party) factors in the investigation of ethnic voting. We find it plausible that much of the institutional variation in the strength of ethnic voting could be explained by candidates' perception of when and where to run for office, rather than by individuals' voting preferences.

Electoral system: A large literature suggests that proportional electoral systems politicize ethnic identities as particularistic parties face few obstacles to representation and may even join governing coalitions (e.g., Lijphart 2004). In contrast, some majoritarian electoral rules arguably incentivize cross-ethnic mobilization (e.g., Posner 2005). Using data on electoral systems from Bormann and Golder (2022), our results in Figure 6 suggest that there are no large or statistically significant difference in the effect of co-ethnicity on co-voting intentions. While PR systems see slightly less shared support for presidential candidates among co-ethnics, this difference is not statistically significant.

Democracy: Democratic institutions, specifically competitive elections, are frequently associated with ethnic mobilization (Rabushka and Shepsle 1972; Horowitz 1985; Eifert, Miguel and Posner 2010). As described above, political leaders seek to mobilize majorities through clientelism and patronage, which often follows ethnic lines (Bates 1974), and might reinforce the effect of cultural differences on diverging policy preferences (Lieberman and McClendon 2013). While elections within dictatorships might also follow a clientelist logic, they are less likely to reveal divergent policy preferences. Countering these considerations, we find that the effect

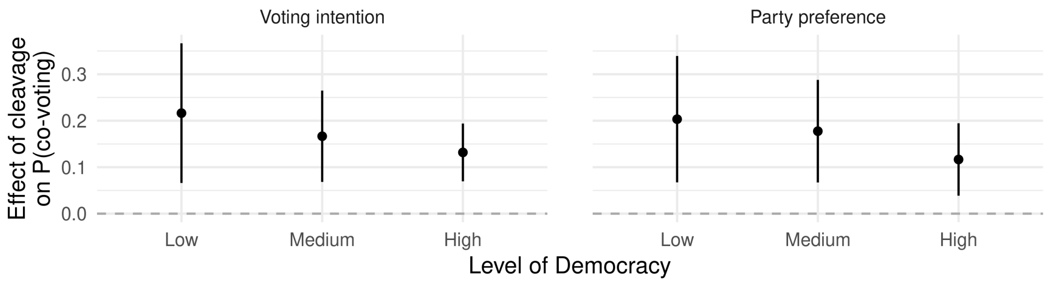


Figure 7: Heterogeneity by countries' level of democracy

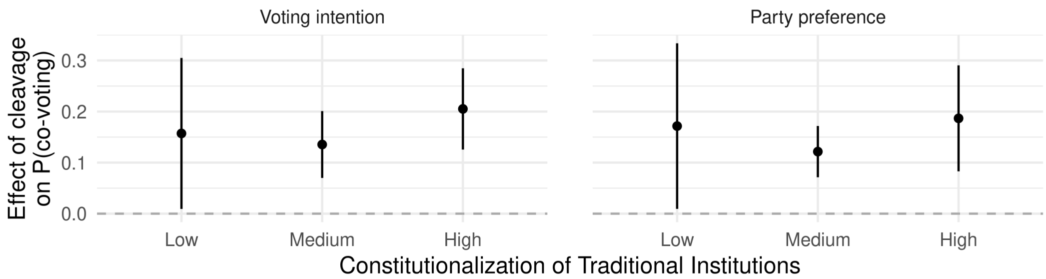


Figure 8: Heterogeneity by country's constitutionalization of traditional institutions

of co-ethnicity on co-voting preferences is, if at all, smaller in countries with higher levels of democracy measured via V-DEM's polyarchy index (Coppedge et al. 2016, see Figure 7). Again, the differences we observe are not statistically significant.

Traditional institutions may co-produce local public goods (Baldwin 2016), and act as complements to the state where they are institutionally tied to it (Holzinger et al. 2019; Henn 2022). Therefore, voters have incentives to vote "with their chief" (Baldwin 2013, see also De Kadt and Larreguy 2018). As a result of the entanglement between traditional authorities and ethnic identities, one might expect strong traditional institutions to come with stronger effects of co-ethnicity on co-voting intentions. Using data on the constitutionalization of traditional authorities from (Holzinger et al. 2019), Figure 8 shows relatively little and no statistically significant variation in the effect of co-ethnicity on co-voting preferences.

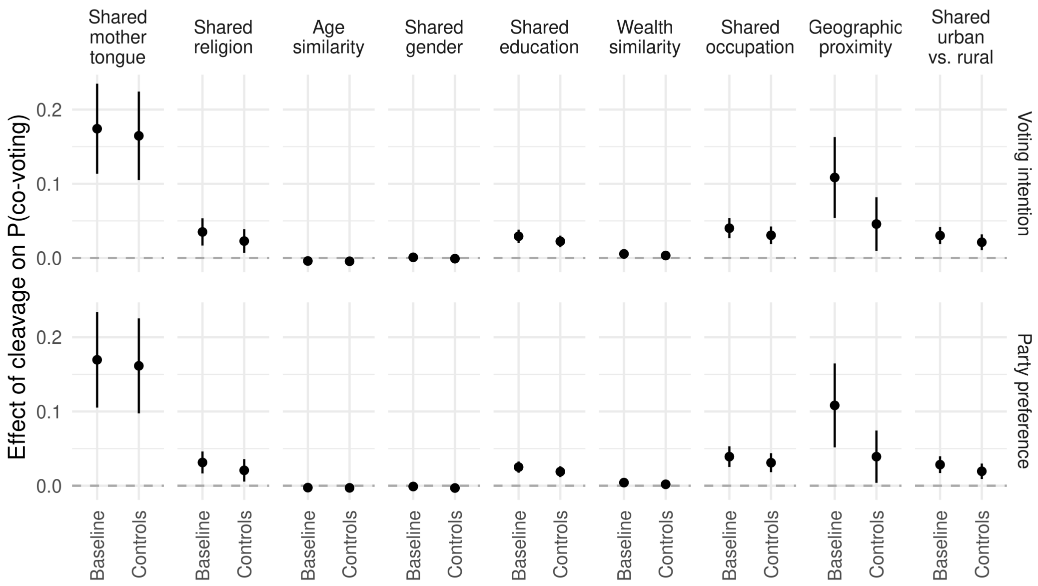


Figure 9: Results by cleavage indicator

Note: Coefficient estimates from (1) baseline model that only include the respective variable and country-fixed effects, and (2) fully specified models with controls (Eq. 8). Error bars denote 95% CIs.

Comparing cleavages

Finally, we compare the effect of co-ethnicity and that of other socio-economic similarities between respondents with respect to their rate of co-voting intentions. To facilitate a fair comparison that takes account of differences between conditional and unconditional effects, Figure 9 plots the results of baseline models of the effect of each variable without any additional controls, as well as coefficient estimates from the fully specified models (see Table 1, Models 2 and 4).

Among identity cleavages, shared mother-tongues seem to be by far the strongest and most stable predictor of co-voting preferences. The first column in Figure 9 depicts our main results from Table 1. Next, effects associated with shared religion are positive but decrease once we condition on covariates. We presume that the unconditional effect of shared religion captures some of the effect of (correlated) sharing of mother-tongues. Across stated support for presidential candidates and legislative parties, we find no substantive effects of age and gender similarities. For age, we find a small *negative* effect of being close in age on co-voting intentions, which suggests that party preferences within age groups are marginally more diverse than across them.

Economic similarities show some but substantively smaller effects on convergent voting intentions than those associated with shared mother tongues. Shared levels of education and occupation between respondents translate into an increase in the chance of supporting the same party by about 2 and 3.4 percentage points, respectively. These effects are robustly estimated. Interestingly, proximity in wealth-levels between respondents does not relate significantly to co-voting intentions between them. This finding speaks to previous findings on economic voting in Sub-Sahara Africa (Bratton and Kimenyi 2008; Bratton, Bhavnani and Chen 2012).

Lastly, we find geographic proximity to correlate with shared support for presidential candidates and parties. In the unconditional baseline models, increasing geographic proximity by 1'000km comes with an increase in co-voting intentions by 10 percentage points, consistent with the existence of regional voting blocs (Boone et al. 2022).[30] Yet, once we condition on all other cleavage measures, the effect of geographic proximity drops by about 50%. This decrease supports the interpretation that geography correlates with voting preferences because of its reflection of economic incentives and ethnic identities. Shared urban or rural status has a consistent and statistically significant effect on co-voting preferences of approximately 2 percentage points when including controls. This is consistent with literature on rural-urban cleavages on the continent (e.g. Harding 2010; Koter 2013).

Conclusion

In this paper, we introduce the Co-Voting Regression Model (CVR) as a novel analytical approach to study the electoral effects of social cleavages in general and ethnic voting in Africa in particular. Shifting from individual support for co-ethnic candidates towards shared voting intentions between two individuals allows us to address two key methodological weaknesses in existing work. For one, we avoid selection bias that plagues micro-level studies when the supply of candidates does not allow survey respondents to express support for co-ethnic candidates, or forces them to do so in the absence of non-ethnic rival candidates (e.g., Ferree 2022). For another, we avoid ecological inference inherent in meso and macro-level research

[30] Though note that our sample includes many small countries where such a change is unrealistic.

that examines ethnic co-voting preferences but fixes ethnic groups as the main unit of analysis while disregarding other cleavages. Coincidentally, we retain the advantages of micro and macro studies. The CVR model we introduce captures both individual-level effects, *and* recovers country-wide concentration indices such as the effective number of parties and the Herfindahl-Hirschman index of ethnic concentration. Finally, CVR operates at scale and enables broad cross-country comparisons without sacrificing country-specific insights.

Our empirical analysis of 28 countries and five survey rounds from the Afrobarometer indicates that language-based ethnicity continues to be the dominant electoral cleavage across Sub-Saharan Africa. The effect of co-ethnicity on voting intentions is at least five times larger than alternative cleavages including religion, shared urban or rural residence, geographic regions, as well as educational and occupational background. Although we find that co-ethnicity does not influence co-voting preferences equally across all survey rounds and countries, prominent case studies that question the effect of ethnicity in vote choices describe exceptions rather than broader trends across the African continent (Dunning and Harrison 2010; Ferree and Horowitz 2010; Boone et al. 2022). Finally, our analysis reveals little support for factors that moderate the strength of co-ethnic voting intentions, such as the level of democracy, electoral rules, and traditional authorities (Rabushka and Shepsle 1972; Huber 2012; Baldwin 2013).

Our study opens up new avenues for the study of ethnic and more generally cleavage-based voting in Sub-Saharan Africa and beyond. More precisely measured data on economic income and partisanship would enable us to gain much deeper insight into class and psychological explanations of voting – two core concerns of voting research outside Africa. While economic-instrumentalist and psychological factors have already received much attention in the study of ethnic voting, one major theory of voting, its sociological basis (Lazarsfeld, Berelson and Gaudet 1968/1944), has been widely overlooked by students of Sub-Saharan Africa (though see work on traditional institutions and norms, Baldwin 2013; Holzinger et al. 2019). Given appropriate data, the CVR model can easily test the effect of different social networks on voting by capturing the overlap in (the homogeneity of) social contacts.[31]

Beyond Sub-Saharan Africa, our analytical focus on co-voting intentions lends itself to the study of the relative strength of different cleavages, such as the re-emergence of urban-rural divides (Cramer 2016), and the increasingly dominant nationalist-cosmopolitan division across western democracies (Kriesi et al. 2012). Lastly, our method might also benefit existing meso-level analyses of vote shares in small-scale spatial units, such as municipalities (Cagé and Piketty 2023). These analyses face similar challenges as the ones we discussed in the context of research on voting intentions in Sub-Saharan Africa. Rather than estimating the likelihood of co-voting intentions at the individual-level, we would require compositional similarity scores between spatial units in terms of voting results as a function of similarities in their social structure. After all, social and political cleavages are an inherently *relational* concept and should be operationalized and studied as such.

[31]Spillover experiments constitute an attractive but more costly alternative methodology (e.g., Foos and De Rooij 2017).

References

Adida, Claire, Jessica Gottlieb, Eric Kramon and Gwyneth McClendon. 2017. "Overcoming or reinforcing coethnic preferences? an experiment on information and ethnic voting." *Quarterly Journal of Political Science* 12(4):437–477.

Alesina, Alberto, Reza Baqir and William Easterly. 1999. "Public goods and ethnic divisions." *Quarterly Journal of Economics* 114(4):1243–1284.

Aronow, Peter M, Cyrus Samii and Valentina A Assenova. 2015. "Cluster–robust variance estimation for dyadic data." *Political analysis* 23(4):564–577.

Baldwin, Kate. 2013. "Why Vote with the Chief? Political Connections and Public Goods Provision in Zambia." *American Journal of Political Science* 57(4):794–809.

Baldwin, Kate. 2016. *The paradox of traditional chiefs in democratic Africa*. Cambridge University Press.

Bates, Robert H. 1974. "Ethnic Competition and Modernization in Contemporary Africa." *Comparative Political Studies* 6(4):457–484.

Bayart, Jean-François. 2009. *The State in Africa: The Politics of the Belly*. 2nd ed. Cambridge, UK: Polity Press. 64.

Boone, Catherine. 2024. *Inequality and political cleavage in Africa: regionalism by design*. Cambridge University Press.

Boone, Catherine, Michael Wahman, Stephan Kyburz and Andrew Linke. 2022. "Regional Cleavages in African Politics: Persistent Electoral Blocs and Territorial Oppositions." *Political Geography* 99:102741.

Bormann, Nils-Christian and Matt Golder. 2022. "Democratic Electoral Systems around the world, 1946–2020." *Electoral Studies* 78:102487.

Bratton, Michael and Mwangi S Kimenyi. 2008. "Voting in Kenya: Putting Ethnicity in Perspective." *Journal of Eastern African Studies* 2(2):272–289.

Bratton, Michael, Ravi Bhavnani and Tse-Hsin Chen. 2012. "Voting Intentions in Africa: Ethnic, Economic or Partisan?" *Commonwealth & Comparative Politics* 50(1):27–52.

Cagé, Julia and Thomas Piketty. 2023. *Une Histoire du Conflit Politique: Elections et Inégalités Sociales en France, 1789-2022*. Seuil.

Carlson, Elizabeth. 2015. "Ethnic Voting and Accountability in Africa: A Choice Experiment in Uganda." *World Politics* 67(2):353–385.

Carlson, Elizabeth. 2018. "The Relevance of Relative Distribution: Favoritism, Information, and Vote Choice in Africa." *Comparative Political Studies* 51(12):1531–1562.

Casey, Katherine. 2015. "Crossing Party Lines: The Effects of Information on Redistributive Politics." *American Economic Review* 105(8):2410–2448.

Clark, William Roberts and Matt Golder. 2006. "Rehabilitating Duverger's Theory: Testing the Mechanical and Strategic Modifying Effects of Electoral Laws." *Comparative Political Studies* 39:679–708.

Conroy-Krutz, Jeffrey. 2013. "Information and Ethnic Politics in Africa." *British Journal of Political Science* 43(2):345–373.

Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M Steven Fish, Adam Glynn, Allen Hicken et al. 2016. "V-Dem Dataset.".

Cramer, Katherine J. 2016. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker*. Chicago, Il: University of Chicago Press.

De Kadt, Daniel and Horacio A Larreguy. 2018. "Agents of the regime? Traditional leaders and electoral politics in South Africa." *The Journal of Politics* 80(2):382–399.

Dunning, Thad and Lauren Harrison. 2010. "Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in Mali." *American Political Science Review* 104(1):21–39.

Eifert, Benn, Edward Miguel and Daniel N. Posner. 2010. "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science* 54(2):494–510.

Elischer, Sebastian. 2013. *Political Parties in Africa – Ethnicity and Party Formation*. Cambridge University Press.

Ferree, Karen E. 2006. "Explaining South Africa's Racial Census." *Journal of Politics* 68(4):803–815.

Ferree, Karen E. 2010. "The Social Origins of Electoral Volatility in Africa." *British Journal of Political Science* 40(4):759–779.

Ferree, Karen E. 2022. "Choice and Choice Set in African Elections." *The Journal of Politics* 84(4):2261–2265.

Ferree, Karen E, Clark C Gibson and James D Long. 2021. "Mixed Records, Complexity, and Ethnic Voting in African Elections." *World Development* 141:105418.

Ferree, Karen and Jeremy Horowitz. 2010. "Ties that Bind? The Rise and Decline of Ethno-Regional Partisanship in Malawi, 1994–2009." 17(3):534–563.

Foos, Florian and Eline A De Rooij. 2017. "All in the Family: Partisan Disagreement and Electoral Mobilization in Intimate Networks–A Spillover Experiment." *American Journal of Political Science* 61(2):289–304.

Green, Elliott. 2021. "The Politics of Ethnic Identity in Sub-Saharan Africa." *Comparative Political Studies* 54(7):1197–1226.

Harding, Robin. 2010. *Urban-rural differences in support for incumbents across Africa*. Institute for Democracy in South Africa.

Harding, Robin. 2020. "Who is Democracy Good For? Elections, Rural Bias, and Health and Education Outcomes in Sub-Saharan Africa." *The Journal of Politics* 82(1):241–254.

Henn, Soeren J. 2022. "Complements or Substitutes? How Institutional Arrangements Bind Traditional Authorities and the State in Africa." *American Political Science Review* pp. 1–20.

Hoffman, Barak D and James D Long. 2013. "Parties, Ethnicity, and Voting in African Elections." *Comparative Politics* 45(2):127–146.

Holzinger, Katharina, Roos Haer, Axel Bayer, Daniela M Behr and Clara Neupert-Wentz. 2019. "The constitutionalization of indigenous group rights, traditional political institutions, and customary law." *Comparative Political Studies* 52(12):1775–1809.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkely: University of California Press.

Horowitz, Jeremy. 2022. *Multiethnic Democracy: The Logic of Elections and Policymaking in Kenya*. Oxford University Press: New York, NY.

Houle, Christian, Chunho Park and Paul D Kenny. 2019. "The Structure of Ethnic Inequality and Ethnic Voting." *The Journal of Politics* 81(1):187–200.

Huber, John D. 2012. "Measuring Ethnic Voting: Do Proportional Electoral Laws Politicize Ethnicity?" *American Journal of Political Science* 56(4):986–1001.

Huber, John D and Pavithra Suryanarayan. 2016. "Ethnic inequality and the ethnification of political parties: Evidence from India." *World Politics* 68(1):149–188.

Ichino, Nahomi and Noah L. Nathan. 2013. "Crossing the Line: Local Ethnic Geography and Voting in Ghana." *American Political Science Review* 107(2):344–61.

Ishiyama, John. 2012. "Explaining ethnic bloc voting in Africa." *Democratization* 19(4):761–788.

Koter, Dominika. 2013. "Urban and rural voting patterns in Senegal: the spatial aspects of incumbency, c. 1978–2012." *The Journal of Modern African Studies* 51(4):653–679.

Koter, Dominika. 2016. *Beyond ethnic politics in Africa*. Cambridge University Press.

Kriesi, Hanspeter, Edgar Grande, Martin Dolezal, Marc Helbling, Dominic Höglinger, Swen Hutter and Bruno Wüest. 2012. *Political Conflict in Western Europe*. New York, NY: Cambridge University Press.

Laakso, Markku and Rein Taagepera. 1979. ""Effective" number of parties: a measure with application to West Europe." *Comparative political studies* 12(1):3–27.

Laitin, David D. and A. Maurits van der Veen. 2012. Ethnicity and Pork: A Virtual Test of a Causal Mechanism. In *Constructivist Theories of Ethnic Polictics*, ed. Kanchan Chandra. New York, NY: Oxford University Press pp. 341–358.

Lazarsfeld, Paul F, Bernard Berelson and Hazel Gaudet. 1968/1944. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press.

Lieberman, Evan S. and Gwyneth H. McClendon. 2013. "The Ethnicity–Policy Preference Link in Sub-Saharan Africa." *Comparative Political Studies* 46(5):574–602.

Lijphart, Arend. 2004. "Constitutional Design for Divided Societies." *Journal of Democracy* 15(2):96–109.

Lublin, David. 2017. "Electoral Systems, Ethnic Heterogeneity and Party System Fragmentation." *British Journal of Political Science* 47(2):373–389.

Marbach, Moritz. 2021. "Causal Effects, Migration, and Legacy Studies." *American Journal of Political Science* .

McCauley, John F. 2014. "The Political Mobilization of Ethnic and Religious Identities in Africa." *American Political Science Review* 108(4):801–816.

Mor, Maayan. 2022. "Government Policies, New Voter Coalitions, and the Emergence of Ethnic Dimension in Party Systems." *World Politics* 74(1):121–166.

Mozaffar, Shaheen, James R. Scarritt and Glen Galaich. 2003. "Electoral Institutions, Ethnopolitical Cleavages, and Party Systems in Africa's Emerging Democracies." *American Political Science Review* 97(3):379–390.

Müller-Crepon, Carl. 2023. "Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa." *American Journal of Political Science, FirstView* .

Müller-Crepon, Carl, Guy Schvitz and Lars-Erik Cederman. 2023. "Shaping States into Nations: The Effects of Ethnic Geography on State Borders." *American Journal of Political Science, FirstView* .

Müller-Crepon, Carl, Yannick Pengl and Nils-Christian Bormann. 2022. "Linking Ethnic Data from Africa." *Journal of Peace Research* 59(3):425–435.

Nathan, Noah L. 2016. "Local Ethnic Geography, Expectations of Favoritism, and Voting in Urban Ghana." *Comparative Political Studies* 49(14):1896–1929.

Nathan, Noah L. 2019. *Electoral Politics and Africa's Urban Transition: Class and Ethnicity in Ghana*. Cambridge University Press.

Pengl, Yannick I, Philip Roessler and Valeria Rueda. 2022. "Cash crops, print technologies, and the politicization of ethnicity in Africa." *American Political Science Review* 116(1):181–199.

Posner, Daniel N. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi." *American Political Science Review* 98(4):529–545.

Posner, Daniel N. 2005. *Institutions and ethnic politics in Africa*. Cambridge University Press.

Rabushka, Alvin and Kenneth Shepsle. 1972. "Politics in plural societies." *Columbus, OH: Charles E. Merrill* 232.

Resnick, Danielle. 2012. "Opposition Parties and the Urban Poor in African Democracies." *Comparative Political Studies* 45(11):1351–1378.

Sen, Amartya. 1997. *On economic inequality*. Oxford university press.

Simpson, Edward H. 1949. "Measurement of diversity." *Nature* 163(4148):688–688.

Wahman, Michael and Catherine Boone. 2018. "Captured Countryside? Stability and Change in Sub-National Support for African Incumbent Parties." *Comparative Politics* 50(2):189–216.

Wooldridge, Jeffrey M. 2008. *Introductory Econometrics: A Modern Approach*. International edition of the 4th revised edition ed. South-Western.

Supplementary Material

Table of Contents

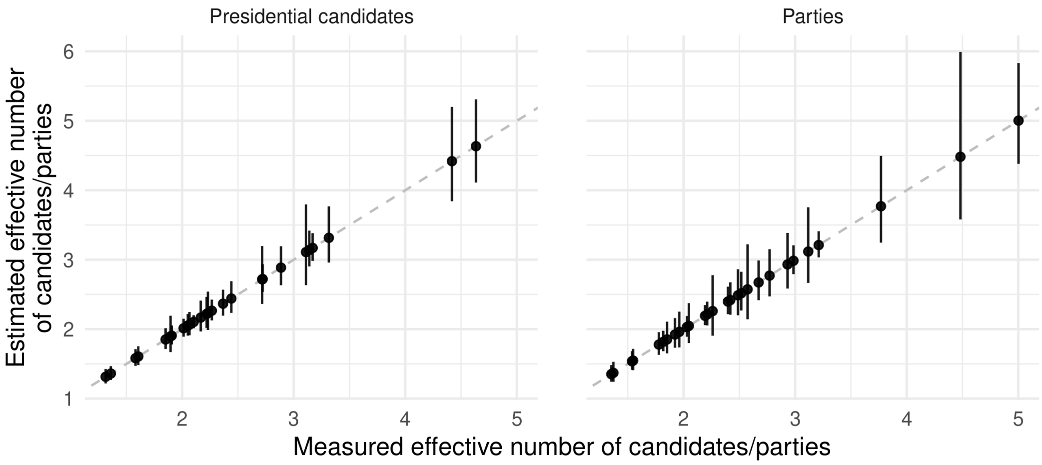
A Model setup and data

Table A1: Summary statistics of variables in presidential voting intention sample.

Statistic	N	Mean	St. Dev.	Min	Max
Co-Voting	14,311,161	0.547	0.498	0	1
Shared mother tongue	14,311,161	0.791	0.406	0	1
Shared religion	14,311,161	0.780	0.415	0	1
Shared gender	14,311,161	0.499	0.500	0	1
Shared education	14,311,161	0.619	0.486	0	1
Shared occupation	14,311,161	0.767	0.423	0	1
Shared urban vs. rural	14,311,161	0.413	0.492	0	1
Age similarity (decades)	14,311,161	1.491	1.224	0.000	8.700
Wealth similarity (sd)	14,311,161	1.506	1.172	0.000	6.147
Geographic proximity (100 km)	14,311,161	3.621	2.849	0.000	18.484

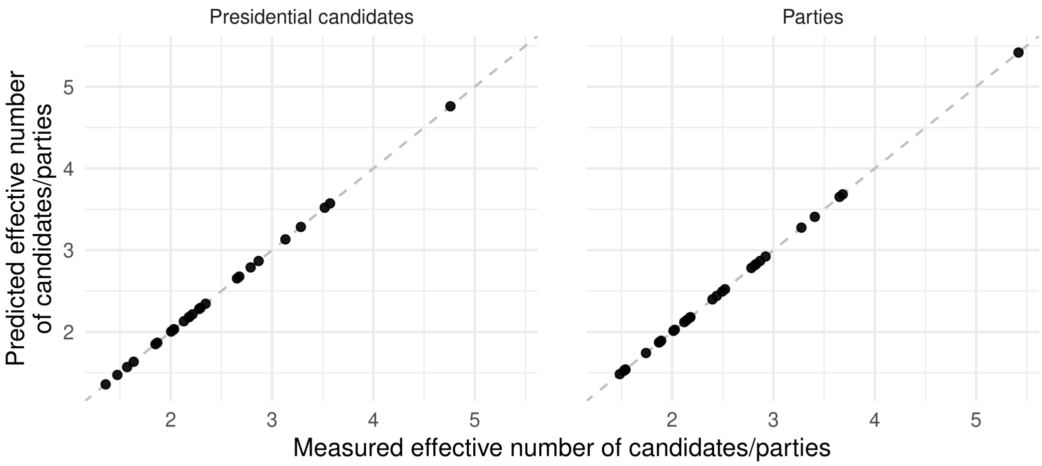
Table A2: Summary statistics of variables in party preference sample.

Statistic	N	Mean	St. Dev.	Min	Max
Co-Voting	10,678,427	0.542	0.498	0	1
Shared mother tongue	10,678,427	0.775	0.418	0	1
Shared religion	10,678,427	0.772	0.419	0	1
Shared gender	10,678,427	0.497	0.500	0	1
Shared education	10,678,427	0.618	0.486	0	1
Shared occupation	10,678,427	0.755	0.430	0	1
Shared urban vs. rural	10,678,427	0.408	0.491	0	1
Age similarity (decades)	10,678,427	1.518	1.234	0.000	8.700
Wealth similarity (sd)	10,678,427	1.511	1.171	0.000	6.147
Geographic proximity (100 km)	10,678,427	3.553	2.830	0.000	18.484



(a) Estimated and measured effective number of parties

Note: Estimates are derived as the inverse intercept of an otherwise empty regression model estimated separately for each country, standard errors are clustered on the level of individuals.



(b) Predicted and measured effective number of parties

Note: Predicted ENP is derived as the inverse average fitted probability of co-voting obtained from the fully specified regression model estimated separately for each country.

Figure A1: Empirical relation between measured Effective Number of Parties and the Linear Probability Model of co-voting

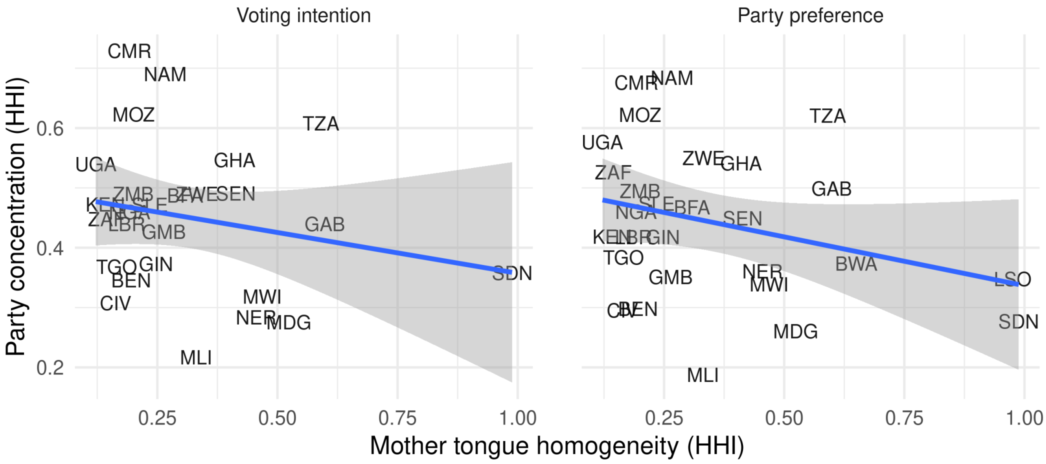


Figure A2: Bi-variate relationship between the national party system concentration and mother tongue homogeneity

Note: Both variables measured as Herfindahl-Hirschman Indices. Based on Afrobarometer Data, round 7.

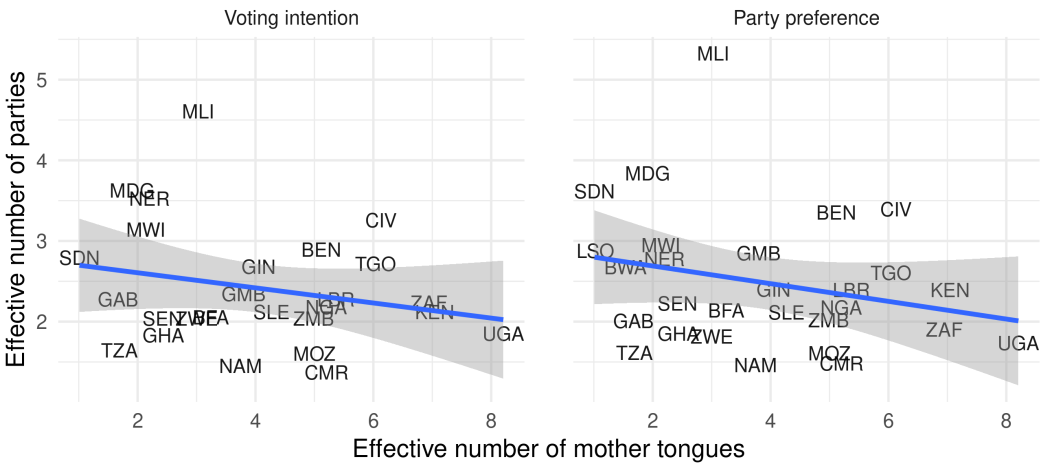


Figure A3: Bi-variate relationship between the national Effective Number of Parties and Effective Number of Mother Tongues

Note: Based on Afrobarometer Data, round 7.

B Robustness checks

B.1 Accounting for potentially endogenous ethnicity

Table A3: Co-voting intentions and shared home language

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared language (0/1)	0.133***	0.124***	0.134***	0.126***
	(0.031)	(0.029)	(0.030)	(0.029)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.457	0.457	0.456	0.456
Countries	26	26	28	28
Respondents	16,368	16,368	12,687	12,687
Dyads	5,484,046	5,484,046	3,177,242	3,177,242
R^2	0.069	0.073	0.069	0.072
Within R^2	0.014	0.017	0.014	0.017

Clustered (lang.round.to & lang.round.from) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table A4: Co-voting intentions and shared ethnicity

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared ethnicity (0/1)	0.203***	0.195***	0.198***	0.192***
	(0.031)	(0.031)	(0.032)	(0.032)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.454	0.454	0.453	0.453
Countries	26	26	28	28
Respondents	16,036	16,036	12,430	12,430
Dyads	5,267,187	5,267,187	3,049,944	3,049,944
R^2	0.081	0.083	0.080	0.082
Within R^2	0.026	0.029	0.025	0.027

Clustered (eth.round.to & eth.round.from) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table A5: Co-voting intentions and linguistic proximity

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Mother tongue proximity (0-1)	0.206***	0.104***	0.205***	0.104***
	(0.040)	(0.029)	(0.042)	(0.029)
Shared mother tongue (0/1)		0.114***		0.112***
		(0.026)		(0.028)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	Yes	Yes	Yes	Yes
Outcome mean	0.452	0.452	0.450	0.450
Countries	26	26	28	28
Respondents	15,882	15,882	12,333	12,333
Dyads	5,164,635	5,164,635	3,003,998	3,003,998
R^2	0.078	0.082	0.075	0.079
Within R^2	0.023	0.027	0.021	0.025

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

B.2 Accounting for geographic variation

Table A6: Co-voting intentions and shared mother tongue: Within Administrative Regions

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.119***	0.113***	0.111***	0.103***
	(0.018)	(0.019)	(0.020)	(0.021)
Fixed-effects				
region.to	Yes	Yes	Yes	Yes
region.from	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.513	0.513	0.516	0.516
Regions	322	322	348	348
Respondents	16,359	16,359	12,662	12,662
Dyads	839,628	839,628	502,388	502,388
R^2	0.133	0.135	0.134	0.137
Within R^2	0.012	0.014	0.010	0.013

Clustered (Language & Language) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table A7: Co-voting intentions and shared mother tongue: Within Enumeration Areas

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.078***	0.076***	0.079***	0.077***
	(0.019)	(0.019)	(0.018)	(0.018)
Fixed-effects				
enumarea	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.584	0.584	0.586	0.586
Enum. areas	2,320	2,320	1,448	1,448
Respondents	12,015	12,015	6,954	6,954
Dyads	27,012	27,012	14,080	14,080
R^2	0.361	0.362	0.381	0.383
Within R^2	0.004	0.006	0.004	0.006

Clustered (Language & Language) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

B.3 Data construction

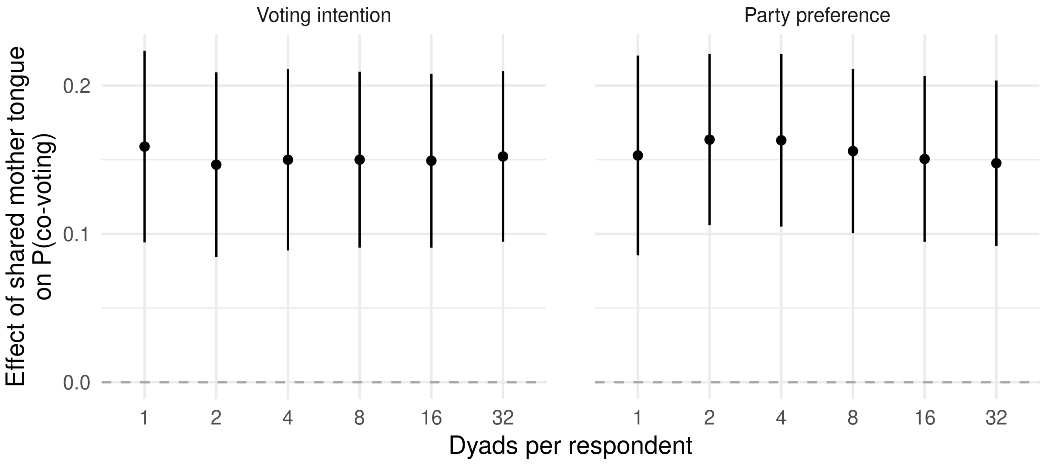


Figure A4: Effect of shared mother tongue by number of comparisons per respondent

Table A8: Co-voting intentions and shared mother tongue: Country-weights

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.184***	0.176***	0.191***	0.184***
	(0.033)	(0.033)	(0.040)	(0.041)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.458	0.458	0.458	0.458
Countries	26	26	28	28
Respondents	16,833	16,833	12,989	12,989
Dyads	5,805,059	5,805,059	3,318,087	3,318,087
R^2	0.071	0.074	0.073	0.076
Within R^2	0.021	0.024	0.023	0.026

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table A9: Co-voting intentions and shared mother tongue: Recoding 'other' parties as single parties

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.156***	0.149***	0.165***	0.156***
	(0.030)	(0.030)	(0.032)	(0.032)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.418	0.418	0.447	0.447
Countries	28	28	28	28
Respondents	17,732	17,732	13,232	13,232
Dyads	6,377,847	6,377,847	3,422,317	3,422,317
R^2	0.079	0.082	0.077	0.080
Within R^2	0.016	0.019	0.017	0.020

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

B.4 Model specification

Table A10: Co-voting intentions and shared mother tongue: Logistic regression

Dependent Variables:	Voting intention		Party preference	
Model:	(1)	(2)	(3)	(4)
Variables				
Shared mother tongue (0/1)	0.754***	0.714***	0.733***	0.699***
	(0.135)	(0.133)	(0.143)	(0.142)
Shared religion (0/1)		0.098**		0.088*
		(0.036)		(0.034)
Age similarity (decades)		-0.019***		-0.013*
		(0.006)		(0.006)
Shared gender (0/1)		-0.004		-0.013**
		(0.004)		(0.005)
Shared education (0/1)		0.098***		0.082***
		(0.017)		(0.016)
Wealth similarity (sd)		0.015*		0.008
		(0.006)		(0.007)
Shared occupation (0/1)		0.134***		0.135***
		(0.026)		(0.028)
Geographic proximity (100km)		0.207*		0.177*
		(0.082)		(0.080)
Shared urban vs. rural (0/1)		0.092***		0.084***
		(0.024)		(0.023)
Fixed-effects				
Country x Round	Yes	Yes	Yes	Yes
Fit statistics				
Controls	No	Yes	No	Yes
Outcome mean	0.458	0.458	0.458	0.458
Countries	26	26	28	28
Respondents	16,833	16,833	12,989	12,989
Dyads	5,805,059	5,805,059	3,318,087	3,318,087

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

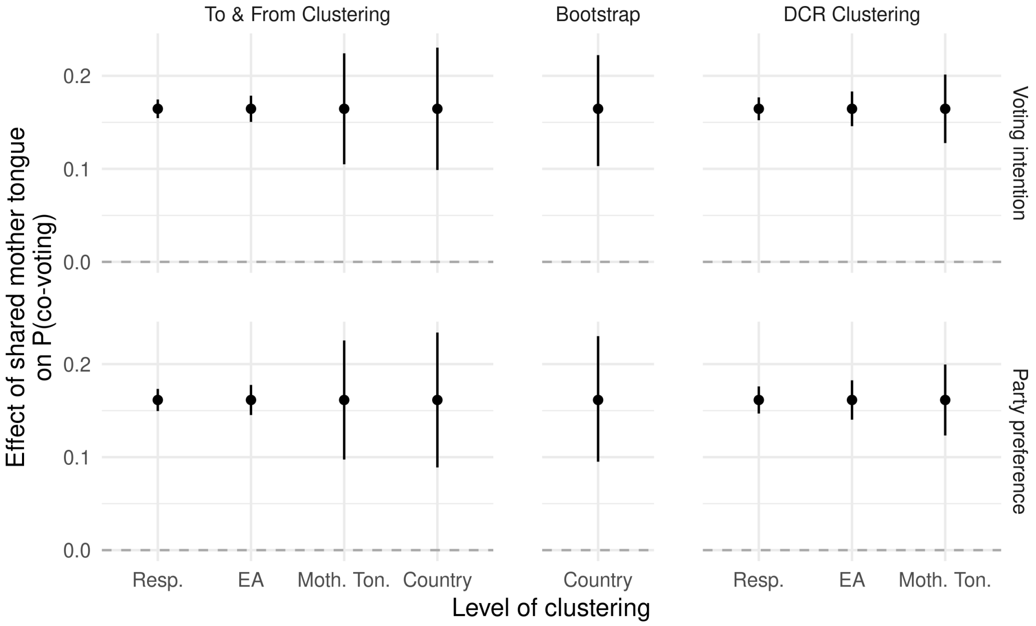


Figure A5: Varying the clustering of standard errors

Table A11: Co-voting intentions and shared mother tongue: Fixed effects specification

Dependent Variables:	Voting intention			Party preference		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Shared mother tongue (0/1)	0.162***	0.152***	0.152***	0.179***	0.162***	0.174***
	(0.032)	(0.032)	(0.033)	(0.036)	(0.034)	(0.036)
Fit statistics						
Fixed Effects	Lang.	EA	Resp.	Lang.	EA	Resp.
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	0.458	0.458	0.458	0.458	0.458	0.458
Countries	26	26	26	28	28	28
Respondents	16,833	16,833	16,833	12,989	12,989	12,989
Dyads	5,805,059	5,805,059	5,805,059	3,318,087	3,318,087	3,318,087
R^2	0.117	0.226	0.421	0.120	0.252	0.432
Within R^2	0.018	0.021	0.025	0.020	0.022	0.029

Standard errors clustered at the level of each respondent's mother tongue in parentheses.
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

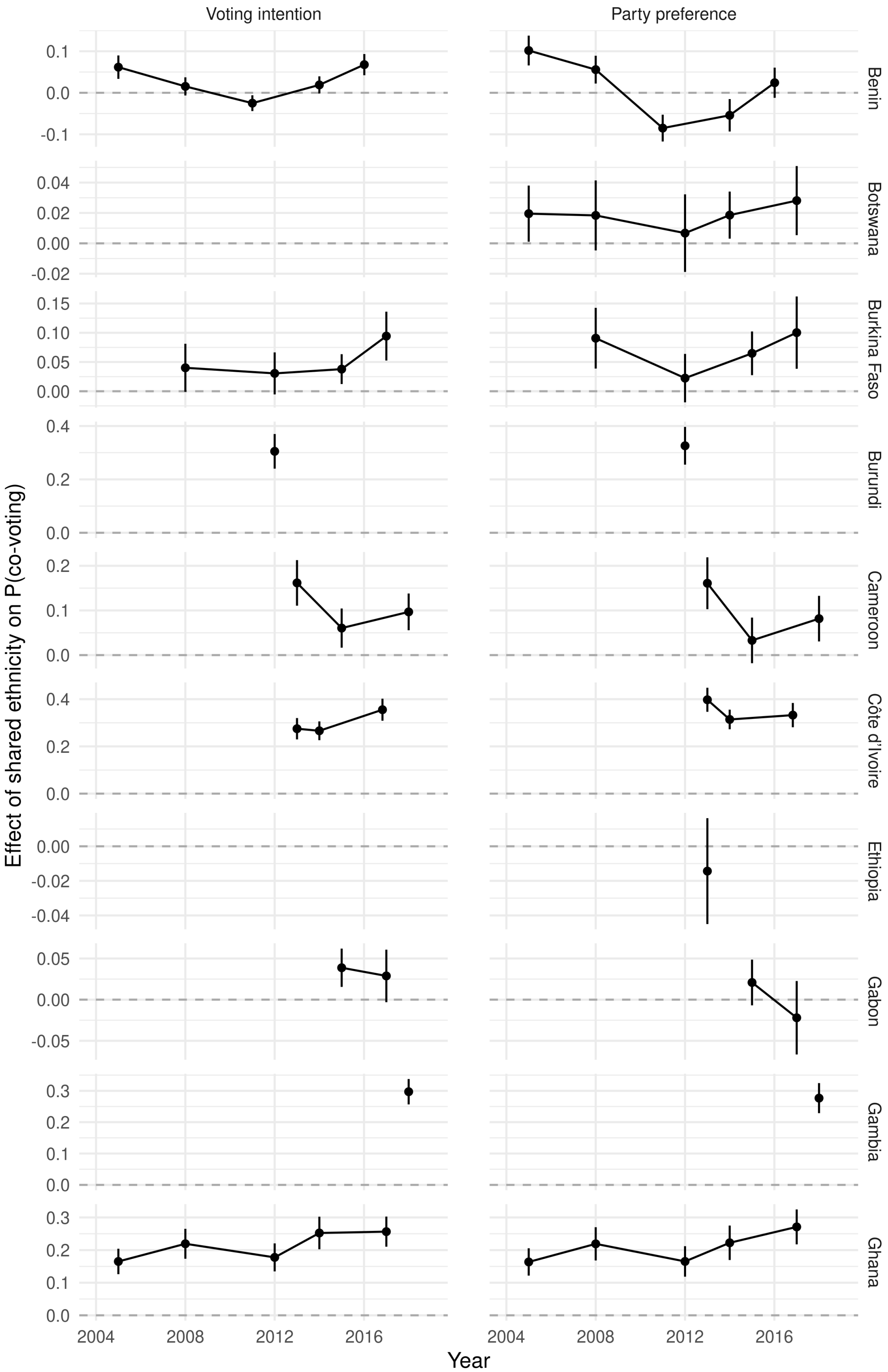


Figure A6: By country, over time, I

Note: Coefficients result from the fully specified model in Eq. 8 estimated separately for each Afrobarometer survey round since Round 3 with respondents' self-identified ethnicity as ethnicity indicator.

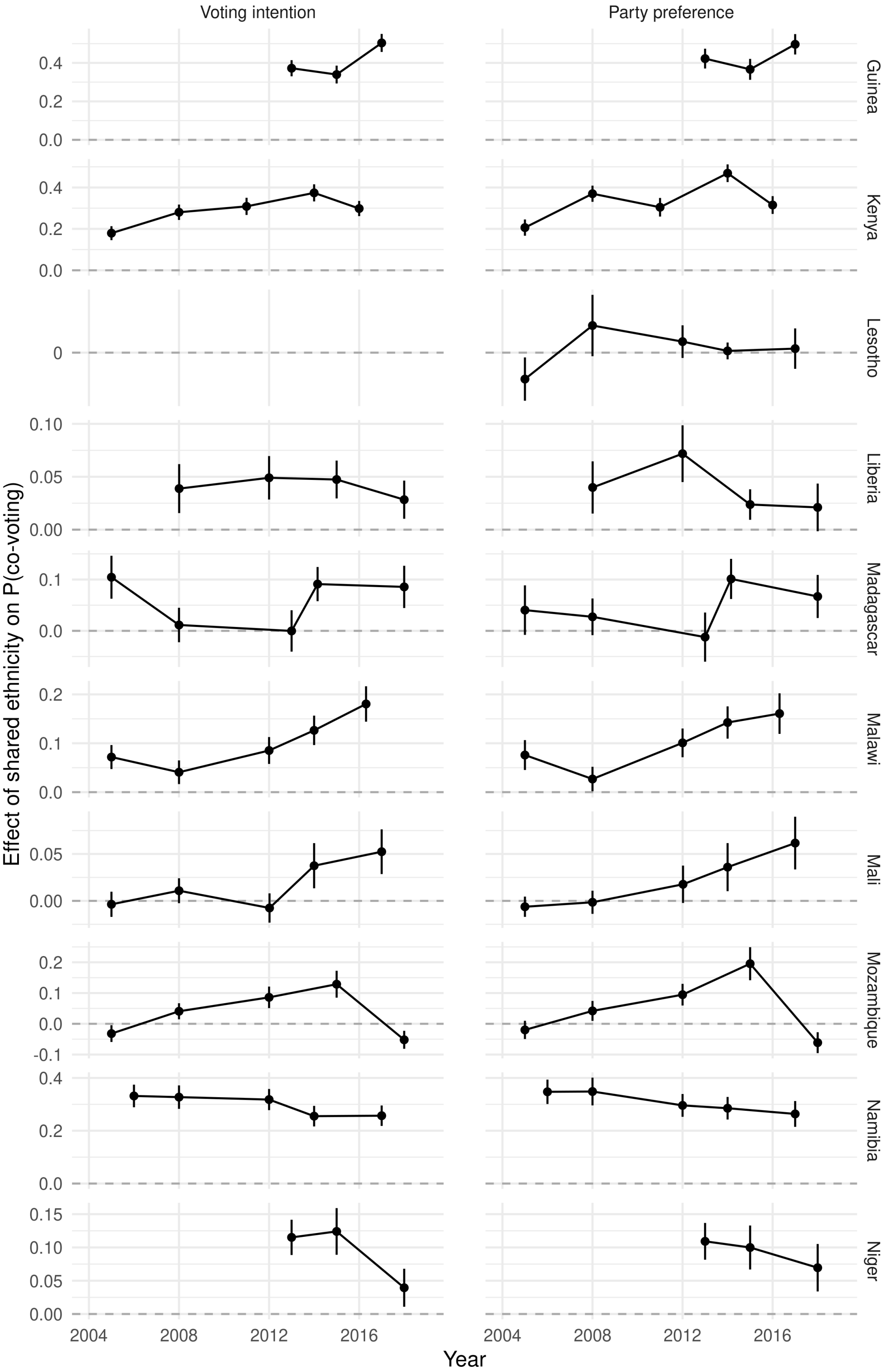


Figure A7: By country, over time, II

Note: Coefficients result from the fully specified model in Eq. 8 estimated separately for each Afrobarometer survey round since Round 3 with respondents' self-identified ethnicity as ethnicity indicator.

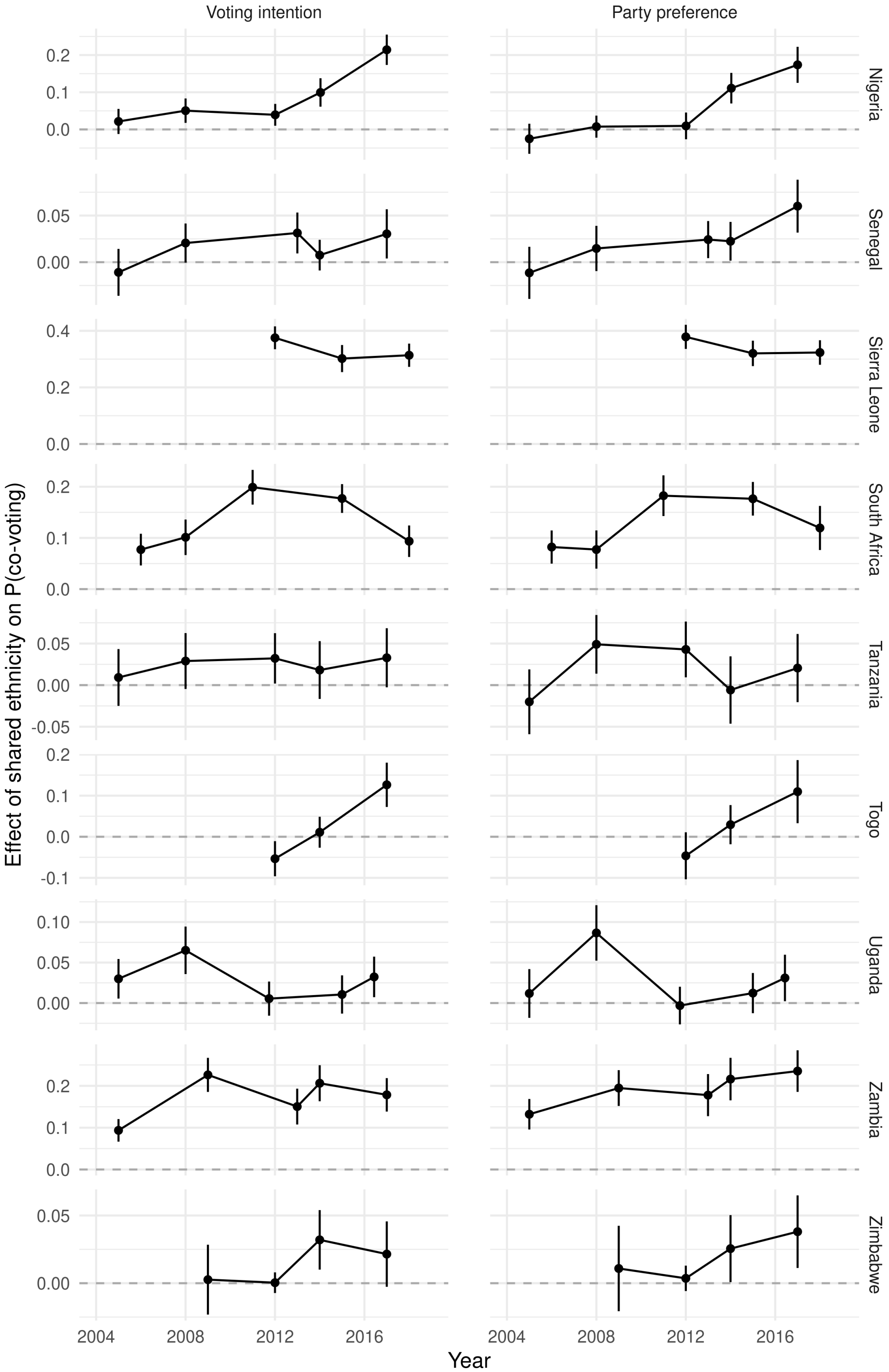


Figure A8: By country, over time, III

Note: Coefficients result from the fully specified model in Eq. 8 estimated separately for each Afrobarometer survey round since Round 3 with respondents' self-identified ethnicity as ethnicity indicator.

C Network-based partition model

We here apply the recently developed *Probabilistic Spatial Partition Model* (Müller-Crepon, Schvitz and Cederman 2023) to the case of the partitioning of voters into parties. The model was originally developed to partition geographical space into state territory. We transform the understanding of space to extend to a multi-dimensional electoral space. To apply the model we need to understand voters as the nodes of a network which is divided into partitions (candidates or parties) based on dyadic differences and similarities between voter characteristics, in short, electoral cleavages.

C.1 Probabilistic Partition Model

Following Müller-Crepon, Schvitz and Cederman (2023), we model the partitioning of voters as a Boltzman distribution

$$Pr(P = p_k) = \frac{e^{-\epsilon_k}}{\sum_{k=1}^{|\mathbb{P}|} e^{-\epsilon_k}}, \tag{A1}$$

where the chance that a given partitioning p_k is realized decreases with its "energy" ϵ_k. This energy can be interpreted as political tensions in a given division of voters into parties: the more voters are dissatisfied with the party they vote for in a given partitioning, the higher the tension and the less likely the partitioning emerges. Partitionings' energy ϵ_k results from attractive and repulsive forces $\epsilon_{i,j}$ between voters i and j. These forces are only realized when i and j support the same party ($\mathbb{1}_{i,j} = 1$) and not otherwise:

$$\epsilon_k = \sum_{i,j \in L} \mathbb{1}_{i,j} \epsilon_{i,j}, \tag{A2}$$

$$\epsilon_{i,j} = \beta_0 + \gamma\, \mathbf{x}_{i,j}, \tag{A3}$$

The political attraction and/or repulsion between pairs L of voters i and j is determined on the one hand by a constant baseline attraction β_0, as well as a vector of dyadic comparisons $\mathbf{x}_{i,j}$ between them. These comparisons can include binary indicators of differing ethnicity or gender, as well as distance measures, such as their wealth difference or geographic distance between them. Intuitively, we expect individuals with different ethnic backgrounds or vastly different incomes to be less likely to vote for the same party – indeed, were the same party trying to attract them, it might end up not succeeding or splitting. The vector of γ parameters indicates the effect of each dyadic voter comparison on the attraction and repulsion between voters and thus ultimately the partitioning of voters into parties. Estimating parameters in γ is therefore our ultimate goal.

As can be seen, similar to our setup in the baseline analysis, this formulation of vote choice is entirely dependent on comparisons between voters and does therefore not pre-suppose the existence of any party or set of parties. These emerge endogenously as the result of co-voting between voters. This allows for estimating the model across countries or country-periods with differing sets of parties and candidates.

We estimate Eq. A3 using the same data as used in the main analysis. In fact, the set of dyadic comparisons constructed for each country-round of the Afrobarometer can be represented as a graph $G_{c,t}$ of voters $i, j \in N$ who are associated with party or candidate preferences. The edges L of G encode the covariates $\mathbf{x}_{i,j}$ in Eq. A3 that determine whether voters i and j are likely to vote for the same (attraction) or two different (repulsion) parties. These co-variates are the same as used in the main analysis. Instead of a separate fixed effect for each country, we add one variable which stores the average attraction between nodes from each country. This is derived as the intercept of an otherwise empty model estimated separately for each country.

C.2 Results

Table A12 presents the main estimates from the partition model, derived – as in the main analysis – from unconditional and conditional models of the effect of share mother tongues on respondents' joint support for presidential candidates and legislative parties. We find relatively large estimates which are stable across specifications and outcomes and associated with little uncertainty. Importantly and as in our main analysis, the effect associated with a shared mother tongue does not significantly change with the introduction of other covariates

Table A12: Shared mother tongue and respondents' partitioning into candidates and parties

	Presidential candidates		Party support	
	(1)	(2)	(3)	(4)
Constant	0.0000	0.0044*	0.0003	0.0036
	[−0.0008; 0.0013]	[0.0015; 0.0069]	[−0.0003; 0.0025]	[−0.0003; 0.0062]
By-country intercept	0.8142*	0.8172*	0.8466*	0.8690*
	[0.7190; 0.9533]	[0.6911; 1.0118]	[0.7889; 0.9879]	[0.7819; 1.0359]
Shared mother tongue	−0.0091*	−0.0087*	−0.0095*	−0.0092*
	[−0.0119; −0.0063]	[−0.0116; −0.0057]	[−0.0162; −0.0057]	[−0.0155; −0.0059]
Countries	26	26	28	28
Respondents	16833	16833	12989	12989
Edges	5805059	5805059	3318087	3318087
Controls	no	yes	no	yes

Notes: 95% confidence intervals from country-level bootstrap in parenthesis. * Statistically significant at the 95% level.

We take two additional steps to gauge the comparability between the results from the partition model with our main results. First, we sequentially decrease the connectivity in graph G that underlies the model to the point where each respondent is connected only to one other respondent. We ensure that each respondent is connected to a uniform number of edges by constructing the sparse graphs G as the union of ring graphs. Each ring graph contains the full set of respondents in a random order and connects each respondent to their two ring-neighbours. Respondent orders are sampled such that the ring graphs do not contain overlapping edges.

Estimates from the disjoint graph of one dyad per respondent without any overarching network structure (see Figure A9) are very close to the estimates obtained from a logistic regression model using the main specification (see Table A10 above). The partition model indeed reduces to a simple edge-wise logistic regression where edges are not connected to each other. Deviations can be explained by the sampling error incurred when sparsening the graph.

Once the density of the network increases, coefficient estimates naturally decrease – this reflects that the importance (or 'energy') of any one edge in influencing the partition membership of each node decreases with the number of its edges. Yet, Figure A9 shows that the *ratio* between the effect of shared mother tongue and the remaining coefficients remains remarkably stable. We take this as further evidence that the network estimator closely mirrors our main results.

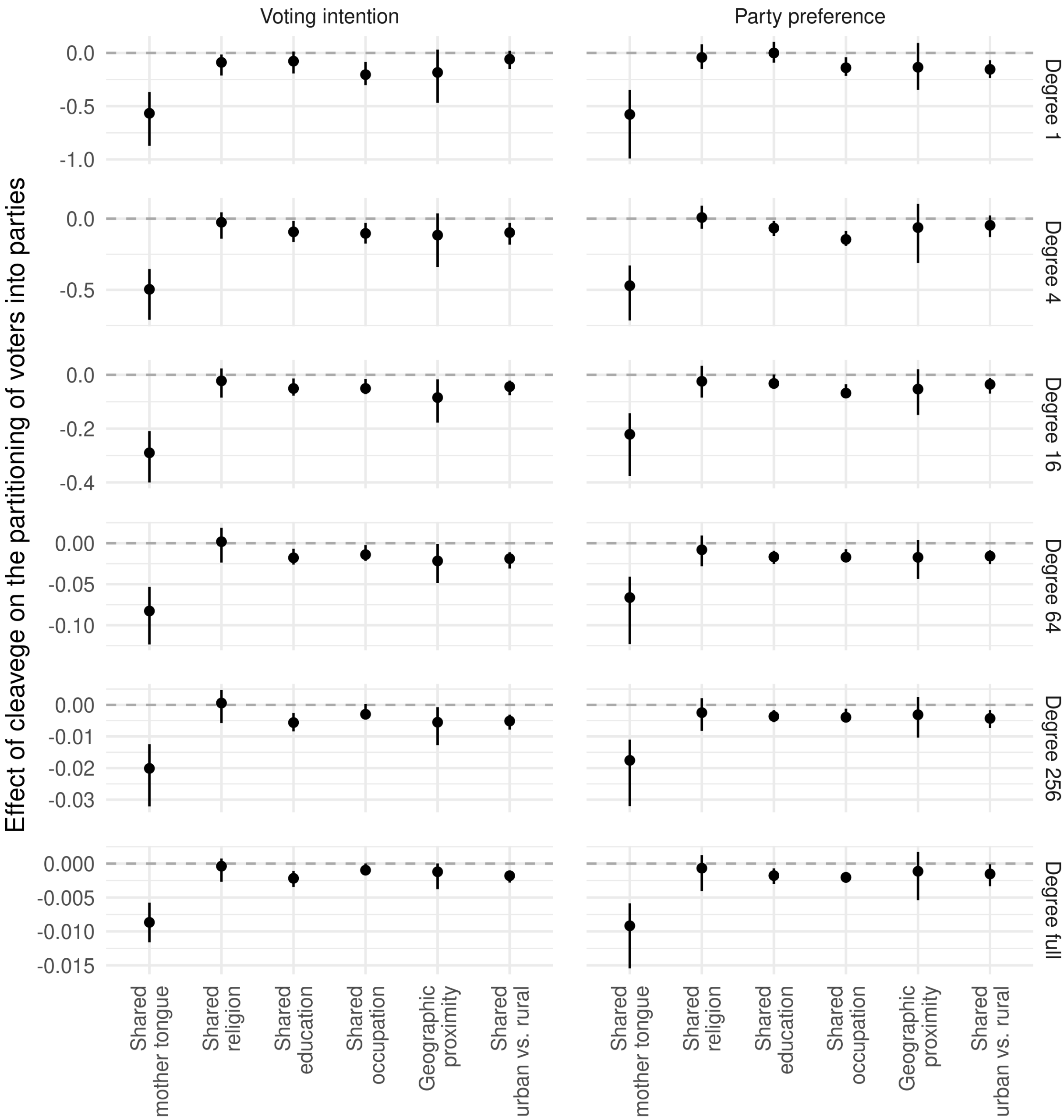


Figure A9: Partition model estimates by degree of connectivity

Note: The figure plots the most important predictors of the partitioning of respondents into presidential candidates and parties, by degree of network connectivity. Each set of coefficients results from estimating Eq. A3 using the full set of control variables.

D References (Appendix)

Müller-Crepon, Carl, Guy Schvitz and Lars-Erik Cederman. 2023. "Shaping States into Nations: The Effects of Ethnic Geography on State Borders." *American Journal of Political Science, FirstView* .