

Coethnics Covote in Africa: Studying Electoral Cleavages with a Covoting Regression Model

Carl Müller-Crepon* Nils-Christian Bormann†

March 20, 2026

Abstract

Ethnicity is an important cleavage in Africa, yet its influence on voting is contested. Existing micro-level analyses often focus on electoral support for a subset of candidates or parties, while meso and macro-level approaches infer individual motivations from aggregate outcomes. The risk of selection bias arises from the former approach, while omitted variable bias may result from the latter. Our new *Covoting Regression* (CVR) tackles several of these challenges. It estimates the effect of coethnicity on the probability that pairs of voters covote for the same party/candidate while conditioning on other characteristics shared between them. Thereby, CVR mirrors the micro-foundations of widely-used aggregate indicators, such as the Herfindahl-Hirschman indices of party concentration and ethnic homogeneity. Our data consists of dyadic comparisons between respondents from Afrobarometer surveys. Pooling across 28 countries, coethnicity increases covoting intentions by 17 percentage points. The effect of coethnicity is driven by politically relevant groups and covoting for ethnic parties. It is consistent across institutionally diverse countries and at least four times larger than that of other cleavages. Beyond ethnicity, we address key issues in studying electoral consequences of socio-economic cleavages and bridge 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, Martha Wilfahrt, and panel participants at the EPSA General Conference in 2023, the APSA Annual Meeting in 2025, and the Political Behavior Work in Progress Seminar at the London School of Economics. Nils-Christian Bormann's work was supported by the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 950359.

*Assistant Professor, Department of Government, London School of Economics and Political Science. Corresponding author: c.a.muller-crepon@lse.ac.uk. ORCID: [0000-0001-8536-7861](https://orcid.org/0000-0001-8536-7861)

†Professor of International Political Studies, Department of Philosophy, Politics & Economics, Witten/Herdecke University. nils.bormann@uni-wh.de. ORCID: [0000-0002-6016-6841](https://orcid.org/0000-0002-6016-6841)

It used to be a truism among political scientists that African voters support co-ethnic candidates and African parties target coethnics 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 alternative 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 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 effects of ethnicity on vote choice and party systems in Africa are in part spurious, simply reflecting underlying geographic clustering or shared economic interests ([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 coethnic candidates, they risk *selection bias* if some voters exclusively face (non-)coethnic 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 analyzing the degree to which ethnic groups differ in their voting patterns ([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 such as geographic or economic differences, this approach might be biased by omitting non-ethnic determinants of vote choice ([Ferree and Horowitz 2010](#); [Boone et al. 2022](#)).

¹Our argument focuses on elections of parties and candidates, which is why we use these labels interchangeably.

We introduce a Covoting Regression (CVR) model as a new analytical approach. The CVR combines the complementary strengths of micro-level voting studies and comparative work at the meso and macro-levels to address several of the problems affecting either. The model is inspired by the micro-level logic of the widely-used Ethno-Linguistic Fractionalization (ELF) index, which yields the probability of two randomly drawn individuals belonging to different ethnic groups. As the ELF's inverse, the Herfindahl-Hirschman Index (HHI), the CVR focus on the likelihood of two individuals to share membership in the same group. Specifically it models the likelihood of covoting among pairs of individuals, which we operationalize as the shared intention to vote for the same party or candidate.² Yet, instead of building aggregate country or group-level measures, we model covoting at the micro-level of individual *pairs of voters*.³ This allows us to estimate the effect of co-ethnicity (or any other cleavage) between individuals conditional on other dyadic socio-economic characteristics with a standard linear probability model. The main methods we present are implemented in the R-package `CVReg`.⁴

Conceptually, the estimated conditional effect of coethnicity parallels Huber's (2012) group- and party-based understanding of ethnic voting. It increases as members of different groups vote for different parties, and as voters of one party increasingly come from one ethnic group. Aggregating across these two dimensions, the CVR thus explicitly considers ethnic bloc voting for candidates from other ethnic groups as an integral part of ethnic voting. Conceptually, this is a major departure from prevalent micro-level approaches, which typically rely on coethnicity between voters and candidates. Beyond the broad micro-level interpretation, we use the CVR to derive covariate-adjusted estimates of group-based and party-based conceptualizations of ethnic voting at the meso-level (Huber 2012). At the macro-level, the model's estimates reflect the classic Herfindahl-Hirschman Index and can be interpreted as marginal changes in party system concentration as a consequence of marginal changes in ethnic homogeneity.⁵ The CVR thus bridges the gap be-

²While we empirically measure covoting intentions and party preferences, our model applies conceptually to actual covoting. We use the shorter term in our theoretical discussion.

³Our dyadic approach is similar to work in international relations on the formation of trade blocs or military alliances.

⁴Available at <https://github.com/carl-mc/CVReg>

⁵This is the elasticity of party system change to infinitesimal changes in another predictor variable.

tween the differing levels of analysis in studies of electoral behavior with possible applications to cleavages beyond ethnicity.

The CVR solves several of the aforementioned methodological problems of micro, meso, and macro-level approaches. By modeling pairwise covoting, parties and candidates disappear from our formulation. This abstraction reduces selection biases from restricted choice sets. Moreover, it allows for comparative analyses that evaluate the relative influence of ethnicity relative to other cleavages across country borders or within countries over time. These advantages, however, come at the cost of the model's inability to yield group- or party-specific inferences. Researchers interested in explaining why members of particular ethnic groups vote for a particular party may use classic micro-level voting models. However, these models do not allow to make general statements about the strength of ethnic voting in a given country. Other forms of selection bias, resulting from missing data and strategic party entry, still affect the CVR.⁶ At the same time, estimating the conditional association between coethnicity and covoting addresses ecological inference problems and bias from confounders which can affect meso- and macro-level studies, while increasing statistical power.

Empirically, we use the CVR to provide comprehensive evidence on the effect of coethnicity on covoting across Sub-Saharan Africa. Drawing on multiple rounds of the Afrobarometer surveys from 28 states across Sub-Saharan Africa, we recast each country sample into pairwise comparisons between respondents. We proxy unobserved actual covoting by measuring our *outcome variables* as covoting intentions in presidential elections and shared preferences for a political party.⁷ Coethnicity, the main *explanatory variable*, is measured as the shared mother tongues of 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 covoting intentions in linear probability models, and adjust standard errors to address the repeated inclusion of individuals from a limited number of ethnic groups across dyads.

Our results show strong support for the dominance of coethnicity in shaping

⁶Note however that the reflection of ethnic bloc voting for non-coethnic candidates alleviates biases from endogenous party formation in some but not all contexts.

⁷The Afrobarometer asks survey participants which party they feel close to.

covoting intentions in most African states in our sample. We estimate that co-ethnicity by mother tongue increases the probability that two respondents have shared voting intentions by 17 percentage points. Zooming in on specific countries and elections, we discuss variation in the estimated effect of coethnicity on covoting intentions over time and across cases. Additional analyses account for cross-cleavage interactions, alternative measures of coethnicity and linguistic distance, different sampling procedures, and modeling choices. We even uncover substantial, if smaller, effects of coethnicity on covoting intentions within comparatively homogeneous villages or neighborhoods, suggesting that coethnicity captures more than shared, spatially clustered political preferences.

In a set of descriptive analyses, we demonstrate how the CVR can be applied to test explanations of ethnic voting or changes therein at the three levels of analysis. At the micro-level, we find that higher ethno-centric trust and shared perceptions of a country's problems or government performance do not robustly explain coethnic covoting. At the meso-level, the effect of coethnicity is augmented in politically relevant ethnic groups and strongest for covoting for parties identified by political scientists as mostly targeting members of one ethnic group. At the macro-level, we find no substantive moderating effects of electoral, democratic, or traditional institutions.

Beyond coethnicity, our results also show positive effects of religious, educational, occupational, and geographic similarities on covoting, thus supporting conjectures on socio-economic characteristics (e.g., [Boone et al. 2022](#); [Bratton, Bhavnani and Chen 2012](#); [Koter 2016](#); [McCauley 2014](#)). Yet, across our broad sample of African electorates, the effects of these alternative cleavages are at least five times smaller than that of coethnicity.

We conclude by discussing the implications of our new approach for the study of electoral cleavages. In doing so, we highlight the importance of bridging levels of analysis to achieve inferences of high internal and external validity and discuss the utility of the CVR in studying the electoral effect of socio-economic cleavage. Potential extensions can apply the CVR to local election results and individual attitudes and choices beyond voting.

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 psychological or instrumental motivations for ethnic voting and the corresponding existence of ethnic parties (Mozaffar, Scarritt and Galaich 2003). In short, the psychological approach entails that voters reaffirm their identity through voting for coethnic 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 coethnics increase electoral support (Rabushka and Shepsle 1972; Horowitz 1985).⁸ In turn, instrumentalists suggest that African voters support coethnic candidates to receive economic benefits through clientelist exchanges during the election period and patronage distribution afterwards, if “their” candidate joins the ruling coalition. Political elites themselves prefer to build ethnically-based support coalitions to limit access to state funds to ethnic insiders (Bates 1974; Chandra 2007).

Political scientists investigate individual-level demand and party-level supply at the micro, meso, and macro-levels. *Micro-level* studies probe both the mechanisms that link individuals’ ethnic identity to vote choice and alternative explanations for party preferences. Experimental research supports an instrumentalist interpretation of ethnic voting that portrays the ethnicity of candidates as an informational shortcut for voters to learn about the likelihood of future economic benefits (Ferree 2006; Ferree, Gibson and Long 2021; Conroy-Krutz 2013; Carlson 2015; Casey 2015; Ferree, Gibson and Long 2021).⁹ In contrast, survey-based research frequently explores alternative voter motivations, such as 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

⁸Sociological explanations of voting following Lazarsfeld, Berelson and Gaudet (1968/1944) are curiously absent from the ethnic voting literature, a theme we return to in the conclusion.

⁹Though see Adida et al. (2017), who highlight motivated reasoning as linking ethnic identity and performance evaluation. Alternatively, Mor (2022) documents grievance-driven demands for a Catholic party in 19th-century Prussia.

countries, including traditional authorities (Baldwin 2013; Koter 2016), and local ethnic geography (Ichino and Nathan 2013).

Comparativists working at the *macro-level* complement micro-level studies by highlighting supply factors. Guided by Duverger's law, cross-country comparisons suggest that the effective number of ethnic groups in a country correlates with the effective number of parties under permissive electoral rules (Mozaffar, Scarritt and Galaich 2003; Clark and Golder 2006; Lublin 2017). In contrast, case studies show that party competition goes beyond ethnicity (Elischer 2013) and that presidential candidates pursue non-coethnic supporters (Horowitz 2022). Larger groups are indeed more likely to have coethnic candidates on the ballot (cf. Posner 2004b) and to see intra-group competition Ferree (2010). Echoing micro-level insights on the importance of alternative cleavages, Boone (2024) argues that the correlation between ethnicity and vote choice hides the effects of regionally concentrated economic interests (see also Ishiyama 2012). Analyses of ethnic voting at the group or *meso-level* fall in between these two end points by highlighting group-level economic inequality and electoral rules as moderators (Huber 2012; Huber and Suryanarayan 2016; Houle, Park and Kenny 2019).

Studies at each level of analysis complement each other. Substantively, research at all levels raises questions about the influence of other cleavages including urban-rural differences (Nathan 2016; Wahman and Boone 2018; Harding 2020), regional coalitions (Ferree and Horowitz 2010; Boone 2024), and economic class interests (Bratton, Bhavnani and Chen 2012; Resnick 2012; Huber and Suryanarayan 2016; Houle, Park and Kenny 2019). Micro-level research trades in generalization against detailed, case-specific insights with the potential of identifying causal underpinnings of vote choice. Meso and macro-level comparisons potentially offer greater external validity but rely on *ecological inference* and risk omitted variable bias.¹⁰ Taking groups as units of analysis or building blocks of aggregate indices, meso and macro-level analyses assume the primacy of ethnicity rather than testing it.¹¹

Importantly, micro, meso, and macro-level approaches differ in their concep-

¹⁰Appendix B2 illustrates this bias via simulations.

¹¹Fixing group boundaries also precludes a more nuanced understanding of potentially variable ethnic boundary markers or even endogenous processes of identity change (Posner 2005; McCauley 2014; Green 2021; Müller-Crepon 2025).

tualization of ethnic voting itself. Both macro and micro-level studies commonly analyze the degree of ethnic voting as the probability that an individual chooses a coethnic candidate or party. Meso-level research, in turn, analyzes the extent of ethnic divergence in voting patterns based on the electoral distance between groups and ethnic distance between parties (Huber 2012). Respectively, we label these alternative strategies the *voter-candidate correspondence* and the *divergence* approach. In the correspondence approach, ethnic voting increases as members of a group increasingly vote for parties that represent their group *and* their proportion among all voters of these parties increases. In contrast, ethno-political divergence increases under two alternative conditions: as a political party increasingly draws support from one ethnic group in what Huber (2012) coins the “party-based approach”, or as voters from one ethnic group vote together (“group-based approach”).

This contrast brings into relief three important conceptual and methodological issues. First, empirical analyses that employ the correspondence approach prioritize a party- rather than group-based understanding of ethnic voting. This can lead to an underestimation of ethnic voting when members of a group vote en bloc for a non-coethnic candidate. While the correspondence approach would not regard those votes as “ethnic”, the divergence approach would detect this ethnicization of political preferences.¹² Second, the correspondence approach classifies votes for coethnics as ethnic even if they were cast for a “non-ethnic” reason or even due to pure randomness. This produces an upward bias in aggregate estimates of ethnic voting.¹³ Third, constrained choice sets can leave voters with either exclusively coethnic or non-coethnic candidates (Ferree 2022). This produces selection bias in the correspondence approach.¹⁴

This selection bias arises in three analytical situations. First, potential candidates do not run in expectation of low support from their group, for example, where groups are too small to lift candidates to victory (Posner 2004b). When excluding such groups from an analysis, analysts will potentially overestimate ethnic voting in a society. Second, the utility of ethnic voting might be stable across

¹²Appendix A formalizes the differences between the various approaches.

¹³See Appendix A.3 for a simulation of this bias.

¹⁴Selection bias affect experimental studies less, but biases may occur if they are inconsistent with the real-world experiences of participants (see Dafoe, Zhang and Caughey 2018).

groups, but some groups might still have no coethnic candidate. Here, a “0” classification of ethnic voter-candidate correspondence mixes the absence and the impossibility of ethnic voting. If members of such “unrepresented” groups voted en bloc for a “second-best” non-coethnic who represents their interests, correspondence approaches would underestimate the overall extent of ethnic voting. Huber’s divergence approach, however, can capture group-based voting for non-coethnic candidates. In a third situation, a broad choice set is available to voters 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](#)). In this case, the degree of ethnic voting may be over- or underestimated depending on the utility that individuals derive from ethnic voting and the distribution of votes for excluded candidates.¹⁵

Because we only observe real but not hypothetical alternative candidate choice sets, those analyzing observational data currently have two options.¹⁶ First, voting studies that rely on the correspondence approach would need to adjust for factors that influence the existence of “ethnic” candidates ([Ferree 2022](#)), and candidates’ expected vote shares under non-ethnic voting. However, this strategy is difficult to implement in broadly comparative work. Second, the divergence approach developed by [Huber \(2012\)](#) offers the possibility to model group-based voting across contexts even where choice sets are constrained. It frees the researcher from specifying correspondences between voters and parties, and from taking arbitrary decisions on which groups to include in the analysis. However, the aggregation of individual choices to the group level makes this approach vulnerable to omitted alternative cleavages.

In sum, a new approach to observational research on ethnic voting should (1) account for alternative motivations and identity categories that might serve as a basis for voting to avoid omitted variable bias; (2) avoid the pitfall of artificially limited choice sets to overcome selection bias, and (3) facilitate coverage of a broad number of countries over time to ensure external validity.

¹⁵Appendix A.3 simulates scenarios 1 and 2. We show that the correspondence approach entails much larger selection bias than our new covoting approach.

¹⁶Researcher-controlled experimental work is better placed to address the problem of endogenous candidate supply resulting from strategic entry and exit decisions in actual electoral races.

Covoting Regression: A new way to estimate the effects of (ethnic) cleavages

Our new Covoting Regression (CVR) bridges prevailing approaches to ethnic voting by combining their strengths, thus overcoming their individual weaknesses. In short, we follow macro- and meso-level measurement strategies to conceptualize the effect of ethnic cleavages as the effect of coethnicity on covoting intentions. Yet we do not rely on aggregate indicators, which suffer from problems of ecological inference and omitted variable bias. Instead, we study covoting intentions among pairs of individuals contained in survey data with standard regression models. In the following, we introduce our method, 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 societal cleavages. Empirically, they employ aggregate measures to operationalize the main variables of interest, such as the Herfindahl-Hirschman Index (HHI) and the effective number of parties (ENP) to measure party system concentration, or the 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 of party concentration and its inverse, Laakso and Taagepera's (1979) ENP, are often computed with meso-level data on parties' vote shares as

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

where $\frac{N_p}{N}$ is the vote share of party $p \in K$ among voters N . Party concentration thus decreases with more and more equally sized parties, which in turn increases the effective number of 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 individuals' voting probability for that party. The square of that probability then yields the chance that two random voters voted for the same party. We can thus reformulate $HHI(party)$ as the chance that two randomly chosen voters vote for the same party:

$$HHI(party) = \frac{1}{N^2} \sum_{i,j}^N \text{covoting}_{i,j}, \quad (2)$$

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

Rarely discussed, the covoting formulation in Equation (2) includes comparisons within the same individual $i = j$, inducing downward bias when computed on a finite sample of individuals N . This is because comparisons within the same voter $i = j$ must yield $\text{covoting}_{i,j} = 1$.¹⁷ Drawing on [Simpson \(1949\)](#), this bias can be corrected by avoiding "within-individual" comparisons when computing the HHI:

$$HHI(party) = \sum_{p=1}^N \left(\frac{N_p}{N} \frac{N_p - 1}{N - 1} \right) \quad (3)$$

$$= \frac{1}{N^2 - N} \sum_{i,j,i \neq j}^N \text{covoting}_{i,j} \quad (4)$$

Equation 4 shows how the unbiased expectation of $HHI(party)$ is equivalent to the average *covoting* rate between all voters.¹⁸ In turn, ethnic homogeneity among voters as measured through the HHI is equivalent to the average rate of coethnicity between all voter pairs. Moving beyond measures of diversity, measures of dispersion such as the Gini coefficient can be similarly reformulated as comparisons between pairs of individuals.¹⁹

We hone in on these micro-foundations of macro-level approaches and propose to model the effect of ethnic cleavages on voting by estimating the effect of coethnicity on covoting in pairs of individuals i and j . We start deriving our *Covoting*

¹⁷The bias decreases with N .

¹⁸Equation (4) approaches Equation (2) as N increases towards infinity.

¹⁹The 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).

Regression (CVR) model as

$$\text{Covoting}_{i,j} = \beta_0 + \epsilon_{i,j} \quad (5)$$

with $i, j \in N, i > j$.²⁰ In this formulation, estimate β_0 captures the average rate of covoting among all pairs of individuals in the sample. It is thus equivalent to ENP^{-1} in Equation (4).²¹

The pairwise regression model in Equation (5) can be easily extended by adding a covariate matrix $\mathbf{x}_{i,j}$ which measure individuals' i and j 's similarities or differences on important socio-economic cleavage dimensions as well as other control variables such as cross-cleavage interactions.²² We thus propose to estimate the effect of coethnicity on covoting as

$$\text{Covoting}_{i,j} = \beta_0 + \beta_1 \text{coethnicity}_{i,j} + \gamma \mathbf{x}_{i,j} + \epsilon_{i,j}. \quad (6)$$

The focus of the CVR on covoting between i and j abstracts from the identities of candidates and parties. Their disappearance from the formulation circumvents the selection bias that arises in the correspondence approach when linking candidates to ethnic groups. Yet, the CVR remains vulnerable to biases from endogenous candidate supply when analyzing observational data. However, we find that its accounting for ethnic bloc voting decreases the size of this bias in comparison to the correspondence approach.²³ We also note that the CVR can be used to evaluate data from experimental studies with exogenous candidate supply.

Due to the equivalence of the $\text{coethnicity}_{i,j}$ 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 macro-level and conditional on covariates $\mathbf{x}_{i,j}$, it is the elasticity of the party concentration in response to marginal changes in ethnic

²⁰As covoting is not directional, we limit ourselves to one comparison between any two voters. Computationally though, the result is the same as in Equation 4.

²¹The estimate of the intercept in an empty linear regression model, β_0 , equals $\bar{y} = \frac{1}{n} \sum_i y_i$ (Wooldridge 2008, 29). Appendix Figure A7 empirically demonstrates this equivalence. Our estimation of ENP and HHI adds the benefit of uncertainty estimates reflecting survey sampling.

²²Appendix B.2 shows how the CVR successfully accounts for correlated cleavages.

²³See Appendix A.3.

homogeneity²⁴ such that

$$\frac{\delta HHI(\textit{party})}{\delta HHI(\textit{ethnic})} = \frac{\delta \textit{covoting}_{i,j}}{\delta \textit{coethnicity}_{i,j}} = \beta_1. \quad (7)$$

This feature of the CVR extends to other dyadic comparisons between voters that reflect macro-level measures, such as pairwise wealth differences, which together constitute the Gini coefficient.

At the meso-level of pairs of ethnic groups, β_1 captures the difference in average covoting within groups and covoting between groups, with group-pairs weighted by the product of their size. As a consequence, Eq. 6 can also be used to derive the expected rate of covoting among coethnics as well as the expected rate of coethnicity among covoters at average covariate values \bar{x} .²⁵

$$Pr(\textit{covoting} \mid \textit{coethnicity} = 1) = \beta_0 + \beta_1 + \gamma\bar{x} \quad (8)$$

$$Pr(\textit{coethnicity} \mid \textit{covoting} = 1) = \frac{\textit{coethnicity}}{\textit{covoting}} (\beta_0 + \beta_1 + \gamma\bar{x}), \quad (9)$$

where the rate of coethnicity among covoters in Eq. 9 is the rate of covoting among coethnics scaled by the relative share of coethnic pairs compared to covoters in the population. These measures closely resemble the inverse of Huber's (2012) measures of ethnic Group Vote Fractionalization and Party Vote Fractionalization (see Appendix D.2). Importantly, the conditional rates of coethnicity and covoting are driven not only by the effect of coethnicity on covoting β_1 , but also by the intercept as well as average covariate values and effects.

Finally, at the micro-level, β_1 can be interpreted as the marginal effect of coethnicity on the probability of covoting between individuals, again conditional on covariates. In principle, this covoting probability can also be recovered from voting probabilities derived from a multinomial logistic or linear regression model (Appendix A.4). Doing so yields largely equivalent results but does not yield standard uncertainty estimates and comes with estimation problems where data is sparse.

²⁴Interpretations that extrapolate beyond small margins neglect the possibility of corresponding strategic party and exit and its knock-on effects on covoting among (non-)coethnics. See Appendix B.1.

²⁵See Appendix A.2 for further discussion and parallel derivations of the rate of covoting among non-coethnics and non-covoting among coethnics which have comparative value.

Reformulating covoting as the product of two individual probabilities to vote for the same party further reveals how to address biases that stem from the endogeneity of one cleavage x_1 to another x_2 . Where such endogeneity exists, it becomes necessary to not only model voter characteristics of i and j , but also their interactions $x_{1,i} x_{2,j}$ and $x_{2,i} x_{1,j}$, thereby capturing cross-cleavage covoting.

By providing interpretations at all levels of analysis based on dyadic, micro-level modelling, the CVR combines the advantages of previous approaches. In particular, it avoids making ecological inferences from macro-level modelling but at the same time allows for inter-temporal and cross-country comparisons of the conditional association of coethnicity with covoting. While the conceptualization of ethnic voting as coethnic covoting entails losing information about which actual parties voters vote for, it prevents biases that arise from focusing only on a subset of parties. We show in our analysis of heterogeneous effects and mechanisms how the CVR can be used to test theoretical arguments about the drivers of ethnic voting at all three levels of analysis. To facilitate understanding of our main empirical strategy, we first introduce our data structure and then present our regression model.

Dyadic data on covoting, coethnicity, and other cleavages

To operationalize our analysis of covoting intentions among individuals, we transform survey data into pairs of individual respondents. For each pair, we encode whether respondents share voting intentions 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-Saharan Africa since 1999. For the most part, our main analyses rely on the survey's seventh round fielded between 2015 and 2018 in 28 states with 1'200 respondents per country.²⁶ To gauge

²⁶Benin, 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. We randomly subset the larger surveys from Ghana, Kenya, Mozambique, Nigeria, and South Africa to 1'200

variation in the effect of ethnic cleavages over time, we draw on rounds 3–7.²⁷ 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 covoting intentions along with a wide range of non-ethnic cleavage dimensions discussed by existing work.

Unit of analysis: Following the logic introduced in Equation (4), we transform survey-rounds for each country into undirected dyadic comparisons between all respondents $i, j \in N_{c,t}$ with $i \neq j$. 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 717 (533) respondents and 256,686 (141,778) dyadic comparisons between them per country surveyed.²⁸

Measuring covoting intentions: Because actual voting behavior remains unobserved in the Afrobarometer, we encode our main two measures of covoting intentions by drawing on answers to 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?²⁹

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

respondents to give each country roughly equal weight.

²⁷Rounds 1 and 2 do not include an item on preferences over candidates in potential presidential elections.

²⁸Appendix Figure A18 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 A19.

²⁹Missing values are recorded for respondents who do not feel close to any party.

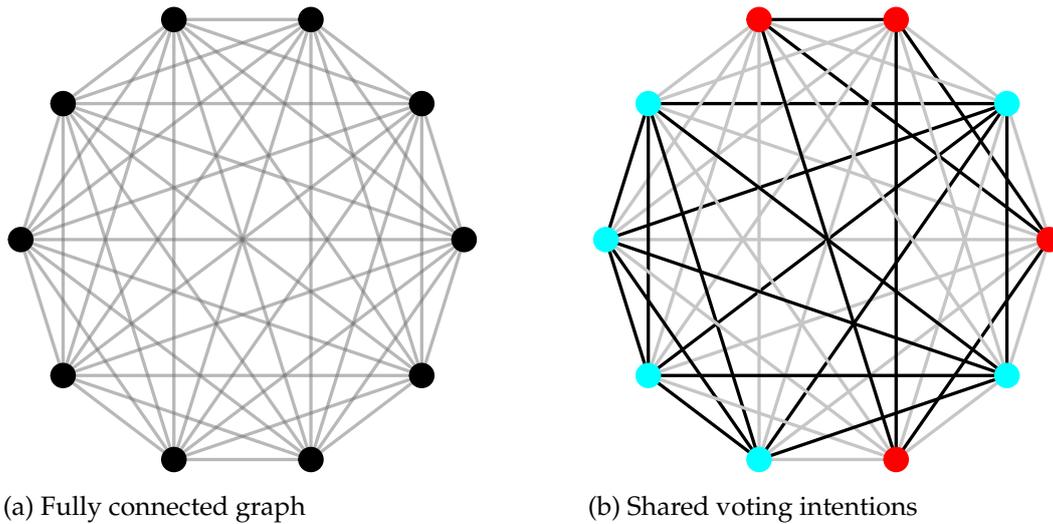


Figure 1: Covoting dyads from 10 respondents in Ghana, Round 7

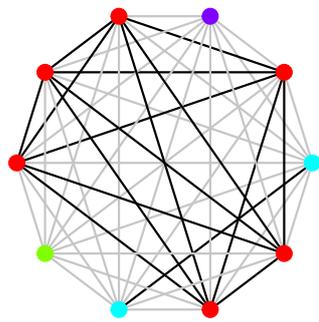
Note: In (b): NPP blue, NDC red. Covoting as black edges.

0 otherwise.³⁰ The result is visualized for a sample of 10 respondents from Ghana in Figure 1. 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. On average, we observe that respondents share their voting intention and party preference in 46.0 and 45.8 percent of all dyads, which corresponds directly to the cross-country average HHI of party concentration or the inverse ENP (see Equation 4).

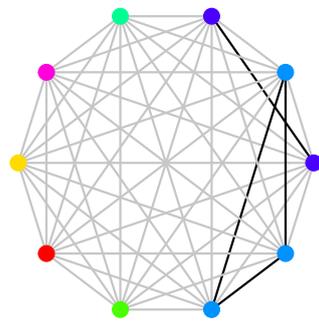
Coethnicity: We capture our main explanatory variable of interest – respondents' pairwise coethnicity – in a binary variable that records whether they share the same mother tongue (1) or not (0).³¹ As visualized for the randomly drawn 10 Ghanaians in Figure 2a, this leads to many coethnic dyads among respondents from large language groups (e.g., the Akan in red). On average, respondent pairs share their mother-tongue at a rate of 24 percent, which corresponds to the average HHI of

³⁰We drop individuals with missing responses and "other" parties as missing. While sensitive party identification in polarized environments can lead to selection bias (Davis and Wilfahrt 2025), results are robust to coding "other" parties as separate parties for each respondent (Appendix Table A20).

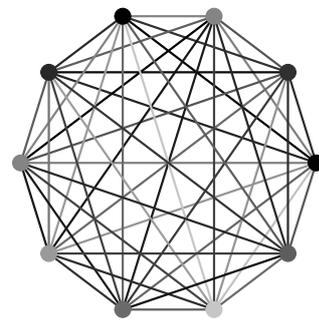
³¹The respective question reads: "Which [enter nationality] language is your mother tongue or language of origin?" Note that Afrobarometer round 7 is the first to ask specifically about respondents' mother tongue as separate from the language spoken in their home *now*. Hampering comparisons over time, all previous rounds ask about respondents' "home language", which leaves this crucial distinction open.



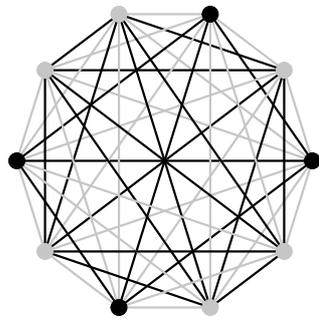
(a) Shared mother-tongues
 Note: Colors denote mother-tongues, intra-group edges in black



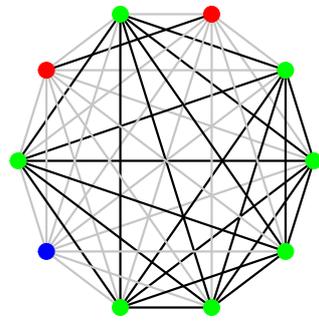
(b) Shared religion
 Note: Colors denote religious groups, intra-religion edges in black



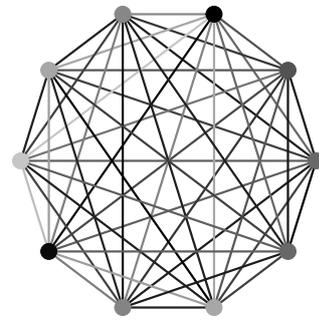
(c) Age similarity (decades)
 Note: Grey-scale denotes older age and greater similarity



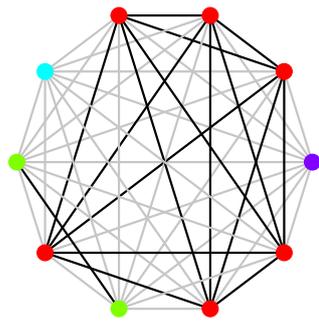
(d) Shared gender
 Note: Female in black, male in grey, intra-gender edges in black



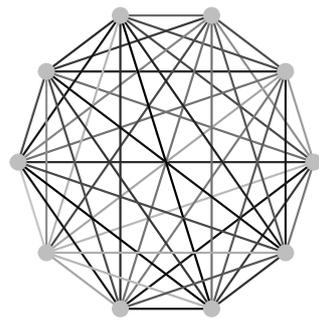
(e) Shared education
 Note: Colors denote education levels, intra-education level edges in black



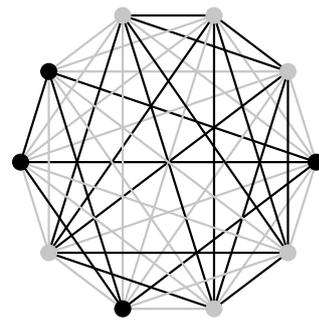
(f) Wealth similarity
 Note: Dark shades denote wealthier respondents and greater similarity



(g) Shared occupation
 Note: Colors denote occupations, intra-occupation edges in black



(h) Geographic proximity
 Note: Darker shades denote greater proximity



(i) Shared urban-rural status
 Note: Urban black, rural grey, shared status edges in black

Figure 2: Encoding of main explanatory variables on example graph of 10 respondents from Ghana

ethnic homogeneity across the countries in the sample.

Respondents' reported mother-tongue is among the least malleable ethnic identity indicators and therefore least likely affected by reverse causality or omitted variable bias. In contrast to the language spoken at respondents' home or their self-proclaimed ethnic identity, reported mother tongues should be least affected by respondents' political concerns (e.g. [Green 2021](#)) and assimilation (e.g. [Müller-Crepon 2025](#)).³² 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 consider a continuous rather than a categorical understanding of ethno-linguistic identity by analyzing variation in effects at short and large linguistic distances. Arguably, larger linguistic distances are harder to overcome through assimilation or misrepresentation, and thus less vulnerable to reverse causation. Third, a set of robustness checks zooms in on covoting 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 2b to 2i. For 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 covoting 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

³²We test the alternative measurements of ethnic identity in Appendix F.3.

minus absolute wealth differences.³³ Lastly, we capture purely *geographic cleavages* by including as-the-crow-flies proximity between respondents (in 1'000km)³⁴ and a dummy variable capturing whether respondents share their urban or rural status.

Combining data across countries and rounds: Because our measures of covoting intentions, coethnicity, 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 not possible with standard analytical approaches. Since party systems differ between countries or over time, it is impossible to estimate the importance of ethnicity for voting decisions for two or more different contexts in one regression model. Harmonization across contexts, for example, by estimating voter support for one type of party that exists in different contexts, such as revolutionary or communist parties, runs the risk of selection bias by excluding voting decisions for or against other parties. In contrast, the CVR enables such comparative work at scale.

Modeling the effect of ethnic cleavages on covoting

With the undirected dyad of respondents i and j as our main unit of analysis, we employ a linear probability model to estimate the probability of covoting intentions between respondents along the lines of the CVR proposed in Equation 6.³⁵ We move beyond the confines of one survey sample of individuals and generalize the model across countries and time as

$$\text{Covoting}_{i,j,c} = \alpha_c + \beta_1 \text{coethnic}_{i,j} + \gamma \mathbf{x}_{i,j} + \epsilon_{i,j} \quad (10)$$

where dyads are constructed only among individuals observed in the same country c at time t . α_c is a fixed effect for each country c , capturing the average rate of covoting in a country and the effects of unobserved country-level factors, such as

³³We 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. An extended version of the index adds respondents' level of education, employment type, and availability of fuel and yields very similar results (Appendix F.5).

³⁴The average distance between respondents amounts to approximately 300km.

³⁵Logistic regression models yield equivalent results. See Appendix Table A21.

states' population size, history, or electoral institutions. β_1 then captures the *average* effect of coethnicity by mother tongue on covoting intentions among individuals across all countries in the sample, conditional on the co-variates \mathbf{x} introduced and visualized in Figure 2 above.

Since almost 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 model with controls. Although it is theoretically possible that the potential biases from omitted variables and post-treatment controls lead us to overestimate the effect of coethnicity in both cases, small differences between the two specifications and subsequent robustness checks would suggest that we approximate the true effect of coethnicity.

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. We explore alternative clustering strategies in Appendix F.7: Clustering at the country-level yields just as conservative standard errors, while clustering on the level of individuals or their enumeration area yields smaller error bands. Lastly, we show that Aronow et al.'s (2015) cluster-robust variance estimator for dyadic data yields comparatively small uncertainty estimates when employed at the level of individuals, their ethnicity, or their locations of residence.³⁶

Results

Our analysis yields strong support of the hypothesis that coethnicity increases the rate of shared voting intentions and party preferences among Afrobarometer respondents. Table 1 presents our main estimates, showing the unconditional and

³⁶Beyond its effects on uncertainty estimates, unit-interdependence may bias point estimates in our setting. When explicitly addressing this concern with a *Probabilistic Partition Model* recently developed by Müller-Crepon, Schvitz and Cederman (2025), we obtain similar results.

conditional effect of coethnicity on respondents' covoting intentions, measured as shared voting intentions and preferences for parties. We find that pairs of respondents who share their mother tongue are between 16.2 and 17.7 percentage points more likely to have the same voting intention and party preference than pairs that do not. The effect is very stable across specifications, does not vary much between our two outcomes, and is associated with little uncertainty ($p < .001$).

Substantively, these effects are large. We observe shared covoting intentions in 46 percent of all survey respondent dyads. The conditional increase in covoting intentions resulting from shared mother-tongues of around 17 percentage points thus amounts to 37 percent of the average rate of covoting. The estimated effect of coethnicity also swamps that of any other pairwise similarity measure between respondents, the substantively largest being that of shared occupation, with an estimated effect of 2.6 percentage points. We compare the effect of coethnicity with other cleavage dimensions more thoroughly below.

At the meso-level, the coefficient of coethnicity together with the base rates of covoting intentions and coethnicity allow for computing the group- and party-based measures of ethnic voting (Eqs. 8 and 9). Based on Model 4, 57.7 percent of coethnics covote compared to 41.5 of non-coethnics. Taking the perspective of parties, we estimate that 34.3 percent of covoters come from the same ethnic group. In contrast, the estimated share of coethnics among pairs of individuals that vote for different parties is only 21.6 percent.

At the macro-level, our results indicate that, on the margin, increasing ethnic homogeneity translates into a higher concentration of parties with an elasticity of .17 (see also Equation 7).³⁷ This positive elasticity stands in drastic contrast to the *negative* bivariate relationship observed when using country-level data (see Appendix Figures A.8), highlighting the caveats of ecological inferences drawn from aggregate data.³⁸

We observe little systematic change in the aggregate effect of coethnicity on covoting intentions over time. When repeating our analysis for Afrobarometer rounds

³⁷The 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.

³⁸Our estimated country-level coethnicity effects correlate positively with Posner's (2004a) fractionalization index (Appendix Figure A10).

Table 1: Covoting and shared mother tongue

Dependent Variables: Model:	Voting intention		Party preference	
	(1)	(2)	(3)	(4)
<i>Variables</i>				
Shared mother tongue (0/1)	0.177*** (0.030)	0.168*** (0.030)	0.170*** (0.033)	0.162*** (0.034)
Shared religion (0/1)		0.020** (0.007)		0.019** (0.007)
Age similarity (decades)		-0.004** (0.001)		-0.002 (0.002)
Shared gender (0/1)		0.000 (0.001)		0.000 (0.001)
Shared education (0/1)		0.019*** (0.004)		0.015*** (0.003)
Wealth similarity (sd)		0.004+ (0.002)		0.002 (0.003)
Shared occupation (0/1)		0.026*** (0.006)		0.022*** (0.005)
Geographic proximity (1'000km)		0.041* (0.020)		0.035+ (0.019)
Shared urban vs. rural (0/1)		0.022*** (0.005)		0.021*** (0.006)
<i>Fixed-effects</i>				
Country x Round	Yes	Yes	Yes	Yes
<i>Fit statistics</i>				
Outcome mean	0.460	0.460	0.459	0.459
Countries	26	26	28	28
Respondents	17,827	17,827	13,892	13,892
Dyads	6,500,096	6,500,096	3,814,306	3,814,306
R ²	0.077	0.079	0.077	0.079
Within R ²	0.020	0.023	0.019	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

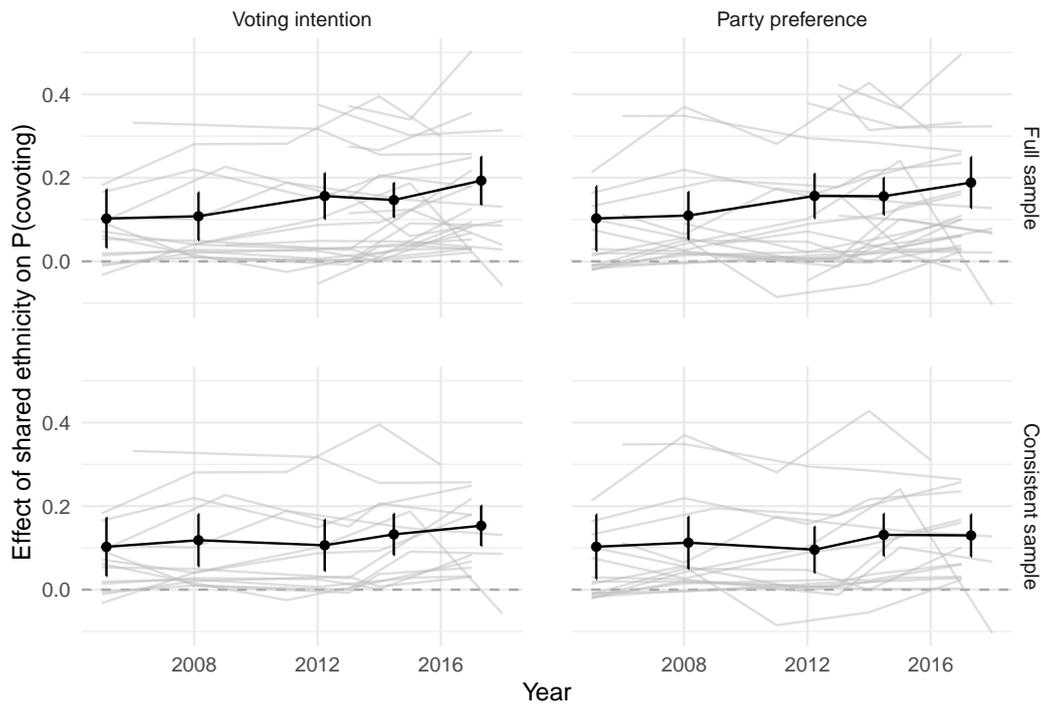


Figure 3: Effect over time, by Afrobarometer survey round

Note: Coefficients result from the fully specified model in Equation 10 estimated separately for each Afrobarometer survey round, using respondents' self-identified ethnicity to construct the coethnicity 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 A21 to A23 for full results.

3 to 7 in Figure 3 we find a slight upwards trend in the full sample, which includes increasingly many countries.³⁹ Yet, there is no significant increase in the effect of coethnicity 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 3 results mostly from the increasing sampling of countries with more extensive ethnic voting.

In contrast, the estimated effects of coethnicity on covoting 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. Figure 4 displays results across Afrobarometer rounds 3-7 for Kenya, Malawi, and Mali. Political scientists typically describe Kenyan elections as classic cases of ethnic voting (Bratton and Kimenyi 2008). Our analysis sup-

³⁹For this inter-temporal analysis, we construct the coethnicity indicator based on respondents' self-identified ethnicity, which has been consistently asked since round 3.

ports this interpretation. Over time, the estimated effect of coethnicity increased on average, with a small decline in 2016 when shared ethnicity is associated with a 30 percentage point increase in covoting intentions, almost two times our main estimate.

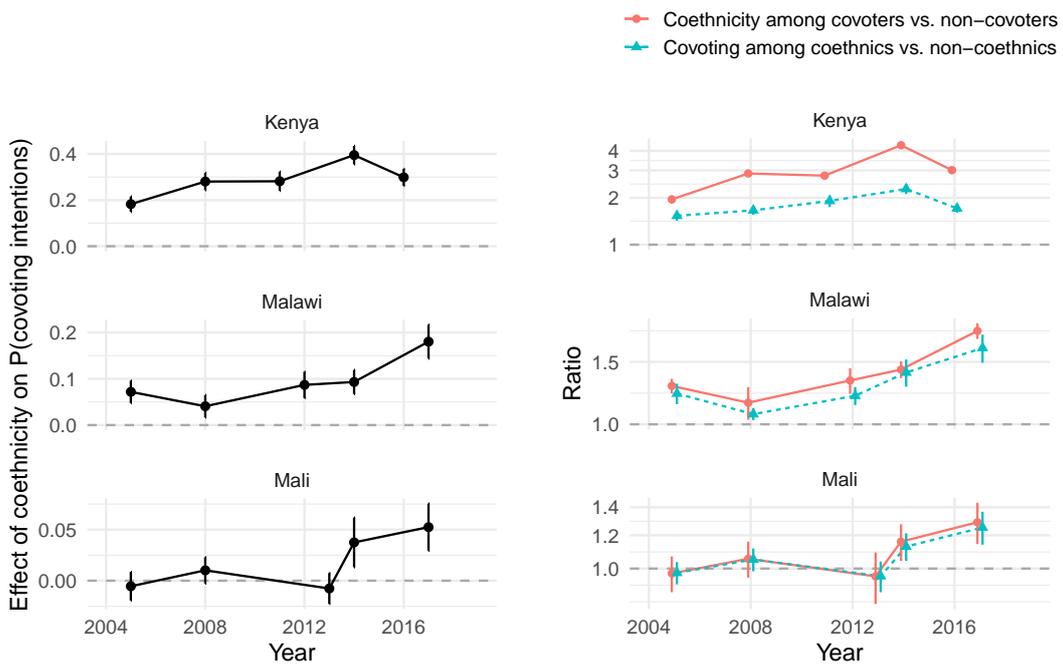
The Kenyan illustration serves well to illuminate both the comparative advantages and remaining limitations of the CVR. The coethnicity coefficient β_1 can be compared across survey rounds to reveal aggregate changes in the patterns of coethnic covoting conditional on covariates. It does not, however, reveal whether these aggregate changes result from individual changes in voting behavior, candidate entry and exit, or both. The right panel of Figure 4 shows the derived group- and party-based components of coethnic covoting. While their values can be affected by changes in other quantities besides coethnicity (β_1), including covoting (β_0) and covariate effects (γ), both measures run largely in parallel to β_1 . The results suggest that the Kenyan increase in coethnic covoting came with a greater increase in the ethnic homogeneity of candidates' voters as compared to a smaller increase in the homogeneity of groups' voting intentions.⁴⁰

Next, we turn to Malawi, a country for which prominent studies diagnosed weak ethnic voting patterns. In line with Ferree and Horowitz' analysis of Malawi, we indeed observe relatively small 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 results indicate a strengthening of the effect of coethnicity that reaches the estimated average effect for all elections in SSA in Afrobarometer round 7. Increasing voting differences between ethnic groups and rising ethnic differences between parties' voters run parallel to this trend. This development challenges recent work that identifies Malawi's persistent regional voting blocs and underlying shared economic interests as the better fitting explanation of covoting intentions (Boone et al. 2022, but see Robinson 2024).⁴¹

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

⁴⁰Such changes can be driven by changing ethnic electoral coalition dynamics and not by changes in voters' underlying preferences to support co-ethnics.

⁴¹Similarly, we estimate some of the strongest coethnicity effects on covoting intentions for Kenya (see above) and Zambia (Appendix Figure A23), Boone et al.'s other cases that serve to illustrate the influence of geography on voting blocs.



(a) Effect of coethnicity on covoting

(b) Party- and group-based derivatives

Figure 4: Ethnic voting over time in Kenya, Malawi and Mali

Note: (a) Coefficients result from the fully specified model in Eq. 10 estimated separately for each Afrobarometer survey round since Round 3. We use respondents' self-identified ethnicity as ethnicity indicator and cluster standard errors at the level of respondents i and j . See Appendix Figures A21 to A23 for all countries in the sample.

(b) Ratios are computed by dividing the covariate-adjusted rate of coethnicity among covoters by the rate of coethnicity of non-covoters (red) and by dividing the rate of covoting among coethnics by the rate of covoting among non-coethnics (blue).

identity is a poor predictor of vote choice.” Our analysis confirms Dunning and Harrison’s verdict when they wrote their study in the late 2000s before the renewed outbreak of civil war in 2012, which also produced increasing communal violence. Since then, the effect of shared ethnicity on Malian citizens’ covoting intentions increased to a comparatively small effect of 6 percentage points in 2017. This development comes with more covoting within ethnic groups and heightened ethnic differences between parties’ voter pools. We estimate similarly small effects of shared ethnicity in Burkina Faso and Senegal (see Appendix Figures A21 and A23), two other countries that Africanists have singled out as examples of weak ethnic voting (e.g., [Koter 2013a](#)). As in Mali, however, both cases see increasing covoting intentions among coethnics.

Together, these trends underline the importance of broad comparative work like ours. It is important to note that changes in the effects of coethnicity on covoting over time and differences between countries can be driven by many factors not discernible from CVR estimates such as general changes in political preferences, party systems, or electoral institutions. Explicit analyses of effect heterogeneity can, however, shed light on such drivers, as we show in exploratory analyses below.

Robustness checks

We systematically test the robustness of our results to the omission of cross-cleavage interactions, 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 F.

Saturated interaction and multinomial regression models: Our main explanatory variable “shared mother tongue” is based on an interaction between linguistic characteristics of respondents’ i and j . The main analysis might therefore be biased by the omission of substantively important cross-cleavage interactions, interactions between all characteristics (ethnicity, wealth, education, etc.) of voter i with that of

voter j (Appendix A.4). As a remedy, we estimate fully interacted models.⁴² We also compute the effects of coethnicity on covoting from multinomial choice models for each country in our sample. Appendix F.1 and F.2, respectively, show that the results closely align with the estimates from our specification with simple control variables (Eq. 6).

Accounting for potentially endogenous ethnicity: Endogenous ethnicity (Müller-Crepon 2025; Pengl, Roessler and Rueda 2022) or identity misreporting (Green 2021) might bias our main estimates. In Appendix F.3, we therefore assess whether our results hold for respondents' current home language and ethnic self-identification, both of which are more malleable to change but potentially more precise. Supporting our main results, we find that a shared home language and ethnic self-identification increase covoting by 12 and 19 percentage points, respectively. We additionally estimate the effect of the pairwise linguistic proximity between respondents, finding that covoting increases with linguistic proximity. Varying the coarseness of language classification on the basis of the phylogenetic language tree yields a positive effect of coethnicity even if it is solely based on speaking a language belonging to the same of 6 language families. In combination, these results suggest that strategic misreporting or endogenous ethnic change are unlikely to substantively affect our results.

Accounting for geographic variation: Because many ethnic groups' feature regionally distinct "homelands", coethnic voting might simply emerge from an alignment of economic interests and political preferences of individuals from the same administrative region or even location (Boone et al. 2022; Boone 2024), geographies which might have partially shaped ethnic geography itself (Posner 2005; Müller-Crepon 2025).⁴³ Appendix F.4 shows that our results are robust to adding more fine-grained controls for geographic distance and joint residence in enumeration areas, districts and regions. These do not substantively interact with coethnicity. In addition, results from estimating our CVR only for the small, non-representative

⁴²We obtain equivalent results when reducing the number of interactions via LASSO (Beiser-McGrath and Beiser-McGrath 2020).

⁴³Individuals' place of residence is, however, also endogenous to (ethnic) migration patterns.

subset of respondent pairs residing in the same region or enumeration area show that coethnicity increases covoting in these selected pairs by 10 to 14 percentage points. This estimate decreases by another 2 to 4 percentage points once we include region and enumeration fixed effects, which soak up significant variation in ethnic and political homogeneity across locations. Overall, these results suggest that ethnic cleavages – while certainly influenced by geographic factors – produce strong electoral effects that are largely independent of geography.⁴⁴

Data construction: We vary a number of choices made in the construction of our dyadic comparisons between survey respondents (see Appendix F.6). 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.⁴⁵ 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 F.7). We first reestimate the main specifications in Table 1 using logistic regression models, which yields equivalent results (Appendix Table A21).⁴⁶ 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 at the level of language groups, enumeration areas, and individual respondents on each side of a comparison. Doing so drastically im-

⁴⁴We do note, however, that this finding is consistent with plausibly strong effects of economic geography on voting patterns (Boone 2024).

⁴⁵This imbalance results from differing rates of missingness in the data.

⁴⁶Similarly, estimating a network-based Probabilistic Partition Model (Müller-Crepon, Schvitz and Cederman 2025) yields consistent results (Appendix F.8).

proves model fit but does not substantively change the estimated effect of a shared mother tongue on covoting.

Mechanisms and heterogeneous effects

We can use the CVR to further test for mechanism and heterogeneity underlying the effect of coethnicity on covoting. Building on previous theoretical and empirical approaches at the micro-, meso-, and macro-level, the following exploratory, correlational analyses show strongest support for meso-level arguments about the importance of political mobilization of ethnic groups by “ethnic” parties. Evidence for micro-level arguments about ethnic biases and shared interest and macro-level theories on the importance of political institutions is mixed. While we focus on explaining cross-sectional heterogeneity, we note that similar analyses can be used to study drivers of temporal changes in the effect of coethnicity on covoting.

Micro-level: Ethno-centrism and shared interests

Arguments in the micro-level literature can be divided into psychological explanations of ethnic voting that highlight preferential bias for coethnics (Horowitz 1985), and instrumental approaches that stress (1) expectations of ethnic favoritism (Bates 1974), (2) ethnicity as an information shortcut in low-information environments (Ferree 2006), and (3) shared preferences among coethnics (Lieberman and McClendon 2013). We find no substantive moderating effect of ingroup bias proxied as respondents’ ethnic rather than national identification (Appendix E.1). Expectations of favoritism are positively but noisily ($p < .1$) related to coethnic voting. Greater political knowledge and news consumption do not decrease the effect of coethnicity on covoting.⁴⁷

Our test of the importance of ethnically structured interests proceeds in two steps. First, we extend the CVR to test whether coethnicity increases shared perceptions of pressing problems in a country and of governments’ performance across 19 issue areas ranging from the economy to youth policy.⁴⁸ As Models 1 and 2

⁴⁷We continue to conservatively cluster standard errors at ethnic group level.

⁴⁸For the former, we assess how much respondents’ choice of the three most pressing problems

in Table 2 show, coethnicity comes with small, statistically significant increases in shared perceptions, yet these effects amount to less than 5 percent of the respective outcome means. Second, we show that shared perceptions increase the likelihood of covoting between respondents. However, these effects do not diminish the estimated effect of coethnicity in Model 3. These results suggest that individual-level explanations do not explain the bulk of coethnic covoting, at least with the proxies available.⁴⁹

Table 2: Shared mother tongue, perceptions

Dependent Variables: Model:	Problems (1)	Performance (2)	Voting intention (3)
<i>Variables</i>			
Shared mother tongue (0/1)	0.009*** (0.002)	0.014*** (0.004)	0.165*** (0.029)
Problems			0.015** (0.006)
Performance			0.193*** (0.037)
<i>Fixed-effects</i>			
Country x Round	Yes	Yes	Yes
<i>Fit statistics</i>			
Controls	Yes	Yes	Yes
Outcome mean	0.258	0.670	0.459
Dyads	7,465,789	7,535,705	6,366,828
R ²	0.035	0.041	0.082
Within R ²	0.006	0.009	0.026

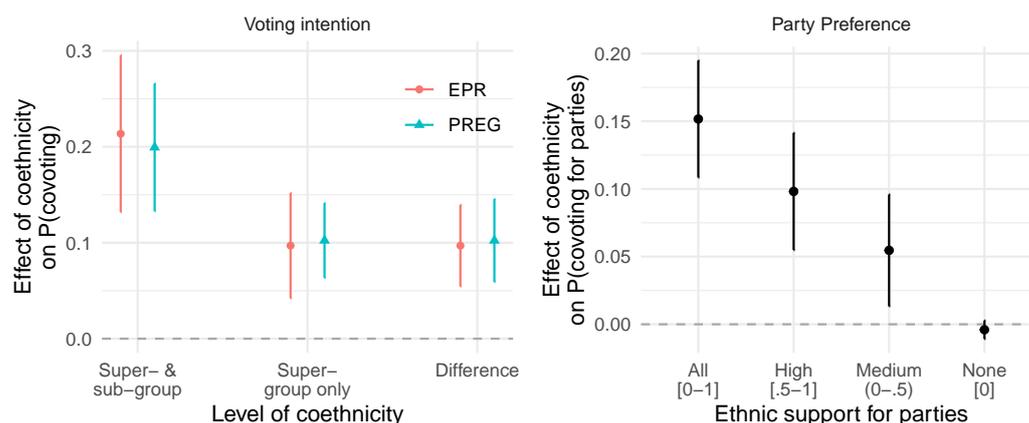
*Standard errors clustered at the level of each respondent's mother tongue in parentheses. Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Meso-level: Politicized ethnic groups and ethnic parties

At the meso-level, we examine heterogeneity in the effect of coethnicity on covoting in two regards. First, we ask whether the political mobilization of (clusters of) ethnic groups as coded by [Vogt et al. \(2015\)](#) and [Posner \(2004a\)](#) increases covoting overlap (0-1). For the latter, we code shared perceptions in a pair as 1 minus the average absolute rating difference (0-1) across issue areas.

⁴⁹Unfortunately, the Afrobarometer does not include detailed batteries on policy ideal points, which could improve the measurement of shared interests.

above and beyond the mere effect of speaking the same mother tongue. Suggesting substantive effects of ethnic groups’ political relevance, Figure 5a shows that joint membership in a politically relevant group increases covoting by 10 percentage points, even if two respondents do not share the same mother tongue. Sharing the same mother tongue increases covoting by an additional 10 percentage points.



(a) Covoting within politically relevant “super-groups” from EPR and PREG

Note: Politically relevant “super-groups” from EPR (Vogt et al. 2015) and PREG (Posner 2004a) linked to mother tongues via Müller-Crepon, Pengl and Bormann (2022). Estimates from Appendix Tables A6 and A7.

(b) Covoting for parties with differing reliance on ethnic support groups

Note: Data on ethnic support groups from the VDEM V-Party data (Lührmann et al. 2020) linked to Afrobarometer Round 6 via Döring and Regel (2019). Estimates from Appendix Table A8.

Figure 5: Variation in the effect of coethnicity on covoting by groups’ political relevance and parties’ reliance on ethnic constituencies

Second, we disaggregate the coding of covoting to reflect whether respondents share a preference for a party that the V-Party data (Lührmann et al. 2020) codes as having a high or medium level of ethnic support, or none at all.⁵⁰ The results in Figure 5b suggest that two thirds of the overall effect of coethnicity on covoting is driven by covoting for parties with a high level of ethnic support, with the remainder being due to parties with a medium level of ethnic support. We find no substantive effect of coethnicity on covoting for parties with no ethnic support base. Their voter bases are spread uniformly across ethnic groups, which shows that our results align closely with experts’ assessments of parties’ core constituencies.

⁵⁰The classification of parties derives from the share of V-Dem experts that agree that a party has an ethnic support group. We drop parties that are not included in the VDEM-Party data.

Macro-level: Political institutions

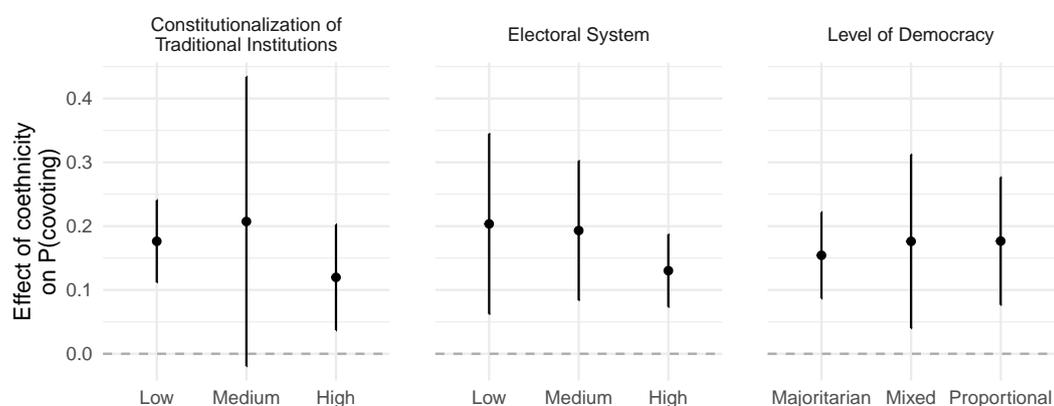


Figure 6: Effect of coethnicity on shared voting intentions by countries' institutional characteristics

Note: Coefficients result from the fully specified model in Eq. 10 estimated separately for countries with different electoral rules, levels of the V-Dem's Polyarchy index as well as constitutionalization of traditional institutions. See Appendix Tables A9 to A11 for underlying models.

Prior research highlights theoretical reasons to expect substantive variation between countries in the extent to which ethnic cleavages structure electoral politics (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 substantively or statistically significant variation in the effect of coethnicity on covoting intentions across these three arguably relevant institutional dimensions (Figure 6 and Appendix E.3). While these findings shed doubt on the importance of institutional variation, our findings can only be understood descriptively as we do not account for potential endogeneity of the moderating factors. It is also possible that existing analyses find empirical support for theoretical mechanisms that operate at the meso and macro-levels. It is particularly important to note that institutions can shape strategic candidate entry. This would introduce differential selection bias into our analysis and bias estimated differences in the effect of coethnicity on covoting.

Comparing cleavages

Finally, we compare the effect of coethnicity and that of other socio-economic similarities between respondents with respect to their rate of covoting intentions. To facilitate a fair comparison that takes account of differences between conditional and unconditional effects, Figure 7 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).

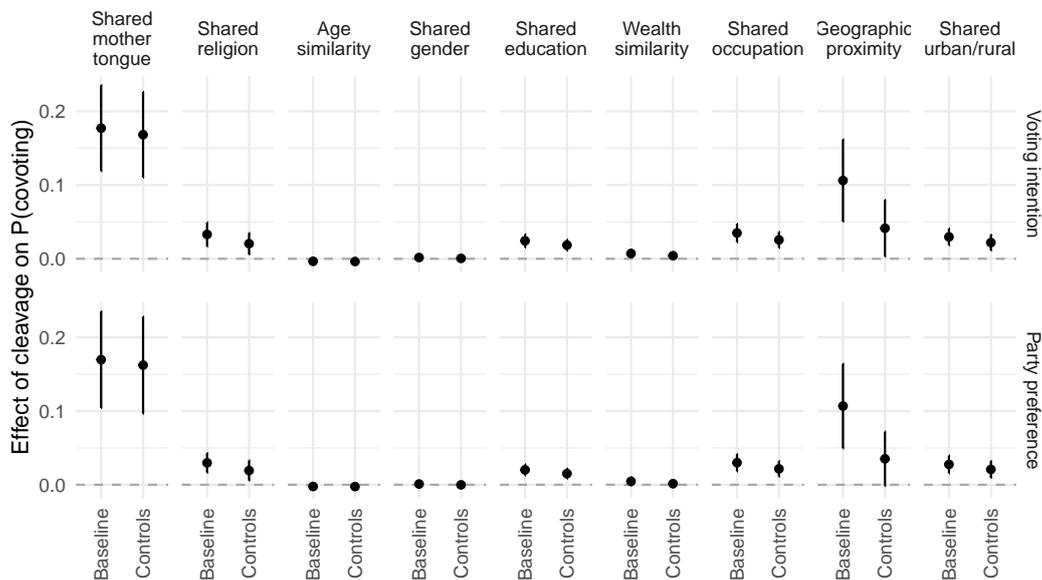


Figure 7: Results by cleavage indicator

Note: Coefficient estimates from (1) baseline model that only include the respective variable and country-fixed effects equivalent to the setup of Models 1 and 3 in Table 1, and (2) fully specified models with controls (Eq. 10), equivalent to the setup in Models 2 and 4 in Table 1. Error bars denote 95% CIs.

Among identity cleavages, shared mother-tongues seem to be by far the strongest and most stable predictor of covoting. The first column in Figure 7 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 covoting intentions, which suggests that party preferences within age groups are marginally more diverse than

across them.

Economic similarities show 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 1.5 and 2.2 percentage points, respectively. These effects are robustly estimated. Interestingly, proximity in wealth-levels between respondents does not significantly relate to covoting intentions between them.⁵¹ This finding speaks to previous findings on economic voting in Sub-Saharan Africa (Bratton and Kimenyi 2008; Bratton, Bhavnani and Chen 2012).

Lastly, we find the geographic proximity correlates with shared support for presidential candidates and parties. In the unconditional baseline models, increasing geographic proximity by 1'000km, or moving from the 97th to the 1st percentile in our data, comes with an increase in covoting intentions by 10 percentage points, consistent with the existence of regional voting blocs (Boone et al. 2022).⁵² Yet, once we condition on all other cleavage measures, the effect of geographic proximity drops by half. 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 covoting of approximately 2 percentage points when including controls. This is consistent with work on rural-urban cleavages on the continent (e.g. Harding 2010; Koter 2013b).

Conclusion

In this paper, we introduced the Covoting Regression (CVR) model 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 coethnic candidates towards shared voting intentions between two individuals allows us to address two key methodological issues in existing work. First, we reduce selection bias that plagues micro-level studies when the supply of candidates does not allow

⁵¹The estimate is robust to expanding the set of variables used to construct the wealth index (Appendix F.5).

⁵²Though note that our sample includes many small countries where such a change is unrealistic.

survey respondents to express support for coethnic candidates, or forces them to do so in the absence of non-ethnic rival candidates (e.g., [Ferree 2022](#)). The CVR does not overcome selection bias that results from strategic choices of candidates (not) to run, a problem that will be more severe under plurality and majoritarian electoral rules. By considering ethnic bloc voting, however, the CVR reduces such bias in many contexts in comparison to previous approaches. Second, we avoid ecological inference inherent in meso and macro-level research that examines ethnic covoting 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 comparisons over time and across countries without preventing 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 coethnicity 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 coethnicity does not influence covoting 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 the importance of group- and party-level characteristics in explaining the incidence of coethnic covoting, but yield no robust evidence for micro-level explanations around ethnocentric bias or policy perceptions, as well as macro-level arguments on the impact of electoral rules, democratic institutions, or traditional authorities ([Huber 2012](#); [Rabushka and Shepsle 1972](#); [Baldwin 2013](#)).

Our study opens up new avenues for studying ethnic and more generally cleavage-based political behavior in Sub-Saharan Africa and beyond. More precisely measured data on economic income, production patterns, and partisanship

would enable us to gain deeper insight into class, economic, and psychological explanations of voting – three core concerns of voting research outside Africa. For example, incorporating local economic production data could be used to approximate residents’ economic interests and model the extent to which they structure the formation of regional voting blocs that can span across ethnic divides (Boone 2024). 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 could, for example, test the effect of different social networks on voting by capturing the overlap in (the homogeneity of) social contacts.⁵³

Beyond Sub-Saharan Africa, our analytical focus on covoting 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). 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 covoting 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. Lastly, our work shows that the CVR might benefit studies of the effects of cleavages on political attitudes and behavior more generally by accommodating more outcomes than just covoting, for example, shared (sets of) political attitudes and policy preferences between individuals. After all, social and political cleavages are an inherently *relational* concept, and their effects on political preferences should be operationalized and studied as such.

⁵³Spillover 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.
- 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.
- Beiser-McGrath, Janina and Liam F Beiser-McGrath. 2020. "Problems with products? Control strategies for models with interaction and quadratic effects." *Political Science Research and Methods* 8(4):707–730.
- 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.
- 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.
- Casey, Katherine. 2015. "Crossing Party Lines: The Effects of Information on Redistributive Politics." *American Economic Review* 105(8):2410–2448.

- Chandra, Kanchan. 2007. *Why ethnic parties succeed: Patronage and ethnic head counts in India*. Cambridge University Press.
- 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.
- Cramer, Katherine J. 2016. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker*. Chicago, IL: University of Chicago Press.
- Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4):399–416.
- Davis, Justine M and Martha Wilfahrt. 2025. "Understanding the Sensitivity of Party Identification Questions in Polarized African Contexts." *Political Behavior* pp. 1–26.
- Döring, Holger and Sven Regel. 2019. "Party facts: A database of political parties worldwide." *Party politics* 25(2):97–109.
- 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.
- 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.
- 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. 2013a. "King Makers: Local Leaders and Ethnic Politics in Africa." *World Politics* 65(2):187–232.
- Koter, Dominika. 2013b. "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.
- 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.
- Lublin, David. 2017. "Electoral Systems, Ethnic Heterogeneity and Party System Fragmentation." *British Journal of Political Science* 47(2):373–389.
- Lührmann, Anna, Nils Düpont, Masaaki Higashijima, Yaman Berker Kavasogly, Kyle L Marquardt, Michael Bernhard, Holger Döring, Allen Hicken, Melis Laebens and Staffan I Lindberg. 2020. "Varieties of party identity and organization (V-Party) dataset."
- 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. 2025. "Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa." *American Journal of Political Science* 69(2):406–422.
- Müller-Crepon, Carl, Guy Schvitz and Lars-Erik Cederman. 2025. "Shaping States into Nations: The Effects of Ethnic Geography on State Borders." *American Journal of Political Science* 69(1):132–147.
- 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. 2004a. "Measuring ethnic fractionalization in Africa." *American Journal of Political Science* 48(4):849–863.
- Posner, Daniel N. 2004b. "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.
- Robinson, Amanda Lea. 2024. *The Political Logic of Cultural Revival: Ethnic Visibility, Linked Fate, and Electoral Politics in Africa*. Oxford University Press.
- Sen, Amartya. 1997. *On economic inequality*. Oxford University Press.
- Simpson, Edward H. 1949. "Measurement of diversity." *Nature* 163(4148):688–688.
- Vogt, Manuel, Nils-Christian Bormann, Seraina Rügger, Lars-Erik Cederman, Philipp M Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution* 59(7):1327–1342.
- 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.