Supplementary Online Materials to the Paper: Cash Crops, Print Technologies and the Politicization of Ethnicity in Africa

Yannick I. Pengl* Philip Roessler† Valeria Rueda‡

June 25, 2021

*International Conflict Research, ETH Zürich. ORCID: 0000-0002-0373-0952
†Department of Government, William & Mary. ORCID: 0000-0002-4935-6468
‡School of Economics, University of Nottingham, and CEPR. ORCID: 0000-0002-4967-3172

Contents

I Data Construction & Descriptives **3**

I.1 Data Sources & Analysis Samples . . . 3

I.2 Descriptive Maps & Figures . . . 5

I.3 Coding Key Variables . . . 8

II Additional Results & Robustness **12**

II.1 Robustness of Group-level results . . . 12

II.2 Magnitude of the Effects . . . 14

II.3 Alternative Definition of the Publications Treatment . . . 17

II.4 Addressing Endogeneity: DHS Inter-Marriage Models . . . 18

II.5 Additional DHS Results . . . 22

III Alternative Explanations & Plausibility of Mechanisms **25**

III.1 Group Size . . . 25

III.1.1 Group Size as Confounder . . . 25

III.1.2 Publications per Capita and Group Size . . . 30

III.2 Mediation Models . . . 31

III.3 Treatment Interactions . . . 35

III.4 Heterogeneity by Colonizer . . . 39

III.5 Disaggregating Crops by Mode of Production . . . 41

III.6 Local Ethnic Diversity . . . 46

III.7 Ethnic or Religious Boundary Making? . . . 50

IV Description of Pre-Registered Analyses and Supplementary Analyses **53**

IV.1 Pre-registered Analyses . . . 53

IV.1.1 Trust . . . 56

IV.1.2 Bloc - Voting . . . 57

IV.2 Ethnic Conflict . . . 58

IV.3 Supplementary, Non-preregistered Analyses . . . 61

I Data Construction & Descriptives

I.1 Data Sources & Analysis Samples

Table I.1: Data Sources

(1)	(2) Name	(3) Purpose	(4) Match to Ethnologue
Unit of analysis	Ethnologue	Unit of analysis	–
Print technologies	Rowling and Wilson (1923)	Treatment	Ethnic
	Mann and Sanders (1994)	Treatment	Ethnic
	Cagé and Rueda (2016)	Control	Spatial
Cash crops	Hance, Kotschar and Peterec (1961)	Treatment	Spatial
Ethnic characteristics	Afrobarometer	Outcomes (ethnic vs. national identity salience, inter-ethnic trust and bloc-voting)	Ethnic & Spatial
	DHS	Outcomes (inter-ethnic marriage)	Ethnic & Spatial
	EPR & PREG	Outcomes (politically salient groups)	Ethnic
	Murdock (1959)	Controls (historical)	Ethnic and Spatial

Notes: This table summarizes the data sources used for this analysis. Column (1) states the type of data, (2) cites the sources, (3) states the purpose, and (4) the type of match required to merge to Ethnologue. A "Spatial" merge refers to a merge to Ethnologue based on the spatial correspondence between geographic information in data (2) and the ethnic homelands mapped in Ethnologue. An "Ethnic" merge refers to a merge between the ethnolinguistic group in (2) and in its linked ethnolinguistic name in Ethnologue.

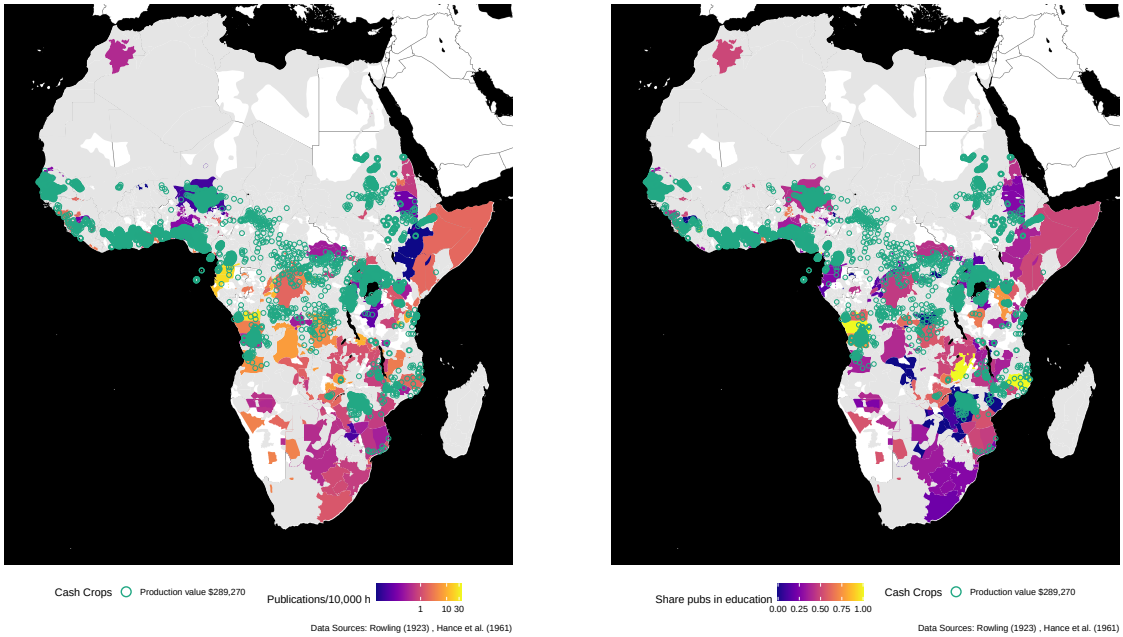
Table I.2: Data Sources

Country	EPR	PREG	Afrobarometer	DHS
Angola	X	X		
Benin	X	X		X
Botswana	X	X	X	
Burkina Faso	X	X	X	X
Burundi	X	X	X	X
Cameroon	X	X	X	X
Cape Verde	X		X	
Central African Republic	X	X		X
Côte d'Ivoire	X	X	X	X
DR of Congo	X	X		X
Ethiopia	X	X	X	X
Gabon	X	X	X	X
Ghana	X	X	X	X
Gambia	X	X		
Guinea	X	X	X	X
Kenya	X	X	X	X
Lesotho	X	X	X	
Liberia	X	X	X	X
Madagascar	X	X	X	X
Mali	X	X	X	X
Mauritius	X	X	X	
Malawi	X	X		X
Mozambique	X	X	X	X
Namibia	X	X	X	X
Niger	X	X	X	X
Nigeria	X	X	X	X
Sao Tome and Principe			X	
Senegal	X	X	X	X
Sierra Leone	X	X	X	X
South Africa	X	X	X	
Swaziland	X	X	X	
Tanzania	X	X	X	X
Tchad	X	X		X
Togo	X	X	X	X
Uganda	X	X	X	X
Zambia	X	X	X	X
Zimbabwe	X	X	X	

Notes: This table summarizes the countries that are covered in the main data sources used to measure ethnic politicization, ethnic salience, and boundary making. Countries present in the respective database and with non-missing information on geographic location and ethnic identities (AB & DHS) are marked with an "X".

I.2 Descriptive Maps & Figures

Figure I.1: Description of Publications in Rowling and Wilson (1923)



(a) Publications per 10,000 inhabitants

(b) Share of education publications

Notes: Language homelands are mapped according to Ethnologue. Greyed regions are those for which there is no record of publications. Publications per 10,000 inhabitants in map (a) are computed using population estimates in Rowling and Wilson (1923). The share of publications in education is based on the categorization in Rowling and Wilson (1923). Each green circle locates $289,270$ USD (1957) cash-crop export value.

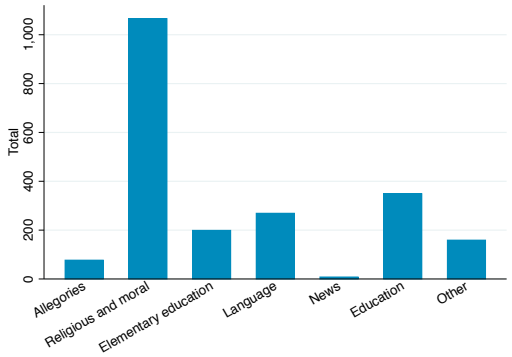
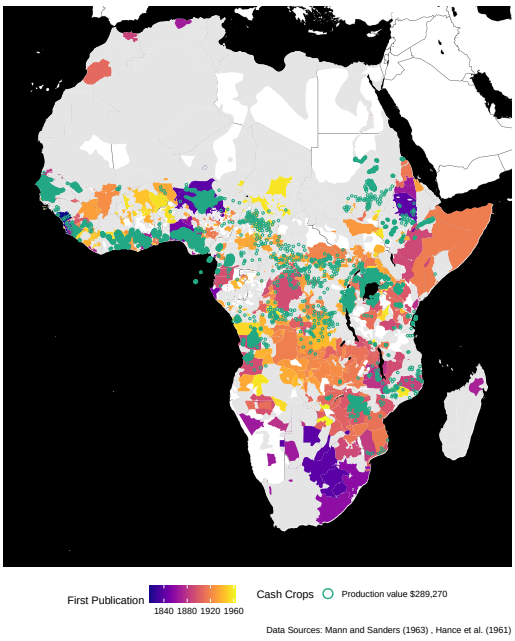


Figure I.2: Christian publications in African languages up to 1923

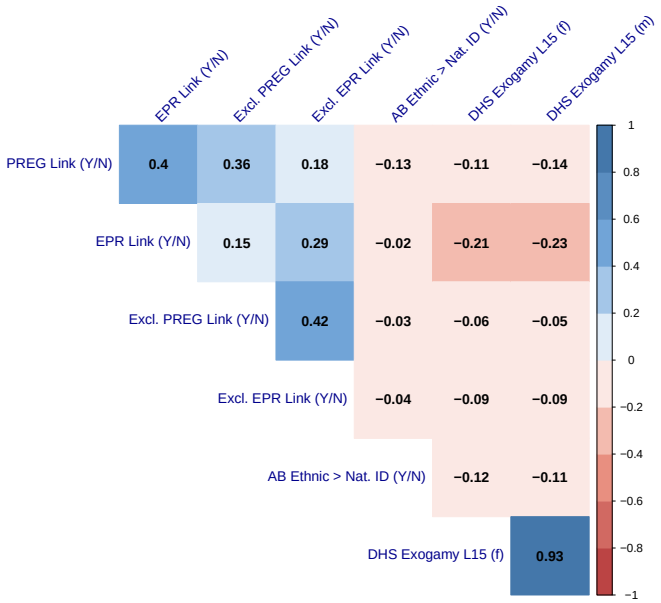
Notes: This chart gives the total number of Christian publications in African languages by type up to 1923, as described in Rowling and Wilson (1923)

Figure I.3: Publications and Cash Crop locations in Mann and Sanders (1994)



Notes: Language homelands are mapped according to Ethnologue. Greyed regions are Ethnologue polygons for which there is no record of publications. The colors indicate the number of publications listed in Mann and Sanders (1994). Each green circle locates 289, 270 USD (1957) cash-crop export value.

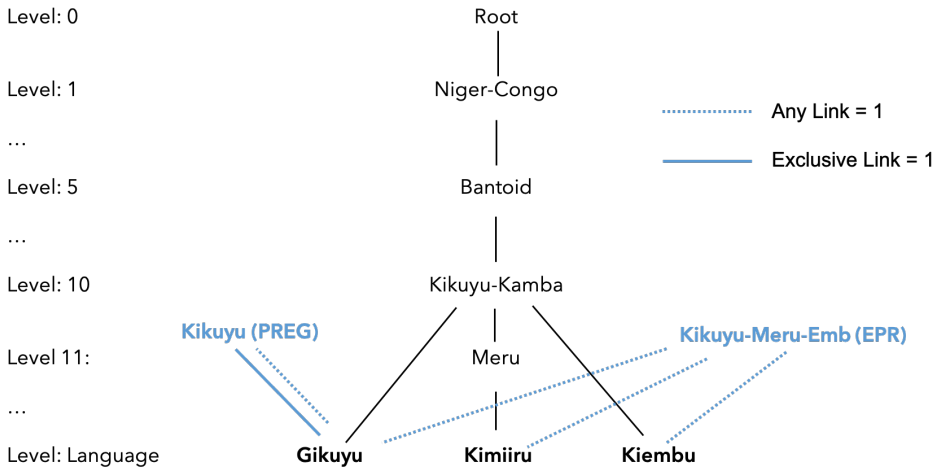
Figure I.4: Political Relevance, Ethnic Salience & Boundaries



Notes: This figure shows pairwise correlations of our ethnicity outcomes with Afrobarometer identity salience and DHS inter-marriage rates averaged at the level of Ethnologue language groups. Both male and female respondents from politically relevant groups (according to PREG and EPR) are on average less likely to marry across group boundaries (rows 1-4, columns 5-6) but correlation coefficients remain relatively small. Similarly, the average DHS respondent from groups with higher average levels of Afrobarometer ethnic identity salience is somewhat less likely to inter-marry (row 5). The expert-coded group-level proxies for political relevance are positively but far from perfectly correlated (rows and columns 1-3). Somewhat counterintuitively, political relevance is negatively (though weakly) correlated with Afrobarometer identity salience (column 4). Overall, these correlations illustrate that our conceptual distinction between ethnic politicization, salience, and boundaries is adequately reflected in our data. At the same time, the consistent results across weakly correlated outcomes increase our confidence that our results are not due to any biases or coding errors in the individual data sources.

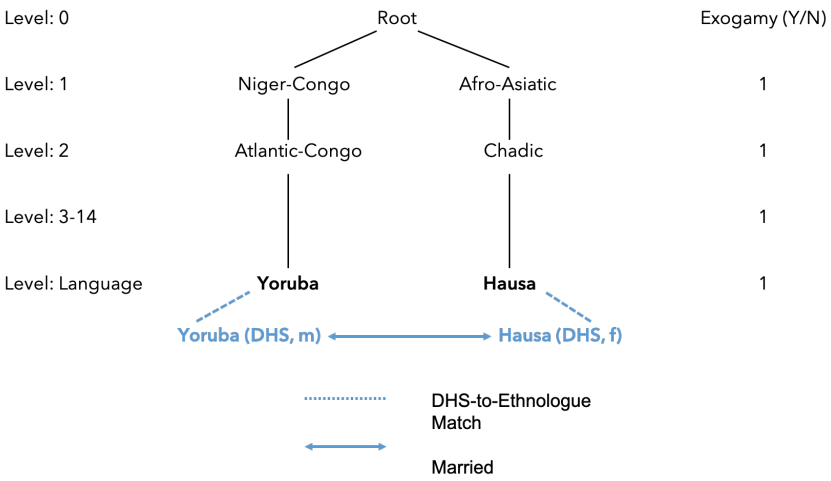
I.3 Coding Key Variables

Figure I.5: Coding Political Relevance at the Group Level



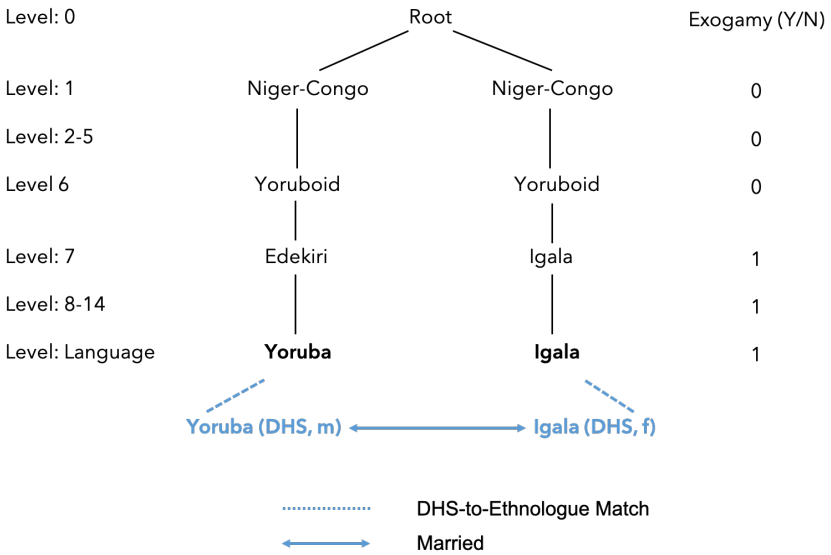
Notes: Kenyan example on how we assign the ethnic relevance outcome from PREG or EPR to Ethnologue language groups. All Ethnologue linguistic groups with no plausible match to a politically relevant group or coalition in PREG or EPR are coded zero on all four outcome variables ("Any Link (PREG)", "Any Link (EPR)", "Exclusive Link (PREG)", "Exclusive Link (EPR)"). Ethnologue groups with a one-to-one match in PREG/EPR receive a 1 on both the "Any Link" and the "Exclusive Link" outcome variable for the respective dataset (e.g. the Gikuyu appearing as Kikuyu in PREG). Where more than one Ethnologue group is matched to a single umbrella group or coalition in PREG/EPR, all of these Ethnologue groups receive a 1 on the "Any Link" outcome but a 0 on the "Exclusive Link" outcome (e.g. the Gikuyu, Kimiiru, and Kiembu which are all matched to "Kikuyu-Meru-Emb" in EPR).

Figure I.6: Coding Inter-Ethnic Marriages 1/2



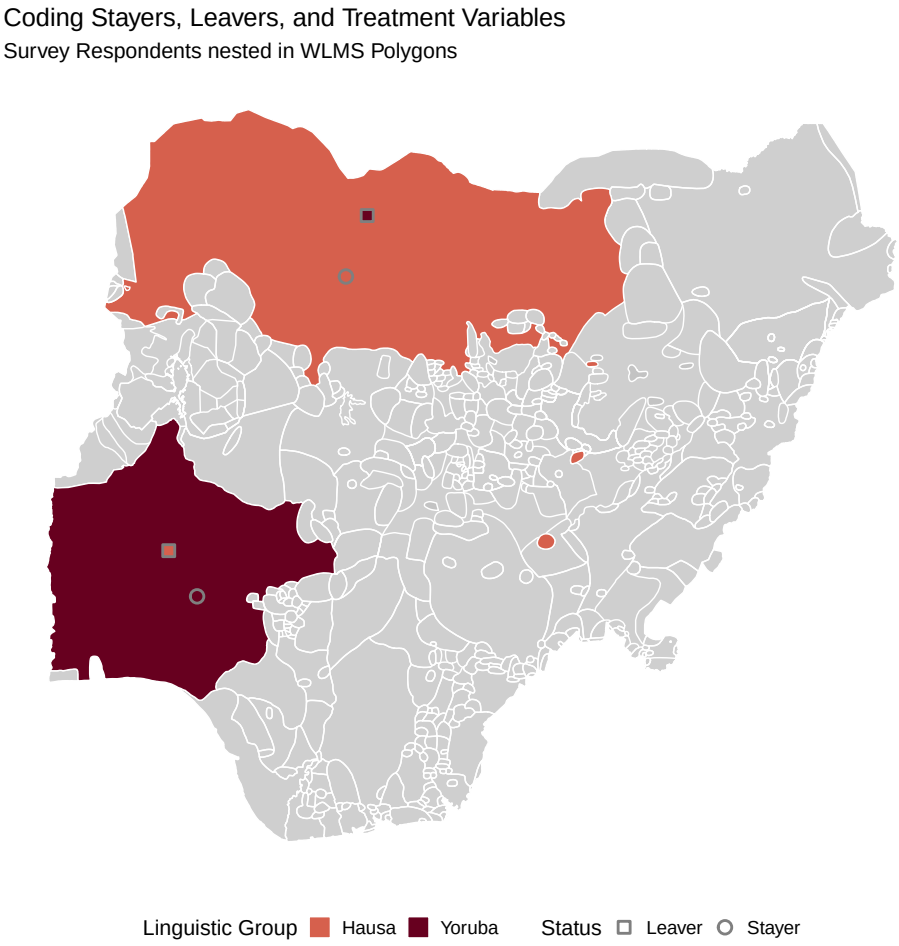
Notes: Nigerian example of how we assign the inter-ethnic marriage outcome across different levels of the Ethnologue language tree: A Yoruba-Hausa marriage is coded as exogamous across all lingusitic levels of aggregation

Figure I.7: Coding Inter-Ethnic Marriages 2/2



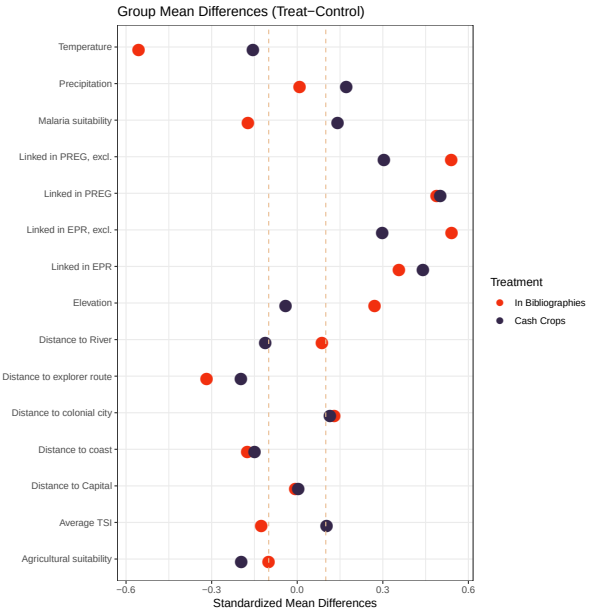
Notes: Nigerian example of how we assign the inter-ethnic marriage outcome across different levels of the Ethnologue language tree: A Yoruba-Igala marriage is coded as endogamous on levels 1-6, but as exogamous on levels 7-15

Figure I.8: Ethnic Stayers & Leavers



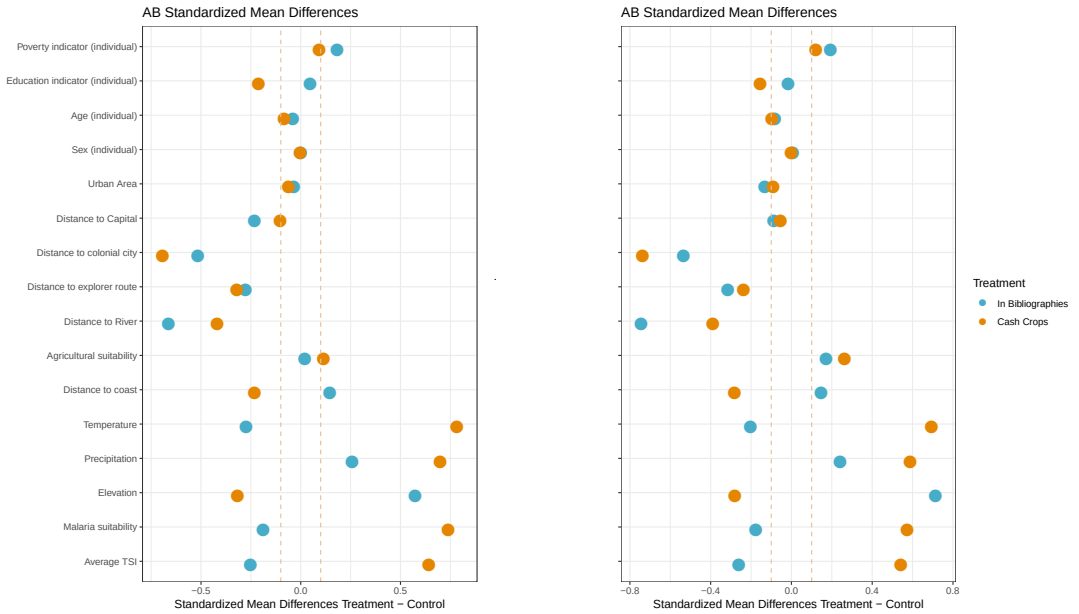
Notes: Nigerian example of how we define ethnic leavers and stayers. Treatments are assigned either based on the polygon a respondent resides in (geographic models) or based on the respondent's stated ethnic identity (ethnic models). In the geographic specifications, both the Yoruba stayer and the Hausa leaver residing in the Yoruba polygon receive the number of Yoruba publications (normalized by historical population) and both respondents living in the Hausa polygon receive the number of Hausa publications. In the ethnic specifications with location fixed effects, both the Yoruba stayer in the Yoruba polygon and the Yoruba leaver in the Hausa polygon are assigned the number of publications in Yoruba, whereas both respondents identifying as Hausa receive the number of Hausa-language publications. In the leavers-only specifications, we drop all ethnic stayers in the sample and compare e.g. the Yoruba leaver in the Hausa polygon to Igbo and Igala respondents surveyed in the same location.

Figure I.9: Balance of Covariates for Group-Level Data



Notes: The figure shows the standardized mean differences in covariates in the group-level analysis. In each graph, dots of different colors represent the different treatments used to compare means: print technology (group listed in bibliographies), or exposure to cash crop technology. The dashed vertical lines represent the conventional 0.1 threshold for assessing covariate imbalance

Figure I.10: Balance of Covariates for Afrobarometer



(a) Geographic level treatment

(b) Ethnic-level treatment

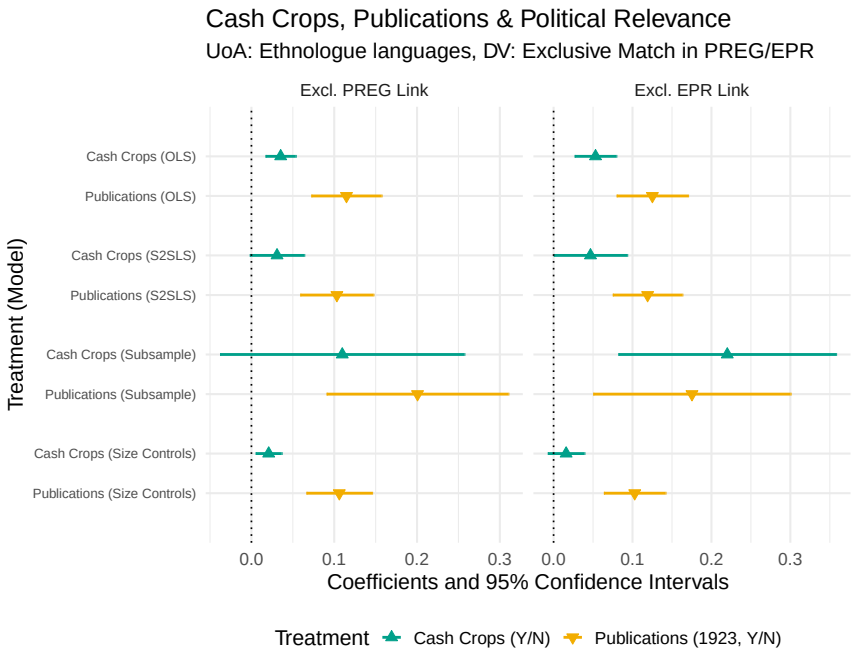
Notes: The figure shows the standardized mean differences in covariates in Afrobarometer. In each graph, dots of different colors represent the different treatments used to compare means: print technology (group listed in bibliographies), or exposure to cash crop technology. The dashed vertical lines represent the conventional 0.1 threshold for assessing covariate imbalance

II Additional Results & Robustness

II.1 Robustness of Group-level results

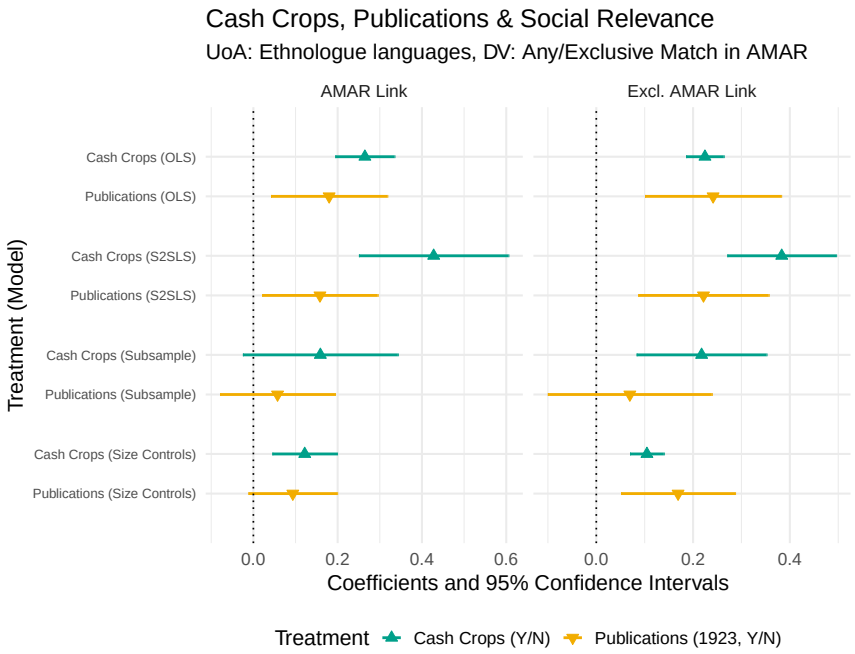
This section presents robustness of the group-level analysis to alternative definitions of the outcome variable.

Figure II.11: Cash Crops, Print Technologies, and Political Relevance - Exclusive Links



Notes: These figures summarize the results of eight regression models. Each column gives the results for one of two binary outcomes. The outcomes flag an exclusive (1-to-1) match between an Ethnologue group and an entry in the EPR and PREG databases, respectively. Different treatment specifications are shown in lines. The first two lines report OLS estimates using binary treatments indicating whether Ethnologue groups were exposed to cash crop production and/or print technologies. In lines 3-4, cash crops are instrumented with the mean agro-climatic suitability for the five most important export crops using the spatial 2SLS approach described in the text. In lines 5-6, the sample is restricted to Ethnologue polygons that experienced missionary activity. Lines 7-8 include a logged estimate of historical polygon population based on the HYDE data set.

Figure II.12: Cash Crops, Print Technologies, and Political Relevance - AMAR



Notes: These figures summarise the results of eight regression models. Each column gives the results for one of two binary outcomes. "AMAR Link" flags whether an Ethnologue group is non-exclusively matched to a group listed as socially relevant in AMAR. "Excl. AMAR Link" flags an exclusive (1-to-1) match between the same databases. Different treatment specifications are shown in lines. The first two lines report OLS estimates using binary treatments indicating whether Ethnologue groups were exposed to cash crop production and/or print technologies. In lines 3-4, cash crops are instrumented with the mean agro-climatic suitability for the five most important export crops using the spatial 2SLS approach described in the text. In lines 5-6, the sample is restricted to Ethnologue polygons that experienced missionary activity. Lines 7-8 include a logged estimate of historical polygon population based on the HYDE data set.

II.2 Magnitude of the Effects

This section presents results that allow comparing treatment effect magnitudes to other covariates. Table II.3 compare coefficients on our treatment variables to individual-level "modernization" proxies that Robinson (2014) has shown to be important predictors of more national identity salience. We observe that in the geographic specification, the effect of a one standard deviation increase in the cash crop treatment is roughly 30% the magnitude of the effect of the urban dummy, 25% that of the female dummy, or 20% that of a dummy indicating formal employment. The effect of a one standard deviation change in the publication treatment is roughly the same size as the urban dummy, 68% the effect of the female dummy, and 56% the effect of formal employment. In the ethnic-level specification, the effect of a one standard deviation change in the publications treatment amounts to roughly 25% of the effect of the female dummy, and 68% of the formal employment dummy (although formal employment no longer has a statistically significant effect in this specification).[1] Across specifications, the effect of education, proxied with a binary variable equal to one for individuals with at least some high-school education is much larger than the rest of covariates.

Tables II.5 to II.7 show the relative magnitude of our cash crop and publication coefficients compared to other important predictors of inter-ethnic marriages. Specific cell values in these tables are calculated by dividing our coefficient of interest (cash crops or publication) by the coefficient of another important covariate in the same model. The covariates that we use in this exercise include standardized female and male education years (*Educ* (f) and *Educ.* (m)), binary indicators for non-agricultural employment (*Modern Occ.* (f), *Modern Occ.* (m)), a standardized asset-based household wealth score, an urban residence dummy, and standardized level of precolonial political centralization based on the Murdock polygon a surveyed couple resides in. Overall, our coefficients of interest are rarely below half the size and frequently larger than those of important other covariates. Thus, our historical treatments have similarly large effects on inter-marriage as important contemporary factors such as education, occupation, or wealth. The main exception is the urban dummy in the geographic specifications. Ethnic exogamy in Sub-Saharan Africa remains a predominantly urban phenomenon and occurs, across all Ethnologue level, about twice as often in urban than in rural survey locations.

[1] As the cash crop treatment does not have a robust effect in the ethnic-level specification in Afrobarometer, we do not comment on its estimated effects in columns (3) and (4).

Table II.3: Afrobarometer - Persistence in Ethnic Identity - Magnitude of the effect

	Geograhic-level		Ethnic-level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.013***	0.013**		
	(0.005)	(0.006)		
Pubs pth pop (1923)	0.035***	0.049***		
	(0.007)	(0.009)		
Urban	−0.036***	−0.027***		
	(0.010)	(0.011)		
Female	0.051***	0.043***	0.052***	0.045***
	(0.006)	(0.007)	(0.007)	(0.008)
Formal Employment	−0.062***	−0.068***	−0.019	−0.022
	(0.012)	(0.014)	(0.013)	(0.015)
Education	−0.121***	−0.126***	−0.123***	−0.129***
	(0.008)	(0.010)	(0.009)	(0.011)
Cash crops USD pkm2			−0.014	−0.039***
			(0.009)	(0.012)
Pubs pth pop (1923)			0.013**	0.013*
			(0.006)	(0.008)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	Yes	Yes
Fixed Effect	Country-Round	Country-Round	Town	Town
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.389	0.38	0.389	0.38
Observations	91,832	65,408	88,962	63,242
R^2	0.042	0.046	0.232	0.224

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a standardized binary variable equal to one if respondent declares a stronger ethnic than national identity. Treatments are defined at the location level (columns (1) and (2)) and ethnic level (columns (3) and (4)). The table reports "beta" coefficients for continuous variables (the cash crop and publication treatments). Binary controls (urban, female, formal employment, and education) are not standardized.

Table II.4: Cash Crop Coefficient Relative to Other Variables (Geographic Models)

Exogamy Level	Educ. (f)	Educ. (m)	Modern Occ. (f)	Modern Occ. (m)	HH Wealth	Urban	Precol. Centr.
L1	6.86	4.15	-0.38	-1.29	-0.95	-0.2	-41.22
L2	-90.15	-2.13	-0.74	-1.52	-0.81	-0.21	2.43
L3	16.46	-3.03	-0.78	-1.53	-0.85	-0.24	2.94
L4	32.02	-1.84	-0.67	-1.12	-0.46	-0.18	1.74
L5	-1.28	-1.13	-0.59	-0.81	-0.43	-0.17	0.83
L6	-0.83	-0.76	-0.57	-0.48	-0.32	-0.12	0.53
L7	-0.95	-0.85	-0.65	-0.56	-0.4	-0.15	0.92
L8	-0.93	-0.83	-0.72	-0.56	-0.41	-0.15	0.89
L9	-0.8	-0.82	-0.71	-0.49	-0.37	-0.12	2.96
L10	-0.57	-0.75	-0.51	-0.45	-0.28	-0.1	5.01
L11-14	-0.56	-0.75	-0.51	-0.46	-0.27	-0.1	9.52
L15	-0.45	-0.61	-0.39	-0.39	-0.21	-0.08	2.2
L16	-0.48	-0.62	-0.4	-0.4	-0.22	-0.09	2.67

Table II.5: Publication Coefficient Relative to Other Variables (Geographic Models)

Exogamy Level	Educ. Years (f)	Educ. Years (m)	Modern Occ. (f)	Modern Occ. (m)	HH Wealth	Urban	Precol. Centr.
L1	8.48	5.12	-0.47	-1.59	-1.17	-0.25	-50.96
L2	-113.49	-2.68	-0.93	-1.91	-1.02	-0.27	3.06
L3	15.6	-2.88	-0.74	-1.46	-0.8	-0.23	2.78
L4	38.36	-2.2	-0.81	-1.34	-0.55	-0.22	2.08
L5	-1.61	-1.41	-0.74	-1.02	-0.55	-0.21	1.05
L6	-1.25	-1.16	-0.86	-0.73	-0.49	-0.18	0.8
L7	-1.17	-1.04	-0.8	-0.68	-0.5	-0.18	1.13
L8	-1.14	-1.02	-0.89	-0.69	-0.51	-0.19	1.09

Table II.6: Cash Crop Coefficient Relative to Other Variables (Ethnic Models)

Exogamy Level	Educ. Years (f)	Educ. Years (m)	Modern Occ. (f)	Modern Occ. (m)	HH Wealth
L1	-6.16	1.33	-0.19	-0.52	-10.36
L2	0.33	0.56	0.06	0.16	0.33
L3	-3.29	-4.86	-0.43	-1.14	-1.99
L4	-8.07	-11.96	-1.16	-4.03	-3.71
L5	-3.56	-25.52	-1.41	-4.38	-6.67
L6	-2.41	-29.51	-1.26	-2.13	-5.09
L7	-2.86	-21.9	-1.78	-2.07	-5.29
L8	-2.84	-22.52	-1.93	-2.09	-5.27
L9	-2.19	-12.21	-2.8	-1.58	-14.44
L10	-1.71	-10.18	-1.87	-1.81	-33.97
L11-14	-1.55	-10.75	-1.84	-1.82	-43.13
L15	-1.53	-7.96	-1.76	-1.84	50.4
L16	-1.53	-8.1	-1.78	-1.84	48.6

Table II.7: Publication Coefficient Relative to Other Variables (Ethnic Models)

Exogamy Level	Educ. Years (f)	Educ. Years (m)	Modern Occ. (f)	Modern Occ. (m)	HH Wealth
L1	-35.05	7.55	-1.07	-2.97	-58.96
L2	-4.62	-7.84	-0.89	-2.23	-4.59
L3	-4.93	-7.29	-0.65	-1.71	-2.98
L4	-3.22	-4.77	-0.46	-1.61	-1.48
L5	-0.89	-6.36	-0.35	-1.09	-1.66
L6	-0.72	-8.88	-0.38	-0.64	-1.53
L7	-0.88	-6.74	-0.55	-0.64	-1.63
L8	-0.89	-7.08	-0.61	-0.66	-1.65

II.3 Alternative Definition of the Publications Treatment

In this section, the publications treatment is defined as the number of publications in Mann and Sanders (1994). As ethnicities present in Mann and Sanders (1994) differ from those in Rowling and Wilson (1923), we cannot use the early estimates of ethnic group size to normalize the number of publications. As a solution, we normalize with the *contemporary* ethnic group size as reported in *Ethnologue*. Therefore, it is important to bear in mind the presence of this source of error when interpreting the results. In the Afrobarometer ethnic salience specifications, the results are only robust in the ethnic-level specification (columns (3) and (4) in Table II.8), which is our most demanding specification (as it includes town-level fixed effects).

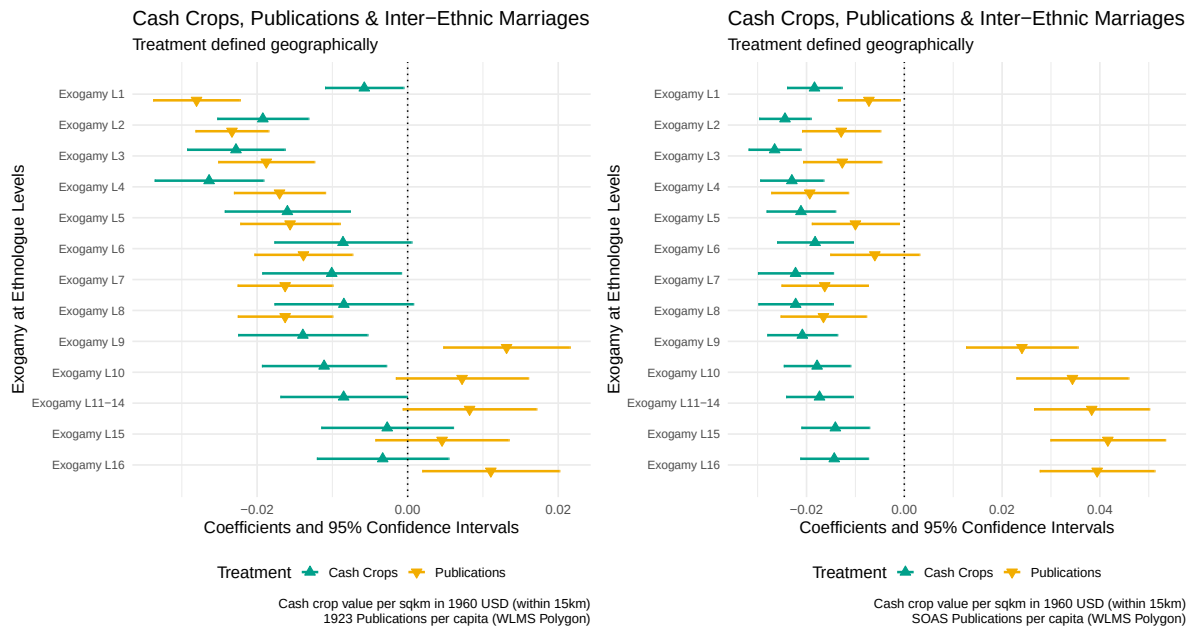
In the DHS, we also test for the robustness of the cash crop and publications effects. Figure II.13 shows two alternative specifications. First, we define cash crops and all control variables at the ethnic polygon level rather than the more local DHS enumeration areas. Second, we use our alternative publishing data based on Mann and Sanders (1994), and described above. The results remain robust. Both of these specifications yield substantively similar results and an even more striking contrast between cash crops and publishing, as the publication coefficients become positive and significant at more disaggregate levels of the language tree.

Table II.8: Afrobarometer - Alternative Publications Treatment

	Geograhic-level		Ethnic-level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.011***	0.008		
	(0.004)	(0.005)		
Pubs pth today (Mann and Sanders)	−0.001	−0.001		
	(0.004)	(0.005)		
Cash crops USD pkm2			−0.011	−0.021**
			(0.007)	(0.010)
Pubs pth today (Mann and Sanders)			0.008*	0.011*
			(0.005)	(0.006)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	Yes	Yes
Fixed Effect	Country-Round	Country-Round	Town	Town
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.1314	0.13	0.1314	0.13
Observations	125,114	87,705	120,630	85,131
R^2	0.042	0.048	0.203	0.194

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. Standard errors are reported in and clustered at the location level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. Treatments are defined at the location level (columns (1) and (2)) and ethnic level (columns (3) and (4)). The treatment is the number of publications listed in Mann and Sanders (1994), normalised by a current estimate of population speaking the language, according to *The Joshua Project* (JP).

Figure II.13: DHS - Alternative Publications Treatment



Notes: The figure reports standardized beta coefficients from 16 OLS models with country-survey-round fixed effects. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the main variables of interest ((i)cash crop production per sqkm within the WLMS polygon the survey location is situated in and (ii) publications per capita in 1923 in the language of the WLMS polygon the survey location is situated in.)

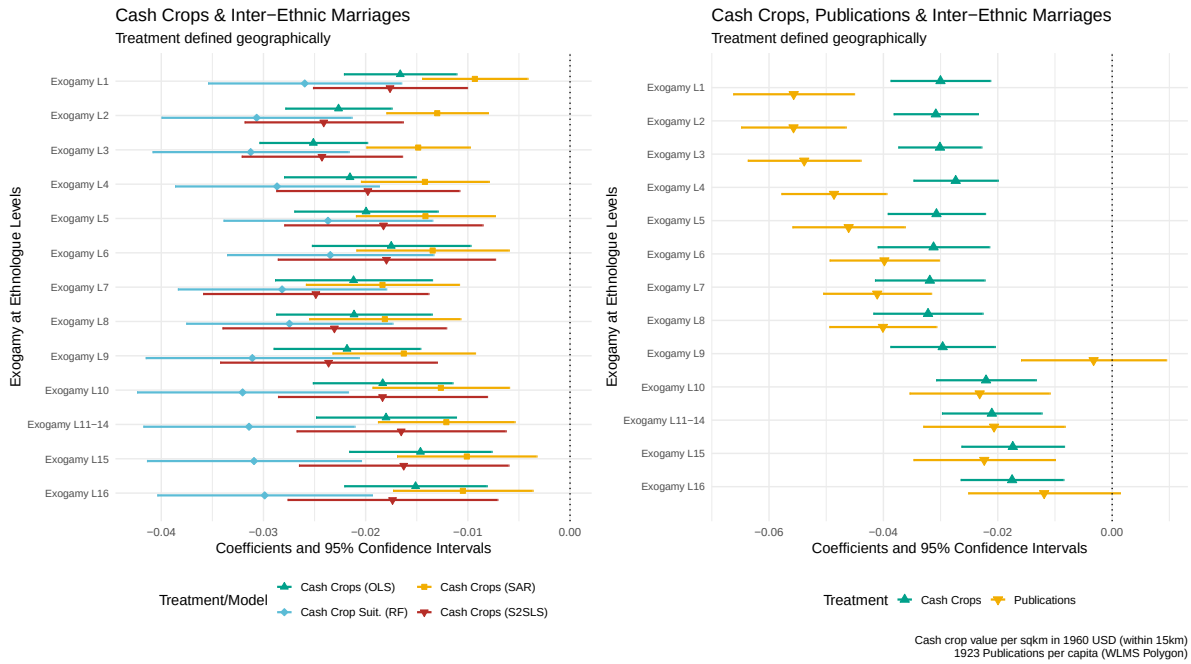
II.4 Addressing Endogeneity: DHS Inter-Marriage Models

We perform additional tests to address endogeneity concerns about our DHS inter-marriage models. First, we replicate the geographic persistence analysis for cash crops using agro-climatic suitability in reduced form and spatial instrumental variable specifications. The spatial IV analyses translate the logic of the group-level approach described in the main text to the DHS setup with individual couples nested in survey locations. We use mean cash crop suitability in survey location ℓ as an instrument and also include a spatial lag of the location mean of inter-ethnic marriages at Ethnologue level d instrumented as described above with first and second-degree spatial lags of geographic baseline controls (**?**). Again, we use a binary neighborhood matrix with a distance cutoff at 100 km. The main identifying assumption is the exclusion restriction, requiring that conditional on all covariates and spatial terms, the suitability instrument only affects intermarriage rates through its impact on observed historical cash crop production as captured in our data. We argue that this restriction plausibly holds, as the suitability measure is unaffected by historical economic activity and is unlikely to pick up non-cash crop related agricultural advantages, as we control for general agricultural suitability. A remaining concern is that suitability may have caused cash crop production in other locations than those depicted on our 1957 map and thereby affected ethnic

marriages through earlier or later adoption for cash crops. We therefore also report results from reduced form suitability models that do allow cash crop potential to affect outcomes through production in areas that are not depicted on our map. The left-hand panel in Figure II.14 presents coefficients from baseline OLS, spatial lag, reduced form, and spatial IV models. Reduced form (standardized mean suitability across the five most important export crops) and spatial IV estimates remain similar if somewhat larger than OLS, whereas coefficients in the spatial lag models get slightly smaller. The robustness of findings in these models suggest that the cash crop results are unlikely to be explained by unobserved confounding or spatially correlated outcomes (Kelly, 2019).

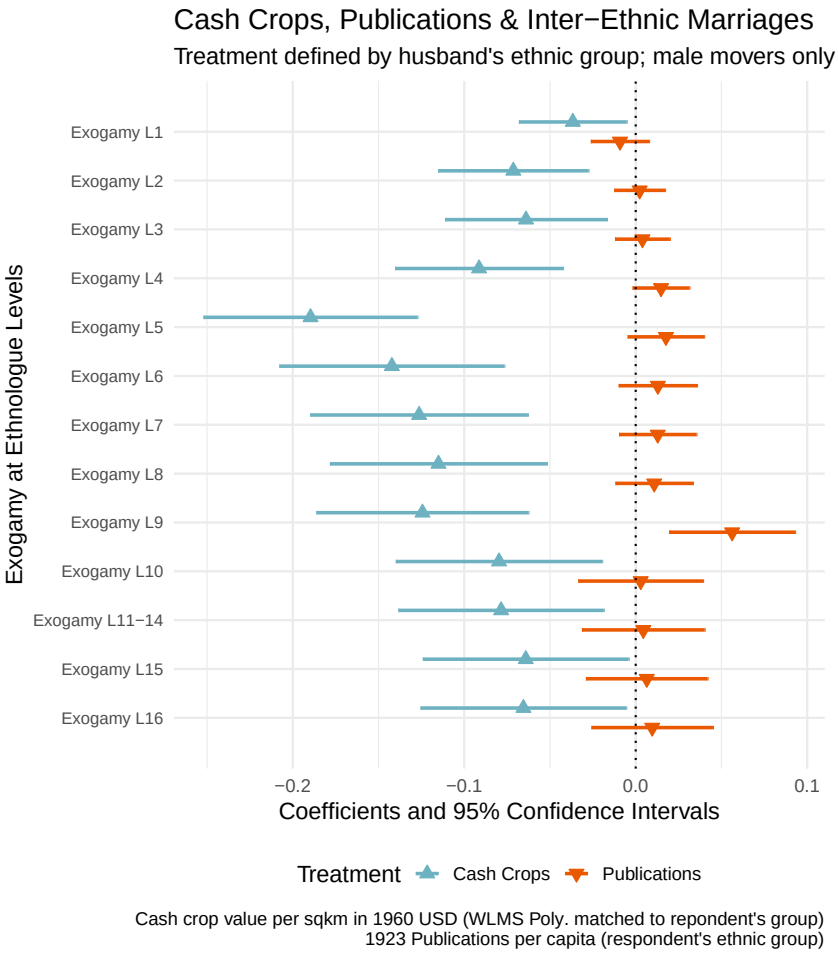
Second, we address the potentially endogenous assignment of missionary language standardization to large and already mobilized ethno-linguistic groups by running intensive margin only analyses. The right-hand panel in Figure II.14 restricts the sample to DHS couples residing in WLMS polygons with at least one Christian text in our 1923 dataset. The publication coefficients are again negative and significant at higher levels of the language tree and get smaller in size but, this time, remain significant at more fine-grained levels of linguistic differentiation. Figure II.15 shows intensive margin version of our leavers only cultural persistence models that assign treatment by respondents' self-reported ethnic identity. The left-hand panel is based on a subsample of male ethnic movers whose self-reported ethnic group had at least one publication in 1923. The publication effects largely disappear in this demanding specification.

Figure II.14: Geographical persistence - Endogeneity?



Notes: The figures probe the robustness of the geographic persistence models in the main text. The left-hand panel present findings from spatial lag models, reduced form models replacing historically observed cash crop production with agro-climatic suitability scores, and spatial IV models instrumenting historical production with agro-climatic suitability. The right-hand panel shows results from intensive margin models that restruct the analysis sample to DHS enumeration areas located within WLMS polygons of languages with at least one vernacular publication in 1923.

Figure II.15: Cultural persistence – Intensive margin models



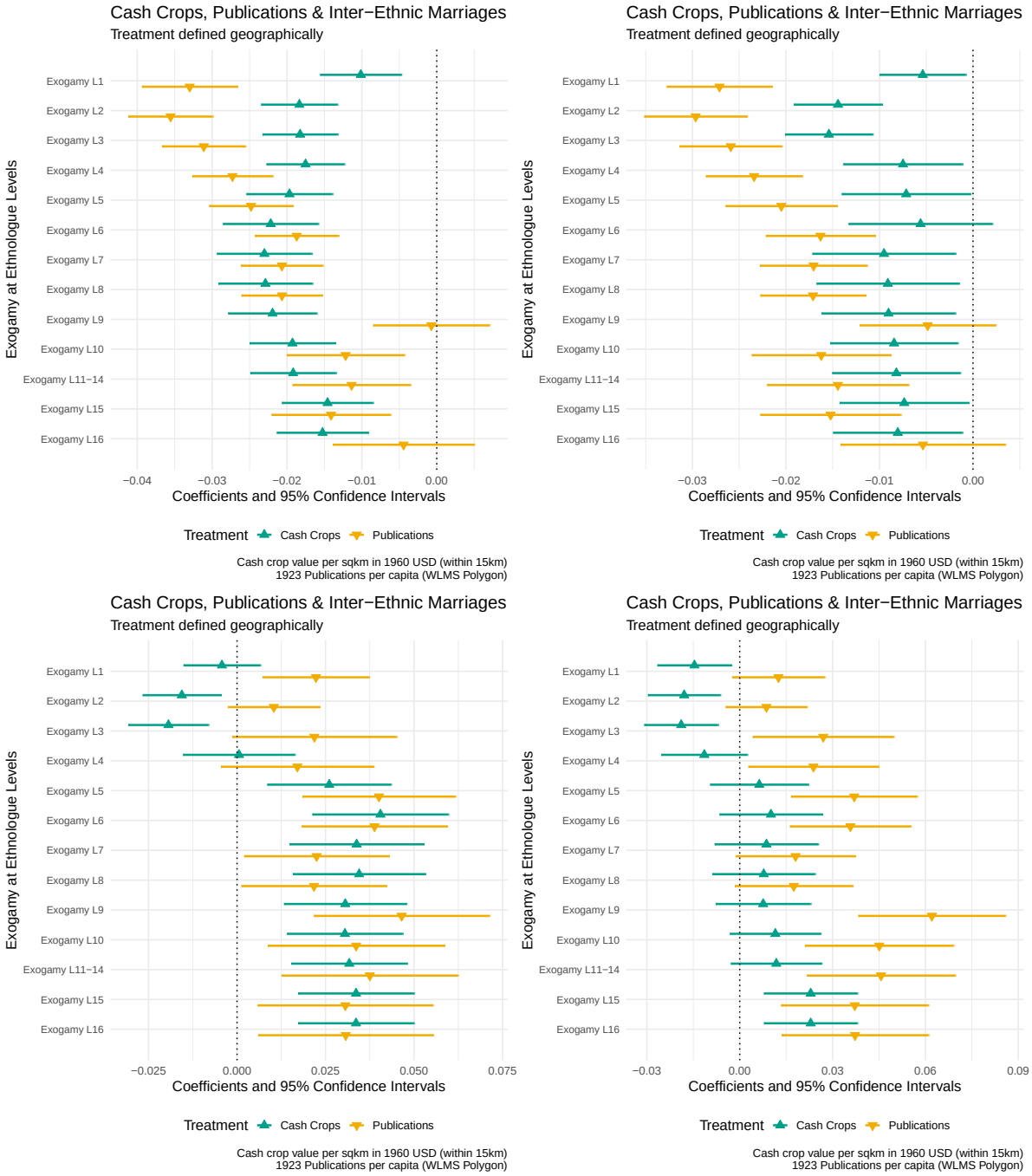
Notes: The figure reports findings from intensive margin versions of our cultural persistence models with location fixed effects. The sample now only includes couples in which the male spouse is an ethnic mover and self-reports to be from an ethnic group with at least one vernacular publication.

II.5 Additional DHS Results

Figure II.16 replicates our baseline analyses on four subsamples of male and female ethnic movers and stayers. Results suggest that the effects of cash crops are driven by historically 'native' ethnic groups rather than respondents who (or whose ancestors) migrated to a given location exposed to cash crop production and/or missionary publishing. These findings are consistent with local sons or daughters of the soil being the main agents of ethnic boundary making and enforcement in historical cash crop and publishing regions.

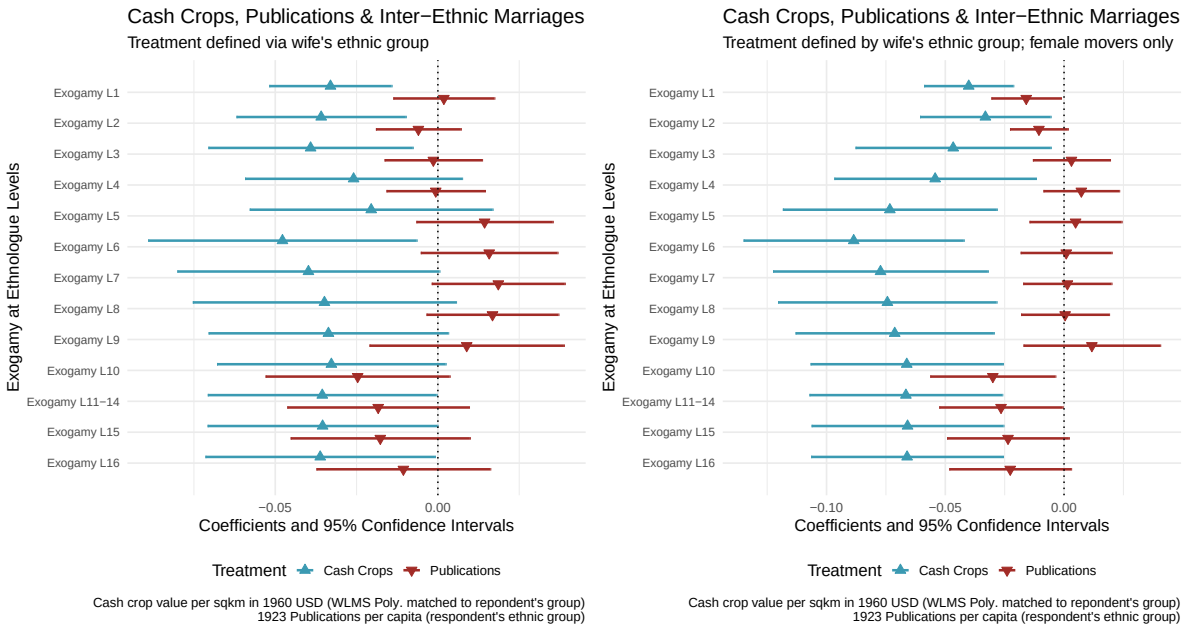
Figure II.17 reports models that assign both treatments based on the wife's instead of the husband's self-reported ethnic group. The cash crop coefficients are comparable to the male movers analysis but the publishing effects are smaller and tend to lose statistical significance. The ethnic persistence effect of vernacular publishing on ethnic boundaries thus seems to be mostly driven by male respondents' marital choices.

Figure II.16: Geographical Persistence - Movers vs. Stayers Subsamples



Notes: The figure reports standardized beta coefficients from 16 OLS models with country-survey-round fixed effects. The dependent variables are binary indicators of inter-ethnic marriages at all levels of the Ethnologue language tree. The analysis sample is restricted to male stayers (top-left), female stayers (rop-right), male movers (bottom-left) and female movers (bottom-right), respectively. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the standardized main variables of interest ((i)cash crop production per sqkm within a radius of 15 km of each survey location is situated in and (ii) publications per capita in 1923 in the language of the WLMS polygon the survey location is situated in.)

Figure II.17: Cultural Persistence - Treatment assigned via wifes' ethnicity



Notes: The figures reports OLS estimates from 16 models with survey location fixed effects. The dependent variables are binary indicators of inter-ethnic marriages at all levels of the Ethnologue language tree. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the main variables of interest: (i) the standardized USD value in cash crop production per sqkm within the WLMS polygon(s) matched to the wife's self-reported ethnic group (ii) standardized publications per capita in 1923 in the African language matched to the wife's self-reported ethnic group. Bars represent 95% confidence intervals. The left panel is based on analyses of the whole sample while the right panel reports results from models run on the subsample of ethnic movers only (i.e. wifes who reside outside of the ethnic polygon of their self-reported ethnic group.)

III Alternative Explanations & Plausibility of Mechanisms

III.1 Group Size

III.1.1 Group Size as Confounder

One concern about our findings is that African language publications and/or historical cash crop production pick up pre-existing group size or cohesion rather than exerting any path-dependent effects of their own. As larger ethnic groups are generally more likely to form viable minimum winning coalitions (Posner, 2004, 2005, 2017), group size may confound our estimates. We address this point in various ways.

First, we normalize the publications treatment (in all specifications across the paper) by the number of language speakers reported by missionaries in Rowling and Wilson (1923) which is arguably closely related to contemporaneous missionaries' perceptions of group size.

Second, we account for the pre-colonial population in each language polygon using estimates from the History Database of the Global Environment (HYDE, Klein Goldewijk, Beusen and Janssen (2010)) across all three analysis parts in the main paper. HYDE provides decadal population rasters since 1700. Our models include the (logged) average population per ethnic polygon between 1720 and 1890. As historical population estimates for Africa are notoriously unreliable (Frankema and Jerven, 2014), this strategy is likely conservative. The HYDE rasters, in most areas, appear as back projections of more robust contemporary population statistics. Cell-level correlations across decades and even centuries are implausibly high ($\geq .9$). While nominally pre-treatment, the HYDE data thus risks picking up post-treatment population dynamics. In addition, aggregating population rasters by ethnic polygons fails to account for historical population diversity at the local level. Despite these clear shortcomings, HYDE seems the best and perhaps only available source on precolonial populations at the subnational level. Our results remain robust to controlling for logged population per polygon. Coefficient sizes remain practically unchanged in the geographic and ethnic AB and DHS analyses, which now always include a population control (Tables 1 and 2, Figures 3 and 4, all AB and DHS specifications in the appendix). They clearly get smaller (cut by approx. 50%) in the group-level relevance models (compare first and last row in Figure 2). This suggests that accounting for size is indeed important yet does not explain away our results. As the political relevance datasets explicitly focus on national-level political competition, Posner's coalition logic may be more relevant for our group-level outcome than for individual-level salience and marital choices which may also respond to more local dynamics.

Third, we control for pre-colonial political centralization at the group level to account for the possibility that already powerful groups were targeted with missionary

publications or had an advantage in cash crop production (Table III.9, Figure III.18)

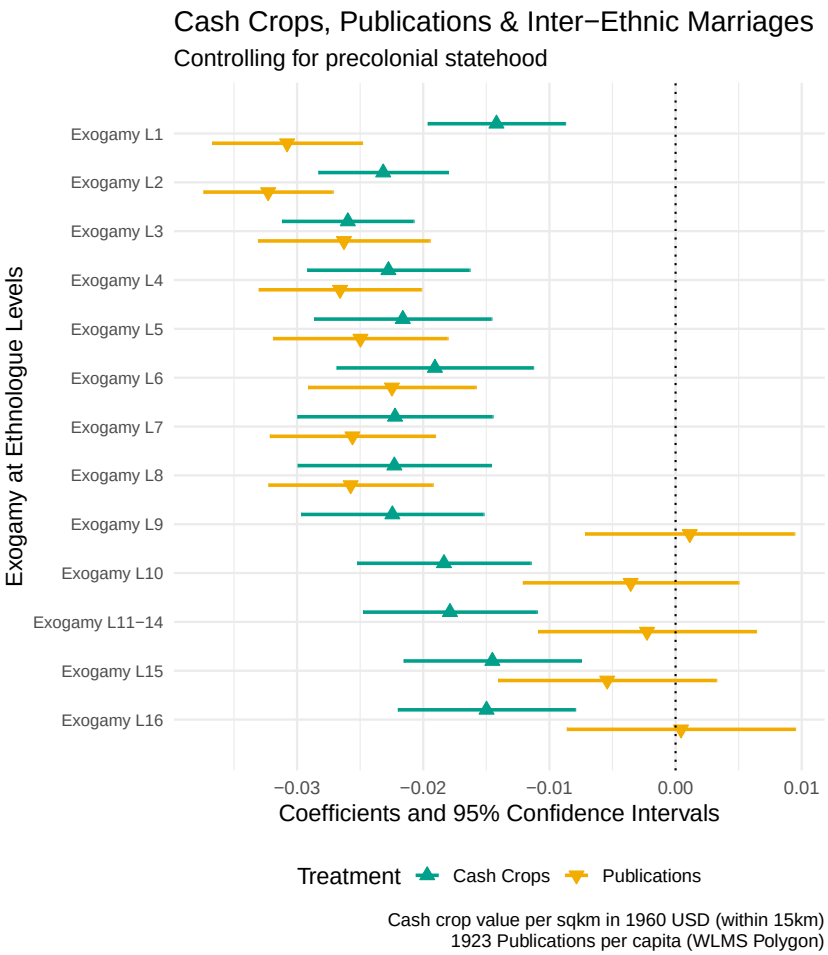
Fourth, we add logged ethnic polygon size to our ethnic leaver models to account for potentially disprortionate measurement error in ethnic leaver status and our geographically matched cash crop treatment among respondents from comparatively small groups. Figure III.19 shows that in our Afrobarometer sample, below-median sized groups have indeed higher shares of ethnic leavers than larger groups. Figures III.20 and III.21 show that our findings in the ethnic leaver specifications remain unaffected when accounting for ethnically matched polygon size.

Table III.9: Afrobarometer-Controlling for Pre-Colonial State Centralization

	Geographic-level		Ethnic-level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.011**	0.011**		
	(0.005)	(0.005)		
Pubs pth pop (1923)	0.041***	0.042***		
	(0.007)	(0.007)		
Murdock Centralisation	−0.017***	−0.027***		
	(0.006)	(0.008)		
Cash crops USD pkm2			−0.019***	−0.009
			(0.008)	(0.012)
Pubs pth pop (1923)			0.012**	0.012*
			(0.005)	(0.007)
Murdock Centralisation			−0.001	−0.002
			(0.007)	(0.010)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No
Fixed Effect	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.131	0.13	NA	NA
Observations	105,639	83,842	104,830	89,058
R^2	0.039	0.043	0.197	0.206

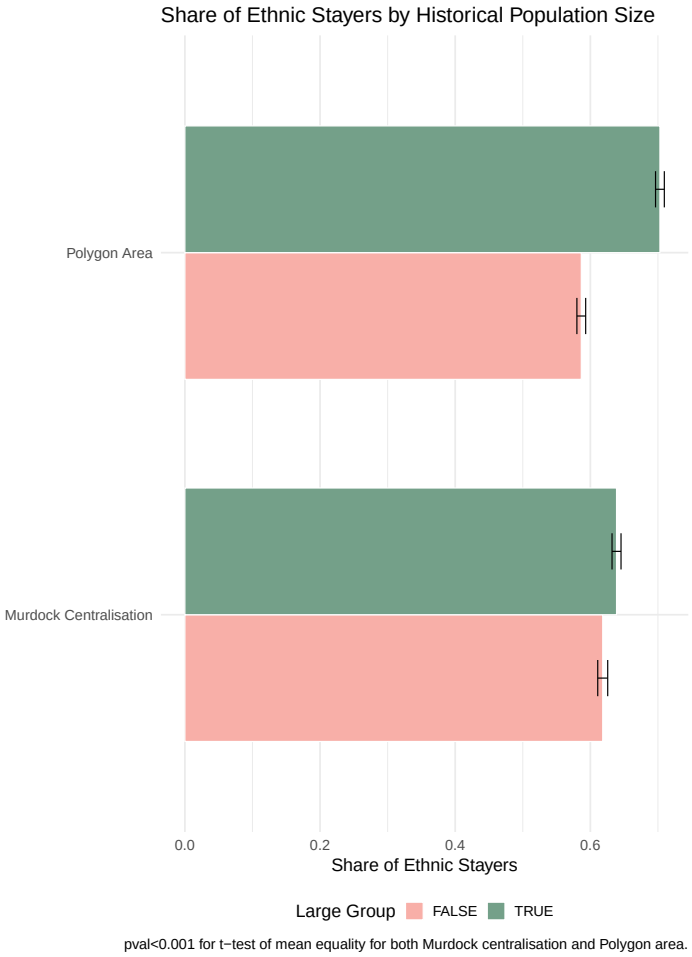
Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments, including Murdock centralisation, are defined at the location level (Columns (1) and (2)), and at the ethnic level (Columns (3) and (4)). Standard errors are reported in parentheses and clustered at the location level.

Figure III.18: DHS-Controlling for Pre-Colonial State Centralization



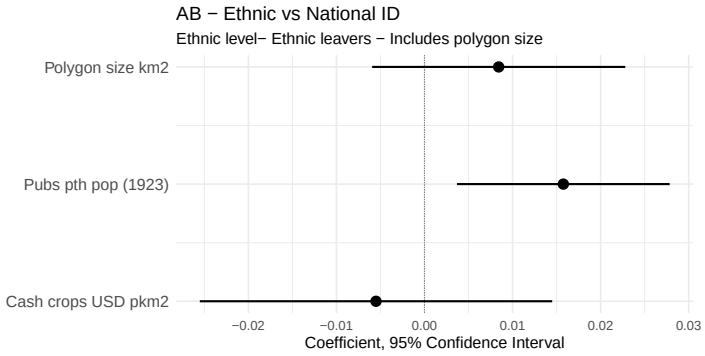
Notes: The figure replicates the geographic persistence analyses of inter-ethnic marriages from the main text. We add a precolonial statehood dummy based on Murdock (1967).

Figure III.19: Average share of ethnic leavers by group size



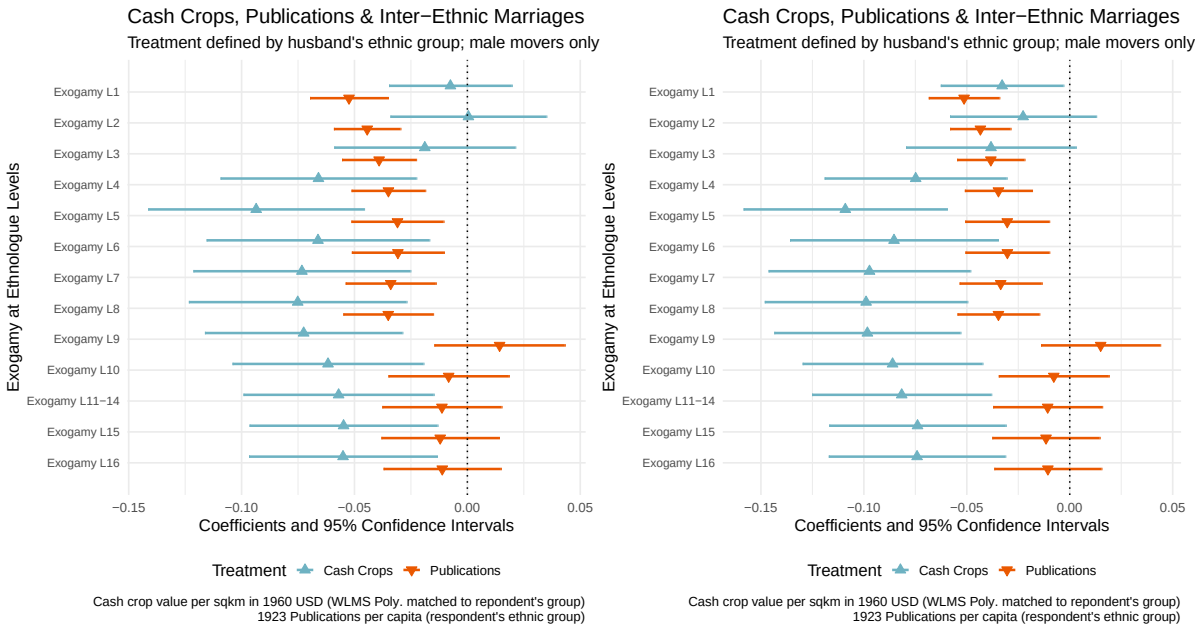
Notes: The figure shows average group size and ethnic polygon size among ethnic leavers and ethnic stayers in the Afrobarometer

Figure III.20: Ethnic leavers results, controlling for polygon size



Notes: The figure shows the effect of cash crop and publication treatment in the ethnic-level specification with Afrobarometer, when including polygon area in square km as a covariate. This corresponds to column (5) in Table 2 in the main paper.

Figure III.21: Group Size Controls in Leaver Models



Notes: The left-hand panel replicates the male leavers only models from the main paper. The right-hand panel adds the logged polygon area of husbands' self-reported ethnic group as as control. Doing so, if anything, increases the size of the crop coefficients.

III.1.2 Publications per Capita and Group Size

Beyond the issue of group size being a confounding factor of publications, we may worry that publications *per capita* are also positively related to group size.

In the presence of economies of scale, one could expect missionaries to choose larger groups first. Knowing the limited number of actors capable of entering the publishing market, that may have left smaller groups without any publications.[2] The subsequent empirical worry is that the number of publications *per capita* would be determined by initial group size, a known determinant of ethnic salience and ethnic politicization (Bates, 1983; Posner, 2004, 2005). This first potential issue is accounted for in the paper by focusing on the intensive margin of the publications effect, in other words, the effect of publications for groups with a record of publications. These are reported in column (5) in tables 2 and 3 and Appendix Figures II.14 and II.15.

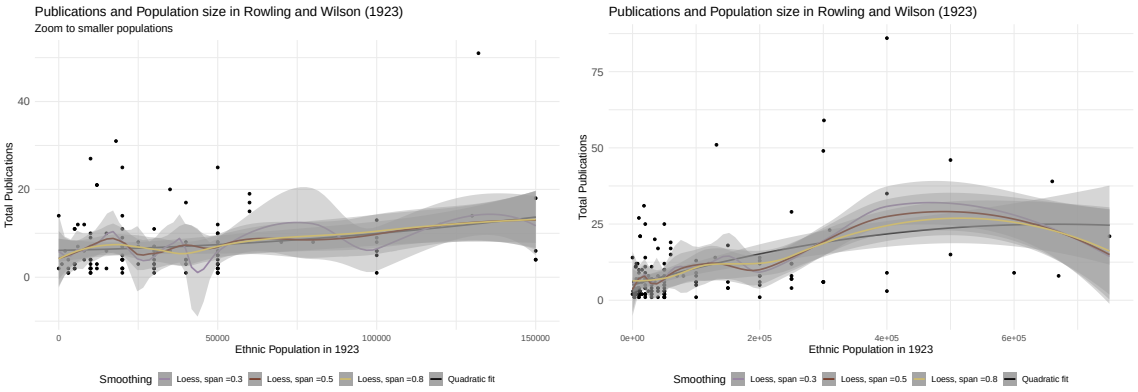
A final worry may arise if the number of publications would grow faster than ethnic group size, for instance as a possible result of economies of scale. In this scenario, publications per capita would still capture population size and bias our estimations. It is beyond the scope of this paper to model how publications are expected to evolve with market size.[3] However, we can assess the plausibility of this threat to our inferences by investigating the relationship between the number of publications and ethnic group size in more detail.

Figure III.22 (respectively III.23) plots the number of publications per ethnic group (respectively publications per capita per ethnic group) against group size as estimated in 1923 by Rowling and Wilson (1923). The figures do not support the assumption that total publications grow faster than ethnic group size. The number of publications per capita is, if anything, negatively related to ethnic group size, and this association is not robust across the sample (see Figure III.23). Moreover, notice that figure III.22 only provides evidence of a weak positive relationship between publications and group size, providing further reassurance on the potential issue of group size as a confounder, discussed in section III.1.1. Overall, we take this as evidence that our measures of publications *per capita* are unlikely to be fully capturing initial ethnic group size.

[2] See Cagé and Rueda (2016) for a discussion on the numerous constraints missionaries faced to start printing and publishing operations. This discussion is key to understand the assumption that a limited number of actors (missions or firms) could actually start publishing in African languages.

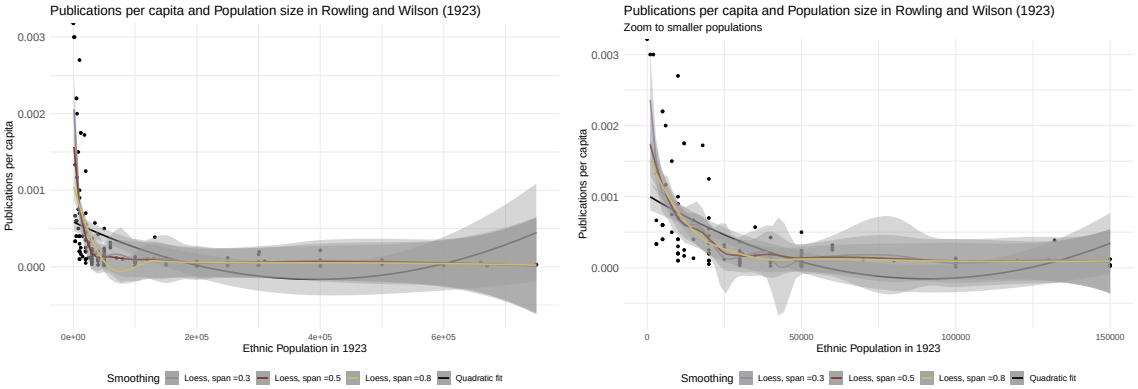
[3] The reader can refer to Desmet and Parente (2010) to overview the complex and diverse theoretical predictions in the relationship between market size and product variety, depending on modelling choices of preferences and competition.

Figure III.22: Publications and Ethnic Group Size



Notes: The figures plot the relationship between ethnic population size and total publications as stated in Rowling and Wilson (1923). Different fitting functions are applied to the data to estimate the relationship between the two variables

Figure III.23: Publications per capita and Ethnic Group Size



Notes: The figures plot the relationship between ethnic population size and total publications per capita as stated in Rowling and Wilson (1923). Different fitting functions are applied to the data to estimate the relationship between the two variables

III.2 Mediation Models

This section investigates potential mediating variables through which our historical treatments may affect contemporary ethnicity outcomes. We follow Acharya, Blackwell and Sen (2016) and estimate average controlled direct effects (ACDE), i.e. the remaining effect after accounting for specific mediators.

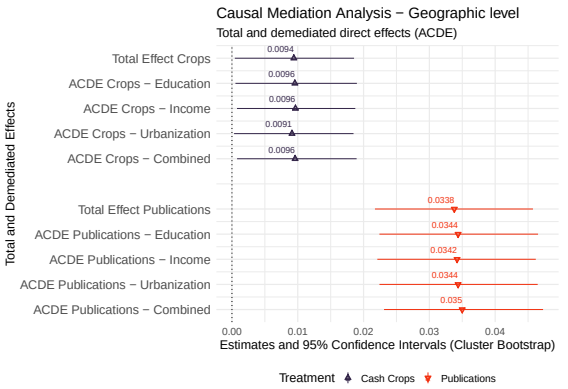
For the effect on ethnic salience as measured in Afrobarometer, we investigate three important mechanisms: (i) contemporary economic modernization (measured with individual education, income, and urban location) which may have resulted from cash crop agriculture and/or missionary investments, (ii) the consolidation of an early educated elite at the ethnic group level, which qualitative accounts see as an important factor in African ethnic group formation and mobilization (measured as the town or ethnic group share of individuals born before 1960 who have at least completed primary

school), and (iii) the development of political engagement and a public sphere (measured with newspaper readership, and different measures of political engagement) (Cagé and Rueda, 2016).[4] The results are reported in Figure III.24. For the effects on inter-ethnic marriage as measured in the DHS, we only account for (i) and (ii) since the DHS does not measure civic and political attitudes. The results are reported in Figure III.25 and III.26.

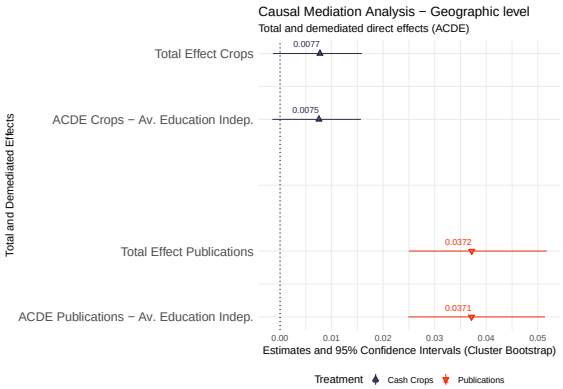
The results suggest that modernization does not explain much of our effects, and in some cases, its effect goes in the opposite direction (lesser ethnic salience and more porous boundaries, see figures III.24a, III.24d, and III.25). Early group-level or location advantages in education explain a small share of the publications treatment for ethnic salience (up to 3%, see figures III.24b and III.24e), but a much larger effect on inter-ethnic marriages (15-43%, see Figure III.26). Finally, political engagement and public sphere variables account for up to 17% of the publications effect on ethnic salience, when all measures are considered simultaneously (see last line of figures III.24c and III.24f).

[4] In terms of variable construction, "Education" is a binary variable equal to one for individuals with at least some high school education, "Newspaper" is defined as a binary variable equal to one for individuals who read newspapers at least once a week, "Pol. Discuss" is a binary variable equal to one for individuals who report discussing public affairs with friends or family at least occasionally . "Pol. Interest" is a binary variable equal to one if people declare at least some interest in politics. "Community Meet" is a binary variable equal to one if people declare attending community meetings at least a few times a year. "Income" is a binary variable equal to one if individuals report below median levels of access to cash income.

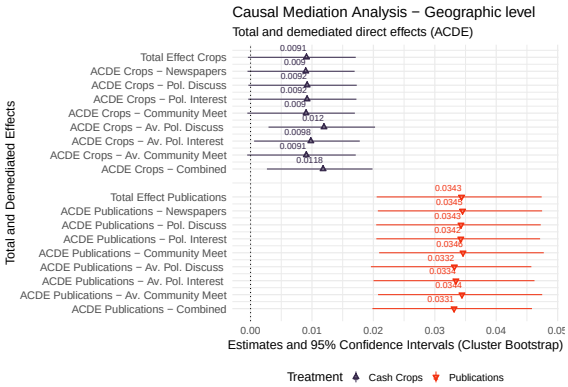
Figure III.24: AB Causal Mechanisms



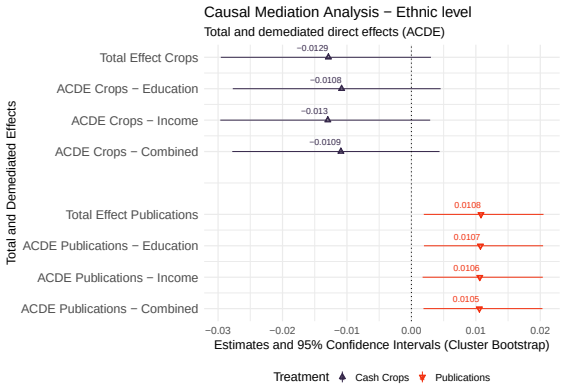
(a) Geographic Level - Modernization



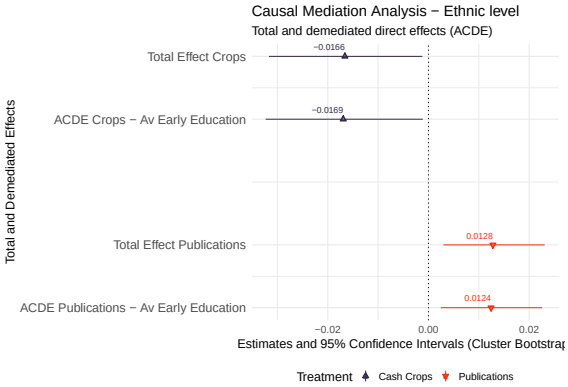
(b) Geographic Level - Early Elite



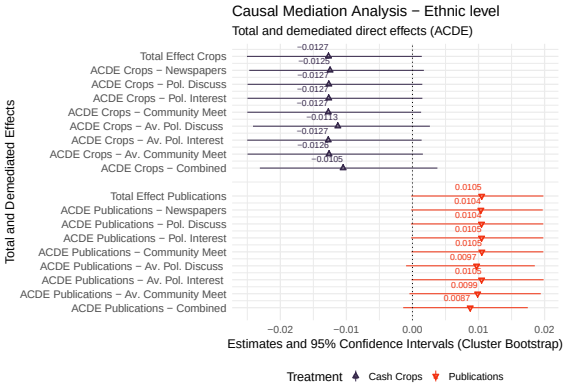
(c) Geographic Level - Public Sphere



(d) Ethnic Level - Modernization



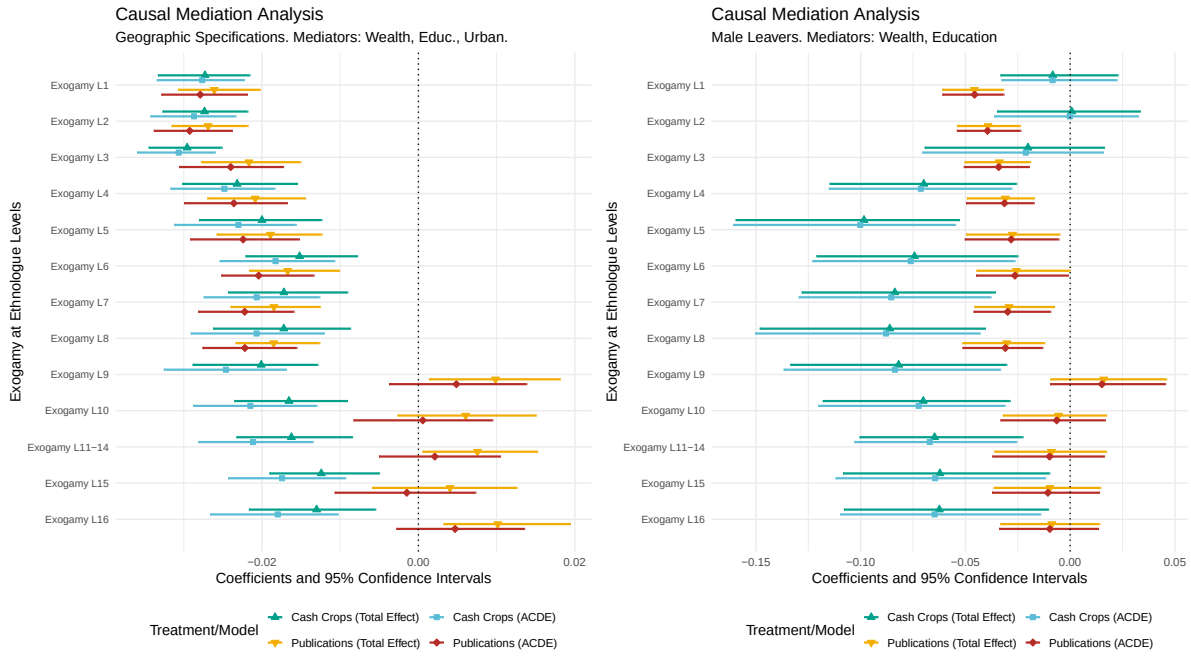
(e) Ethnic Level - Early Elite



(f) Ethnic level - Public Sphere

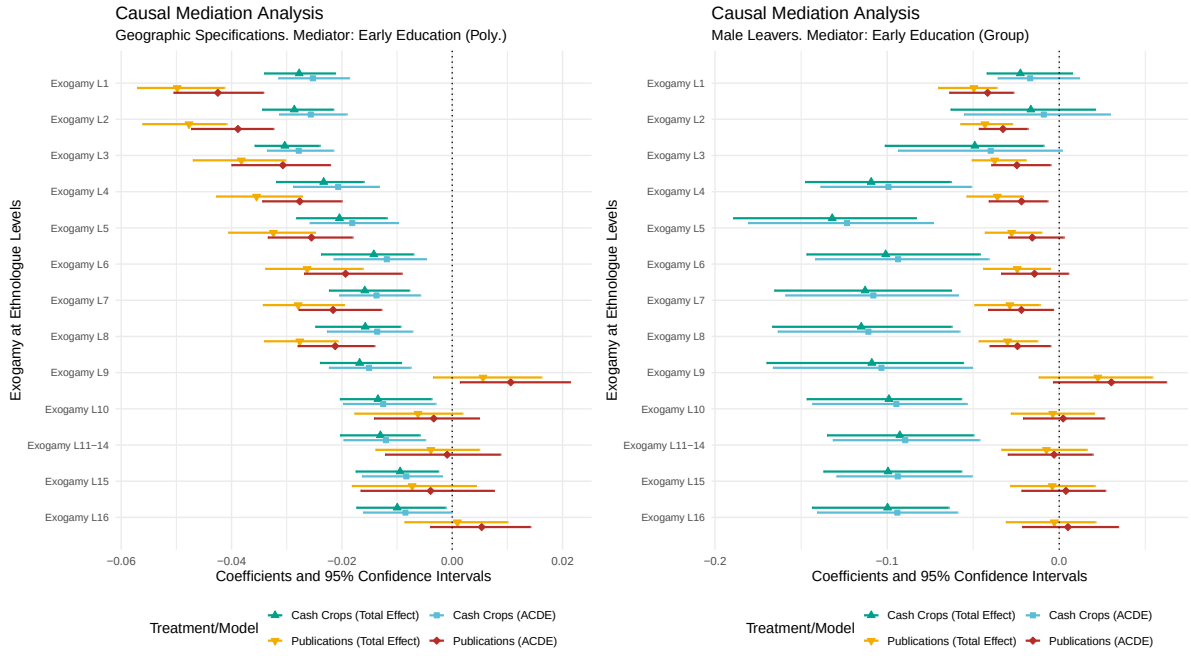
Notes: The figures show the ACDE estimated via sequential g estimation (Acharya, Blackwell and Sen, 2016). Standard errors are estimated through non-parametric bootstrapping with 150 iterations, clustered at the location level. The outcome is a binary variable equal to one if respondent declares a stronger ethnic than national identity. Each line reports either the total effect (un-mediated) of the treatments of interest, or the ACDE when the stated mediating factor is taken into account. The prefix "Av." flags mediators averaged at the town- and ethnic- levels, for geographic- and ethnic- level specifications respectively. "Combined" refers to the ACDE when all mediators stated above in the figure are considered simultaneously. The construction of the mediating variables is described in section III.2.

Figure III.25: DHS Causal Mechanisms: Modernization



Notes: The figures show the ACDEs and total effects. Standard errors are estimated through non-parametric bootstrapping with 150 iterations, clustered at the location level. The ACDEs account for modernization factors, simultaneously including education, wealth, and urban location in the model.

Figure III.26: DHS Causal Mechanisms : Early Elite



Notes: The figures show the ACDEs and total effects. Standard errors are estimated through non-parametric bootstrapping with 150 iterations, clustered at the location level. The ACDEs account for average education at the time of independence, proxied by the ethnic polygon's (left-hand panel) or group's (right-hand panel) share of individuals born before 1960 with at least some high school education.

III.3 Treatment Interactions

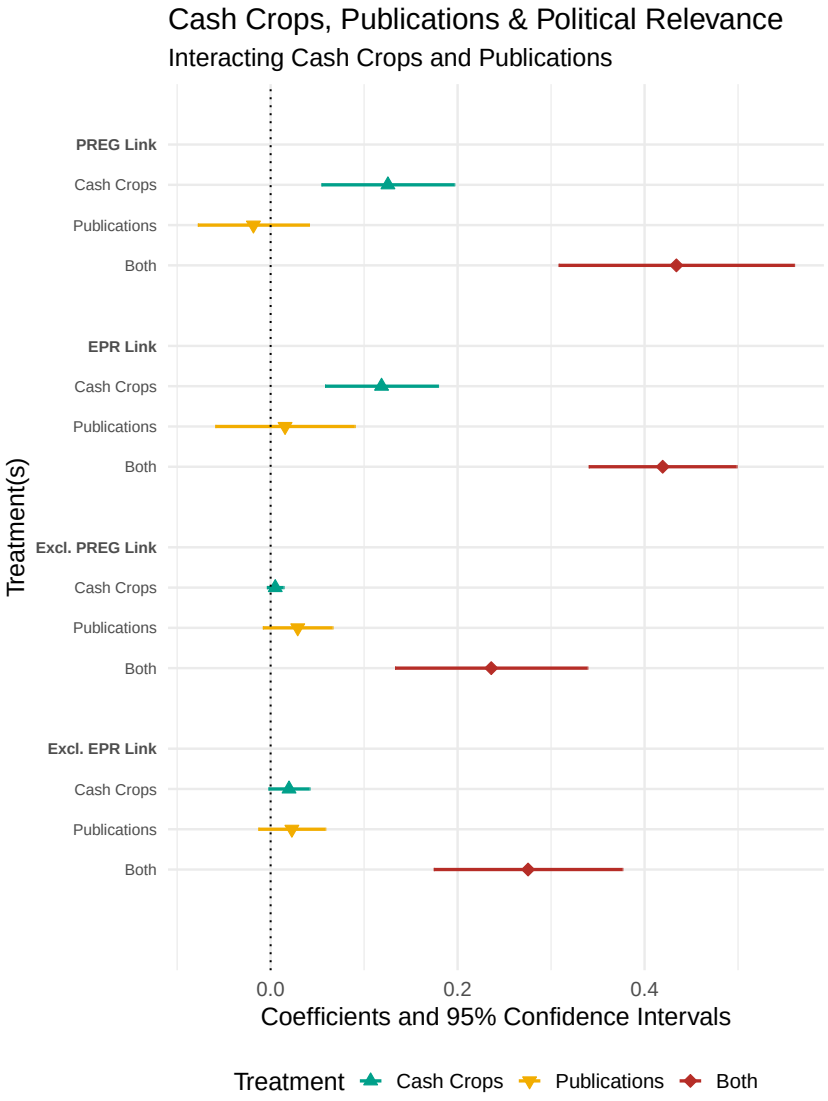
This section investigates potential interaction effects between our treatments of interest, cash crops and publications, across specifications. As our theory predicts similar effects of cash crops and publishing on group-level relevance and individual identity salience, we expect either additive or mutually reinforcing effects from the interaction models. As our theoretical predictions on ethnic boundaries diverge between cash crops and African language publications, we expect this to be reflected in interaction effects that dampen our baseline effects. The results below are broadly in line with these theoretical expectations.

Figure III.27 presents linear predictions from interactive specifications of our group-level relevance models. These models now include binary indicators for exposure to cash crops and publications as well as an interaction term between both treatment indicators. Results indicate that groups exposed to both historical transformations are more than 40 percentage points more likely to be coded as politically relevant in PREG or EPR. The cash crop constitutive terms remain positive, large and significant in the models that use the broader definition of political relevance (Any Link (Y/N)) but get small and lose significance in the "Exclusive Link" models. The pattern is almost reversed for the publication constitutive terms. It is indistinguishable from zero in the "Any Link" models but remains positive (borderline insignificant) when focusing on exclusive links, albeit with smaller substantive size.

The interacted geographic specifications for Afrobarometer similarly suggest that the effects of cash crops and publications magnify each other (first six coefficients in Figure III.28). These coefficients are based on two different models in which we keep one treatment variable in its original continuous form and split the other into three discrete categories (Zero, Low, and High). The distinction between Low and High is based on the sample median of the subset of respondents with at least some publications/crops. The results from the ethnic Afrobarometer specifications are different. Now, the publications effect seem to be mostly driven by survey locations with some rather than none or very intensive historical cash crop production.

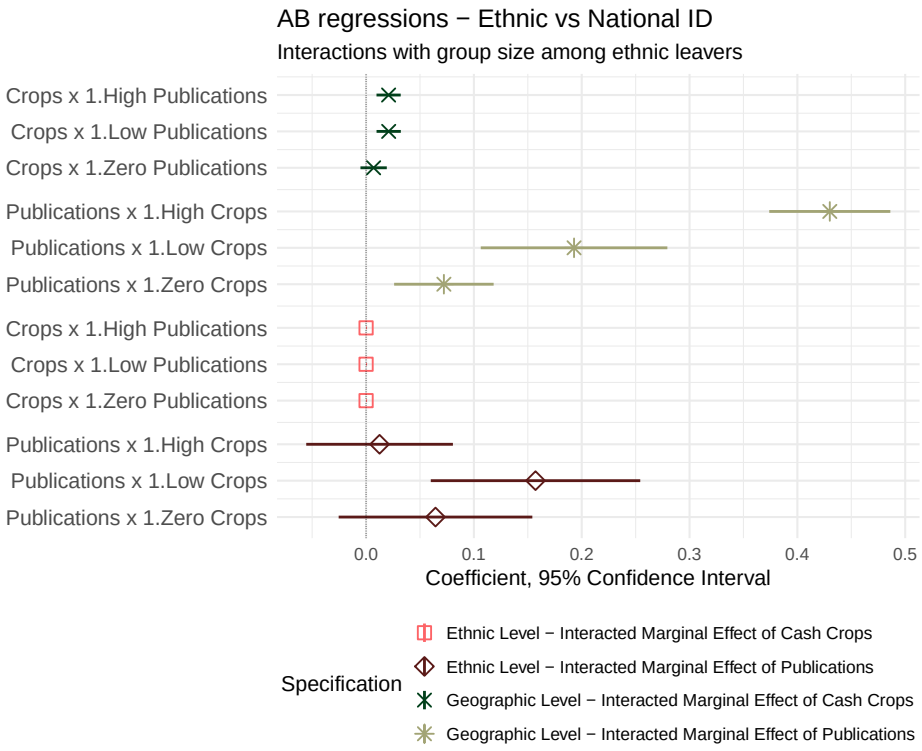
In the DHS marriage analyses Figure III.29), high levels of vernacular publishing tend to dampen the effect of cash crops (left-hand panel). In the absence of any historical cash crop production, missionary publishing is associated with less inter-marriage on levels 1-8 of the Ethnologue language tree and now significantly *more* exogamy on levels 9-16. Consistent with our theoretical expectations, the presence of cash crops counteracts this openness to linguistically related outgroups and leads to less inter-marriage across the board (right-hand panel).

Figure III.27: Cash Crops, Print Technologies, and Political Relevance - Interactions



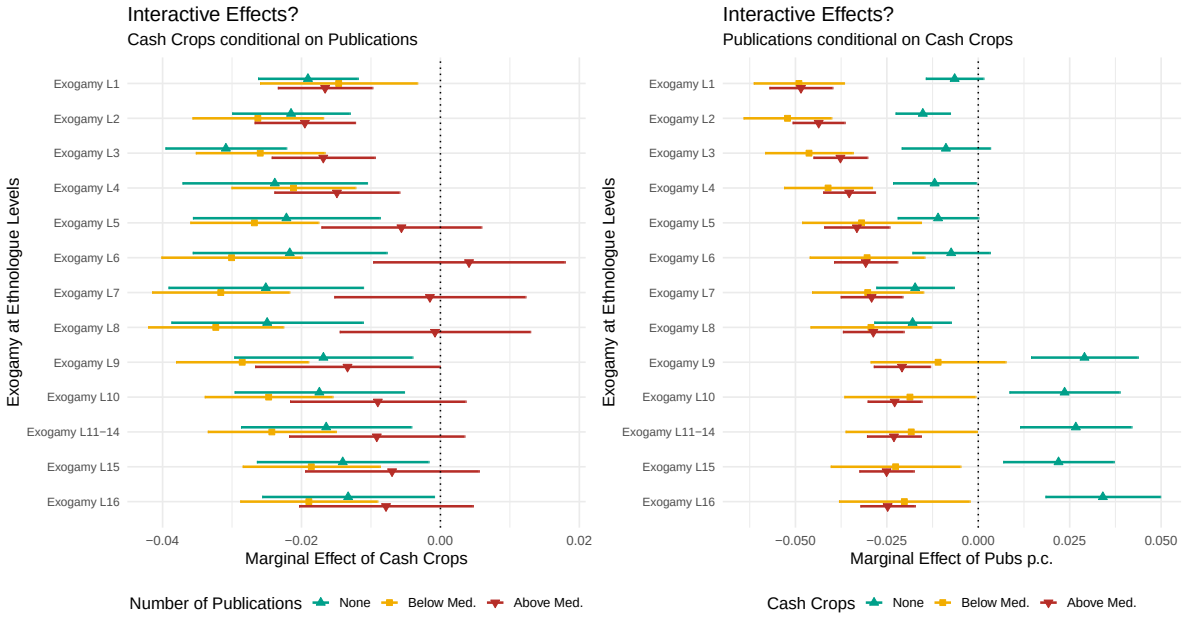
Notes: This figure shows the estimated effects when the two treatments of interest are interacted in the group-level specifications. The regressions are run for the four possible definitions of group-level politicization (exclusive or non-exclusive link in either EPR or PREG). The treatments are binary variables for high-levels of cash crops (resp. publications). "Both" refers to the interaction effect between these two binary treatments.

Figure III.28: Cash Crops, Print Technologies, and Political Relevance - Interactions



Notes: This figure shows the treatment effects of four different regression specification. Each specification is labelled in the legend. Regressions are either run at the location- or at the ethnic-level. For each level, there is one specification that interacts cash crops (resp. publications) with zero, low, or high levels of publications (resp. cash crops). "High" and "low" levels of treatment are defined as above and below the median non-zero level of the respective treatment.

Figure III.29: Treatment Interactions



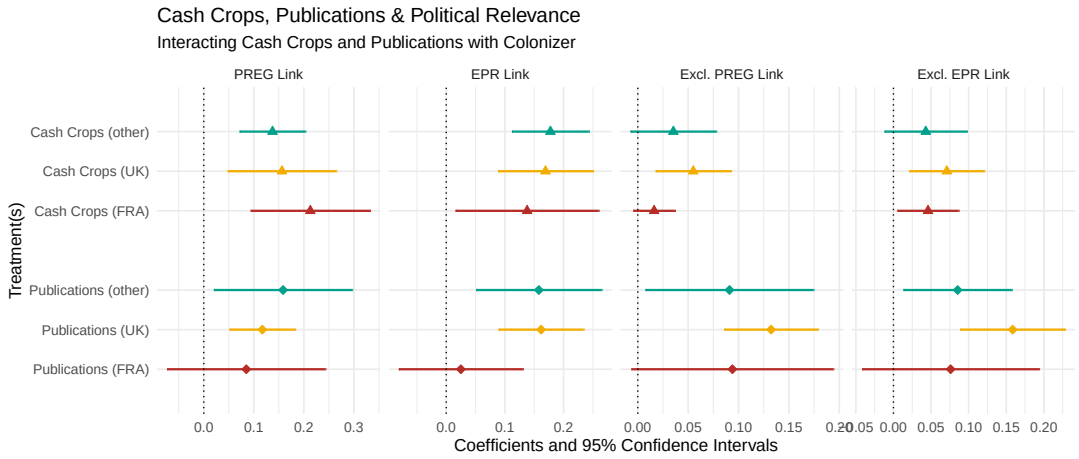
Notes: The figure replicated the geographic models of our DHS-based exogamy analyses but now interacts both historical treatment variables with each other. The left-hand panel shows results from models that interact the continuous cash crop treatment with zero, low, or high levels of publications. The right-hand panel interacts the continuous publication treatment with zero, low, or high levels of publications. "High" and "low" levels are defined as above and below the median non-zero level of the respective treatment.

III.4 Heterogeneity by Colonizer

This section investigates heterogeneous treatment effects depending on the identity of the colonizer. We separate countries depending on whether they were colonized by the UK, France, or any other colonial power. The cash crop effects on group-level political relevance (Figure III.30) and individual identity salience (Figure III.31 remain similar across different imperial powers. In the exogamy models, cash crops remain negatively associated with inter-group marriage in British and French colonies, but enter with positive though mostly insignificant coefficients for other colonies (Figure III.32). More interventionist and coercive modes of cash crop extraction and the relative frequency of plantation agriculture in Portuguese and Belgian colonies may explain these divergent findings (more on different modes of production below).

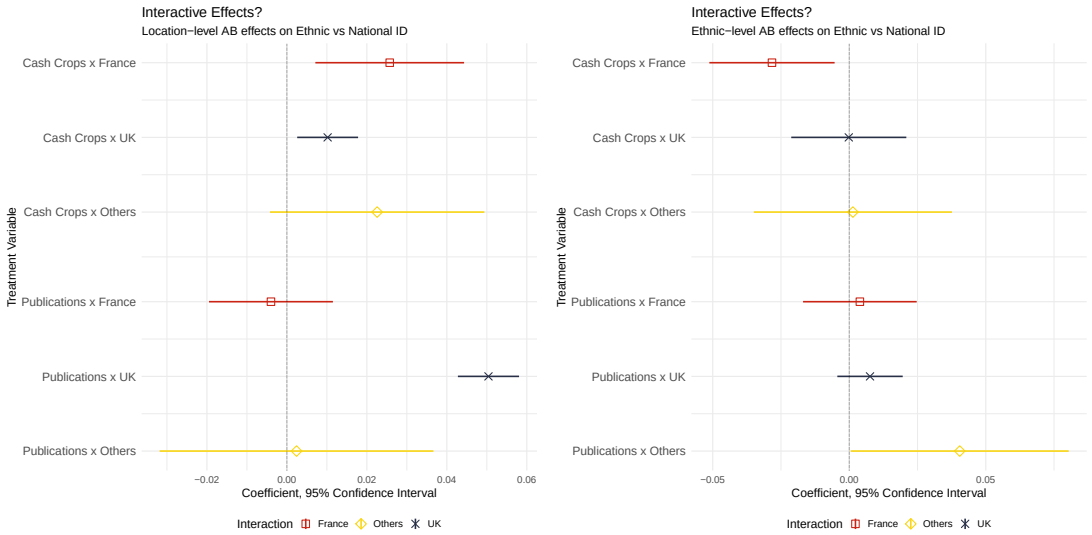
In French colonies, the publication effects on group-level relevance and inter-ethnic marriages are weaker than elsewhere (Figures III.30 and III.32), while the effect on Afrobarometer identity salience disappears completely (Figure III.31). One potential explanation is that French colonial governments put more emphasis on spreading their language than other imperial powers and gave missionaries less of a free hand in language standardization and vernacular education (Albaugh, 2014; Cogneau and Moradi, 2014).

Figure III.30: Group-level Politicization and Imperial Identity



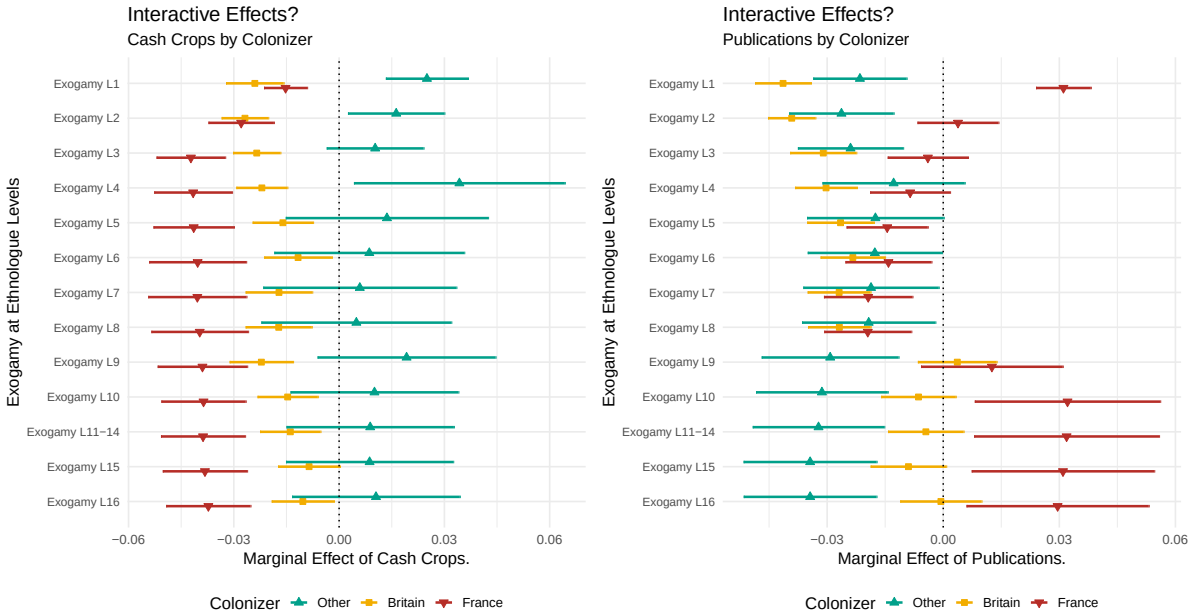
Notes: The figure replicates the group-level specifications with the PREG and EPR-based political relevance outcomes.Our historical treatments are now interacted with binary indicators for the respective country's colonizer (Britain, France, other).

Figure III.31: Ethnic Salience and Imperial Identity



Notes: The figure replicates the analyses of AB ethnic salience from the main text. Our historical treatments are now interacted with binary indicators for the respective country's colonizer (Britain, France, other). The figure reports marginal treatment effect by identity of the colonizer.

Figure III.32: Inter-ethnic Marriage and Imperial Identity



Notes: The figure replicates the geographic analyses of inter-ethnic marriages from the main text. Our historical treatments are now interacted with binary indicators for the respective country's colonizer (Britain, France, other).

III.5 Disaggregating Crops by Mode of Production

Top 5 vs. Other Cash Crops. The cash crop variables in our main analyses include the five most important African cash crops: cocoa, coffee, cotton, groundnuts, and oil palm. Together, these crops constituted approx. 50% off total exports in the African countries covered by our dataset and just short of 80% of all agricultural exports (Hance, Kotschar and Peterec, 1961). We restrict ourselves to these five crops for two main reasons. First, these crops are the only ones in our data with directly matching suitability rasters from the FAO which are needed for our spatial IV and reduced form models. While our spatial data source contains information on other agricultural exports, these are grouped together in umbrella categories which make it impossible to know the exact resource that was produced in a given point location.

Below is the list of individual cash crops and crop groups with their respective shares in total agricultural exports:

- Coffee: 22.45% of all agric. exports
- Cocoa: 17.45% of all agric. exports
- Cotton: 14.73% of all agric. exports
- Groundnuts: 14% of all agric. exports
- Oil Palm: 9.87% of all agric. exports
- Other Industrial Crops (Rubber, Pyrethrum, Sisal, Piassava, Kapok, Cinchona): 7.3% of all agric. exports
- Other Stimulants (Tobacco, Tea, Cloves, Chilies, Khat): 7% of all agric. exports
- Other Food Crops (Bananas, Sugar, Cereals, Pineapples, Peas and Beans): 5% of all agric. exports
- Other Oils (Coconut, Cashew, Castor, Beniseed, Sesame): 2.2% of all agric. exports

Second, we know that the top five cash crops were predominantly produced by African smallholders rather than on foreign-run plantations or European settler estates (with Kenyan coffee as the most prominent exception). If increased ethnic salience results, as we argue, from local ethnic communities' efforts to exclude in-migrants and other out-groups from the benefits of cash crop agriculture, we should only expect effects in production regions where smallholders dominate. Where European settlers, foreign plantations, or the state controlled the most productive land, local populations had little to gain and were in no position to regulate access to land or other resources.

Beyond the top five cash crops, the most important industrial crops (rubber and sisal), stimulants (tobacco, tea), and food crops (sugarcane, bananas) in our data were often but not always produced on plantations.

Smallholder vs. Plantation Crops. The match between crop and mode of production was, however, not perfect. Some typical smallholder crops were historically produced on plantations (e.g. coffee in Kenya and Angola, oil palm in the Belgian Congo), whereas African smallholders frequently outcompeted European settlers in producing typical plantation crops such as tobacco (e.g. in Malawi and Tanzania). To address this potential mismatch, we more carefully disaggregate our cash crop data by production type. Doing so at the level of 5924 unique geocoded production locations of agricultural exports is impossible, not least due to lacking information on the original data sources that went into the Hance map. As an alternative, we code the dominant production mode for all relevant country-crop combinations in our data. More specifically, we consult a wide range of review articles (see especially Gibbon, 2011; Smalley, 2013), historical case studies, and agricultural reports to learn whether at least 50% of the total export value of a given crop in a given country was produced on plantations or settler estates in or around the year 1957. If a crop crossed that threshold, we code all production locations in the respective country as plantation-based.

We identify the following country-crop combinations as dominated by plantations and/or European settlers:

- Coffee
  - Belgian Congo. Source: US Department of Agriculture (1950).
  - Angola. Source: Van Dongen (1961).
  - Kenya. Source: Brown (1968).
- Oil Palm
  - Belgian Congo. Source: Fieldhouse (1978).
  - Angola. Source: US Department of Agriculture (1949).
  - Cameroon. Sources: Ndoye and Kaimowitz (2000); Bederman (1966).
- Rubber
  - Everywhere, but in particular:
  - Belgian Congo. Source: US Department of Commerce (1956).
  - Liberia. Sources: Church (1969).
  - Cameroon. Sources: Ndoye and Kaimowitz (2000); Bederman (1966).

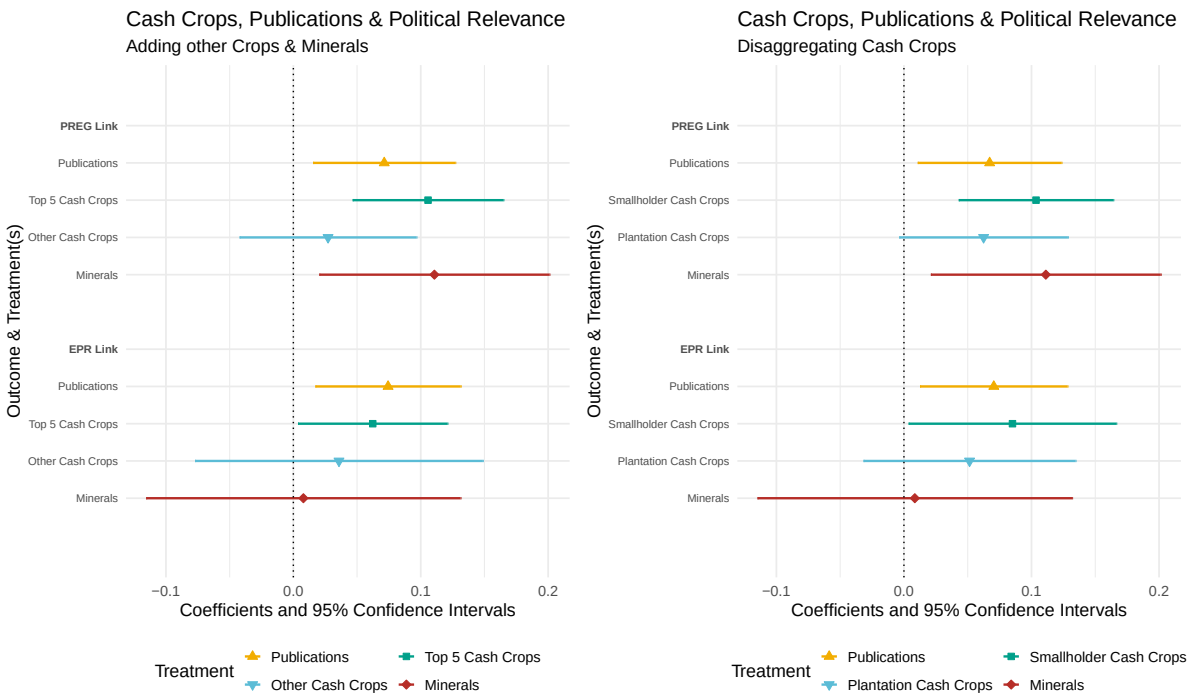
- Sisal
  - Everywhere, but in particular:
  - Tanganyika. Sources: Guillebaud (1959); Iliffe (1979).
- Tobacco
  - Rhodesias. Source: Frankema, Green and Hillbom (2016).
- Tea
  - Everywhere, but in particular:
  - Kenya. Source: Brown (1968).
  - Tanganyika. Source: Nagu (1986).
  - Uganda. Source: Hansen (1970).
  - Nyasaland: Baker (1962).
- Sugar
  - Everywhere, but in particular:
  - Mozambique. Source: Vail and White (1980); Gibbon (2011).
  - Angola. Source: Ball (2003).
- Bananas
  - Everywhere, but in particular:
  - Cameroon. Sources: Ndoye and Kaimowitz (2000); Bederman (1966).

As our extensive review has not revealed any information of quantitatively important smallholder production of rubber, sisal, sugar, or bananas for export, we code the umbrella groups "Other Industrial Crops" and "Other Food Crops" as plantation-based throughout.[5] The "Other Stimulants" group was difficult to code in Tanganyika and Nyasaland, as smallholders dominated tobacco but tea production was a largely European affair. As tea exports surpassed (TZA) or were almost on par with tobacco (MWI) in both cases, we code stimulants as plantation-based in both countries.

Note that we code all cotton points as produced by African smallholders. This was generally the case by the end of the colonial era, although colonial governments or concession companies more heavily intervened in cotton production than in the cases of cocoa, groundnuts, and coffee, especially in French and Portuguese colonies (see

[5]We know from the more detailed export share tables in Hance, Kotschar and Peterec (1961) that rubber and sisal together made up about 90% of the total value of "Other Industrial Crops", and that bananas and sugar were the most important exports in the "Food Crop" category.

Figure III.33: Disaggregating Crops: Political Relevance



Notes: Adding Minerals, other Resources, distinguish Smallholder from Plantation crops. Add sources.

contributions in Isaacman and Roberts, 1995). Forced cultivation, labor coercion, and state control over land, all of which occurred in some cases, should make our preferred mechanism less relevant and, if anything, make our decision for smallholder cotton a conservative choice. Cocoa and groundnut production, on the other hand, were clearly dominated by African smallholders in most if not all important source areas (Clarence-Smith, 1996; Hogendorn, 1978).

Results. Based on the discussion and coding above, we replicate all our baseline models with alternative resource variables. In a first step, we augment the group-level relevance models as well as the geographic Afrobarometer and DHS specifications with two additional variables that come directly from the Hance data. *Other Cash Crops* capture all agricultural export production beyond the top 5 cash crops discussed above and includes industrial crops, stimulants, food crops, and other oils. *Minerals* add the presence and/or value of local mineral production as coded in the Hance map. In total, minerals contributed 27.1% to the total of African exports in 1957 (crops: 64.9%). Copper, which was primarily mined in the Belgian Congo and Northern Rhodesia, made up almost half of the mineral total. In a second step, we replace the somewhat crude distinction between the *Top 5 Cash Crops* and *Other Cash Crops* by our more precise coding of *Smallholder Cash Crops* and *Plantation Cash Crops*.

Results are reported in Figure III.33, Table III.10, and Figure III.34. In the group-

Table III.10: Disaggregating Crops: Afrobarometer Ethnic Identities

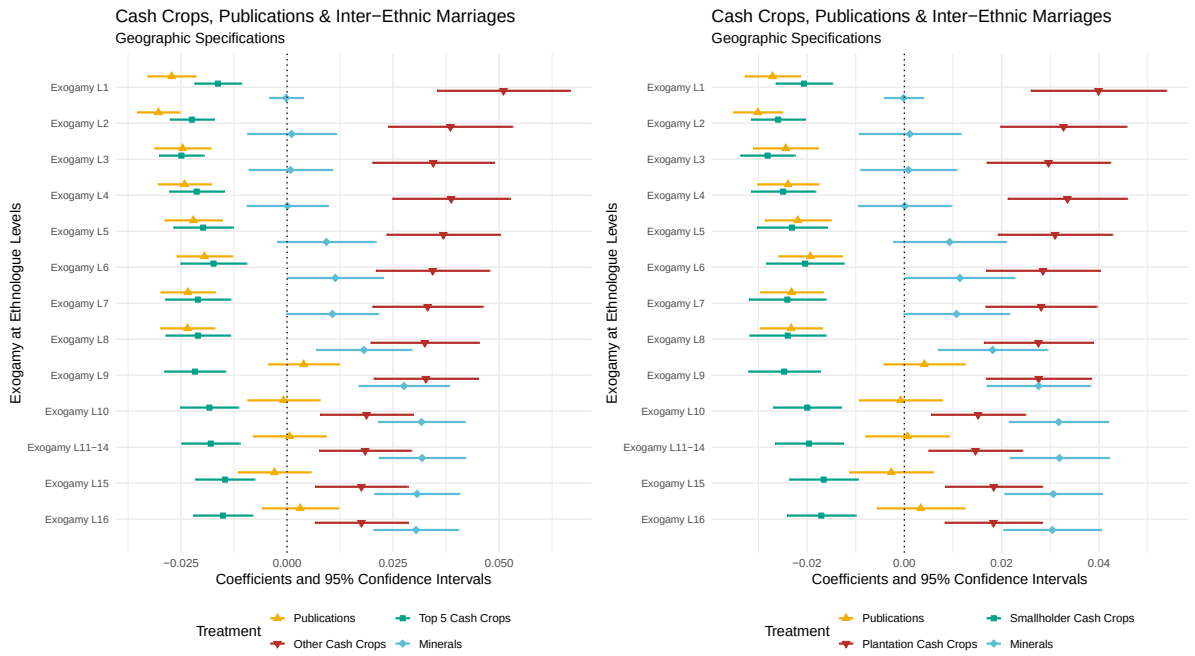
	Geographic- level - Ethnic vs National Id			
	(1)	(2)	(3)	(4)
In bibliographies 23	0.040***	0.040***	0.040***	0.040***
	(0.007)	(0.007)	(0.007)	(0.007)
Top 5 Cash Crops	0.011**		0.011**	
	(0.004)		(0.004)	
All Cash Crops		0.010**		
		(0.004)		
Other Cash Crops			0.002	
			(0.004)	
Smallholder Cash Crops				0.010**
				(0.005)
Plantation Cash Crops				0.004
				(0.004)
Minerals			−0.008***	−0.008***
			(0.002)	(0.002)
Individual controls	yes	yes	yes	yes
Historical and Geo controls	yes	yes	yes	yes
FE	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both
Mean dep. var.	0.1314	0.1314	0.1314	0.1314
Observations	115,962	115,962	115,962	115,962
R^2	0.045	0.045	0.045	0.045

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments, including all resource variables, are defined at the location level. Standard errors are reported in parentheses and clustered at the location level.

level models of political relevance and the Afrobarometer-based analysis of individual ethnic identity salience, the *Other Cash Crops* and *Plantation Cash Crops* variable point in the same direction as the *Top 5 Cash Crops* and *Smallholder Cash Crops*, yet coefficients are smaller and generally insignificant. In our DHS-based analyses of inter-ethnic marriages, however, *Other Cash Crops* and *Plantation Cash Crops* point in the opposite direction and are associated with more rather than less marriages across group boundaries.

We view this as consistent with our theoretical mechanism highlighting the role of local ethnic elites and communities in restricting access to economic benefits which should more directly affect ethnic boundaries than national or individual political salience. Where European companies, settlers, or the colonial state control production, this mechanism is unlikely to play out. On the contrary, land alienation, labor coercion, and other top-down interventions may have dismantled local ethnic institutions and thereby reduced local opportunities for ethnic boundary-making. Consistent with this view, local mining is also associated with more inter-group marriages (Figure III.34), less salient ethnic identities (Table III.10), and mixed effects on group-level political relevance

Figure III.34: Disaggregating Crops: Inter-Ethnic Marriages



Notes: Adding Minerals, other Resources, distinguish Smallholder from Plantation crops. Add sources.

(Null effects in EPR models, positive for PREG). Again, top-down control and coercion by extraction companies and/or the colonial state may have provided cross-cutting incentives for class-based or anti-colonial rather than ethnic mobilization. At the same time, colonial mining had long-term effects on local urbanization and industrialization levels which may have reduced the relevance of ethnic identities post independence.

Overall, these results point to a more nuanced role of colonial-era economic modernization in shaping ethnic cleavages than previously acknowledged. While classic modernization theories expect a gradual shift from ethnic to national identities (Robinson, 2014), "second-generation modernization theory" (Eifert, Miguel and Posner, 2010) expects more salient ethnic cleavages due to inter-group competition for economic benefits (Bates, 1974). Our results highlight how different modes of production and the corresponding ability of ethnic elites to regulate access may matter in shaping how economic change affects ethnic identities.

III.6 Local Ethnic Diversity

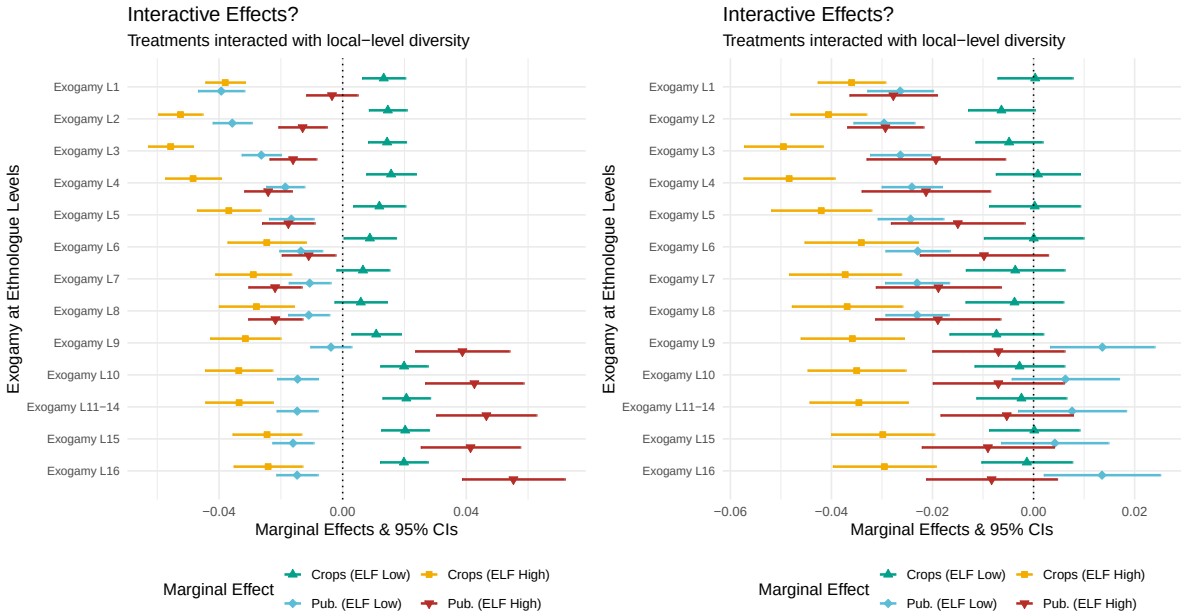
Within-country variation in local ethnic diversity may threaten the validity of our findings. First, lower levels of inter-ethnic marriage may be due to local ethnic homogeneity and lacking opportunities to find an out-group spouse rather than sharp boundaries in heterogeneous areas. Second, historical levels of ethnic diversity might be associated with both our treatments and outcomes and therefore bias our analyses.

We address the first concern by interacting the cash crop and publishing treatments with contemporary ethnolinguistic fractionalization scores calculated at the DHS enumeration area and at the same Ethnologue level of linguistic differentiation as the respective dependent variable. We use the sample median of the respective ELF score to distinguish survey locations with high from those with low levels of local ethnic diversity. The left-hand panel in Figure III.35 reports results indicating that, if anything, our findings are driven by relatively diverse survey locations. This further strengthens our confidence that ethnic competition between sons-of-the-soil and (historical) in-migrants rather than local-level ethnic homegeneity explains lower exogamy in historical cash crop areas.

As for the second concern, we lack explicit data on precolonial population diversity at the subnational level. As a remedy, we use the Ethnologue/WLMS polygons of historical linguistic settlement areas to construct an imperfect spatial proxy. More specifically, we spatially intersect the ethnic polygons with a spatial grid of 0.5 degree resolution (about 56 km at the equator). We then compute indices of ethno-territorial fractionalization in each grid cell based on the area shares of different ethnic polygons within each grid cell. All cells intersected by only one polygon receive a territorial fractionalization score of 0, while cells intersecting with more than one polygon have scores between 0 and 0.93. The main intuition is that areas at the border of different language groups and/or populated by a number of relatively small groups were historically more diverse. We assign each DHS survey location the territorial fractionalization score of its grid cell. The right-hand panel of Figure III.35 again interacts our treatments with a binary indicator of high fractionalization values (above median). Results remain similar to the previous models using contemporary fractionalization scores.

In addition to these interactive models, we also include the cell-based territorial fractionalization index as a control variable in our geographic models of the Afrobarometer and DHS analyses. Doing so slightly reduces coefficient size (more so for the cash crop variable) but does not fundamentally alter our conclusions (See Table III.11 and Figure III.36).

Figure III.35: Local Ethnic Diversity? Interactions



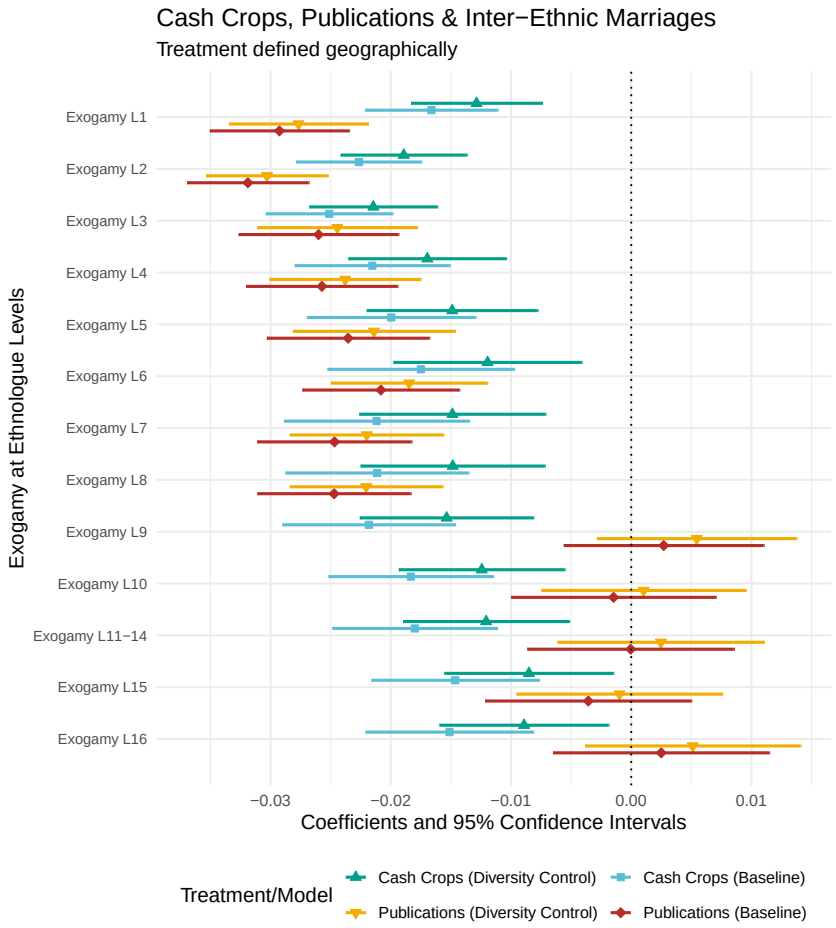
Notes: The figures replicate the geographic persistence analysis of inter-ethnic marriages from the main text. The left-hand panel interacts the cash crop and publishing treatments with contemporary ethnolinguistic fractionalization scores in DHS survey locations. The right-hand panel interacts treatments with a territorial ethnic fractionalization index based on historic linguistic settlement areas from WLMS/Ethnologue (see details in the text).

Table III.11: Afrobarometer (Geographic Models): Local Diversity

	Geographic- level - Ethnic vs National Id					
	(1)	(2)	(3)	(4)	(5)	(6)
Top 5 Cash Crops	0.013***	0.010**			0.011**	0.008*
	(0.005)	(0.005)			(0.004)	(0.005)
In bibliographies 23			0.040***	0.040***	0.040***	0.039***
			(0.007)	(0.007)	(0.007)	(0.007)
Map-based ELF		−0.079***		−0.084***		−0.081***
		(0.017)		(0.017)		(0.017)
Individual controls	yes	yes	yes	yes	yes	yes
Historical and Geo controls	yes	yes	yes	yes	yes	yes
FE	Country-Round	Country-Round	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both	Both	Both
Mean dep. var.	0.1314	0.1314	0.1314	0.1314	0.1314	0.1314
Observations	118,873	116,140	115,962	114,063	115,962	114,063
R^2	0.043	0.044	0.045	0.046	0.045	0.046

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments, are defined at the location level. Standard errors are reported in parentheses and clustered at the location level.

Figure III.36: Local Ethnic Diversity? Control



Notes: The figure replicates the geographic persistence analysis of inter-ethnic marriages from the main text. We now add a map-based territorial fractionalization score as control variable.

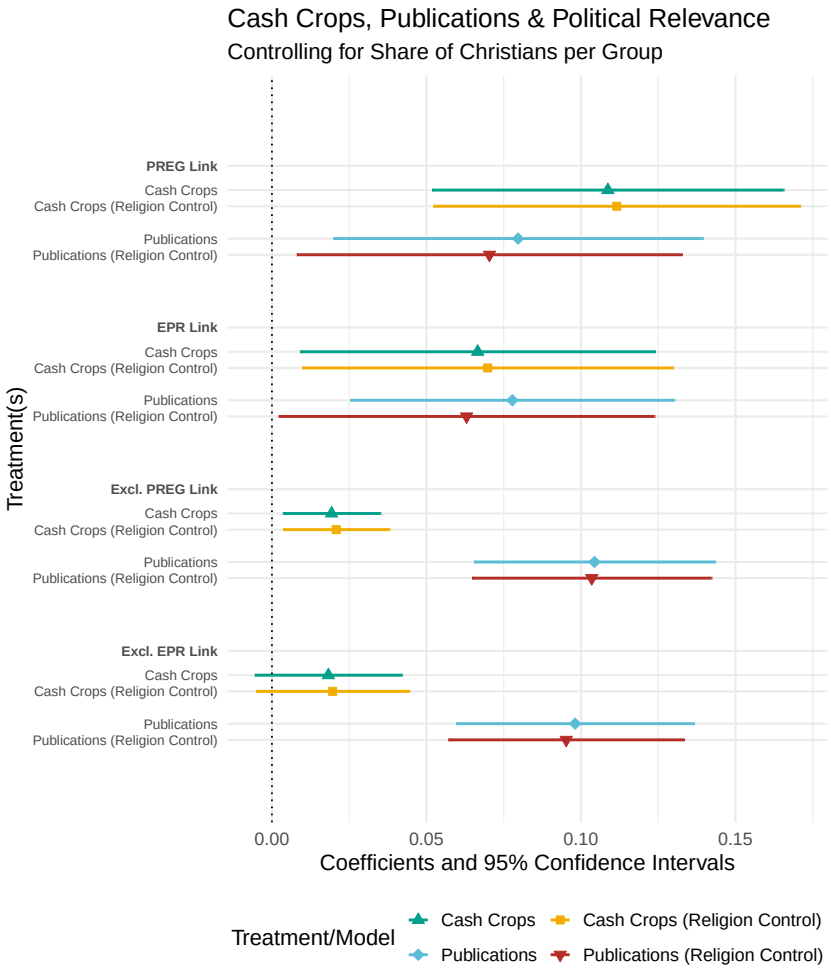
III.7 Ethnic or Religious Boundary Making?

What role, if any, does religious conversion play in explaining the impact of missionary publishing in African languages on inter-ethnic marriage? One concern is that rather than strengthening ethnic boundaries, missionary publishing and other activities instilled stronger religious identities and a preference for religious rather than ethnic endogamy. We address this concern in several ways. First, we add group-level estimates of the share of Christian adherents coded by the Joshua Project to our political relevance models. Figure III.37 indicates that doing so does not affect our conclusions.

Second, we add 857/696 directed religious couple fixed effects to our DHS geographic/ethnic leaver models, using the relatively fine-grained religious classifications from DHS. Directed fixed effects ensure that e.g. a Presbyterian women married to a Pentecostal men is not compared to a Pentecostal women married to a Presbyterian man. The main idea is to test whether we still find effects on inter-ethnic marriages when only comparing within groups of couples with exactly the same combination of religious denominations between both spouses. The results in Figure III.38 remain practically indistinguishable from our baseline models.

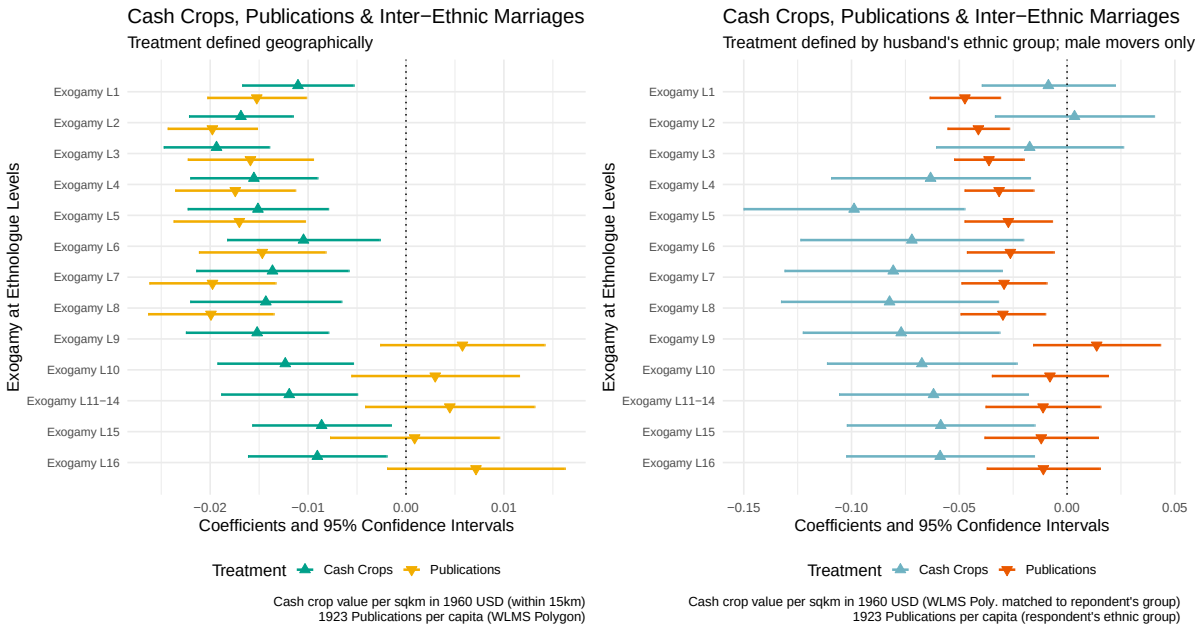
As an alternative strategy, we run mediation models as described above but this time using religious denomination dummies as mediator variables (eight dummies in total: protestant, catholic, muslim, and traditional for wife and husband, respectively). Figure III.39 shows that contemporary religious affiliation as reported by DHS respondents only explains a small share of the total effect of our publications treatment.

Figure III.37: Religious Conversion as Alternative Mechanism?



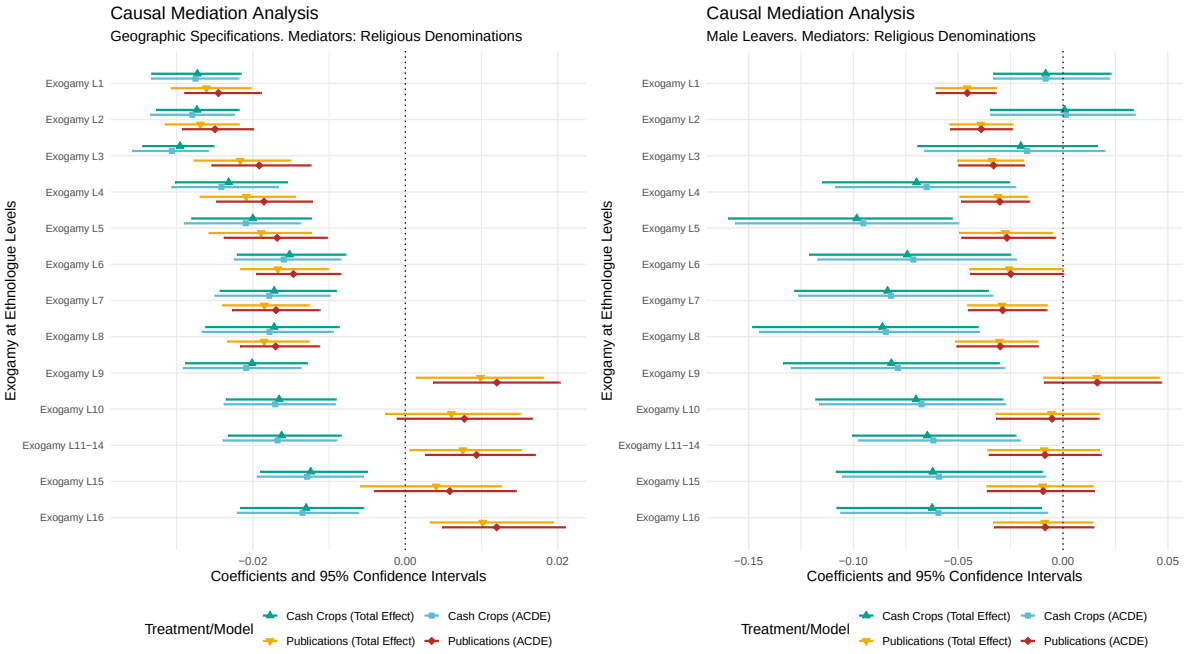
Notes: The figure replicates the group-level political relevance models and adds estimates of the share of Christians per Ethnologue group as control variable (the data comes from the Joshua Project). Adding this control to our most stringent group-level models (with historical size controls) only minimally affects the coefficients of interest.

Figure III.38: Ethnic or Religious Endogamy?



Notes: The figure replicates the geographic and cultural persistence analyses of inter-ethnic marriages from the main text. We add directed religious couple fixed effects to test whether the ethnic marriage effects may plausibly driven by sharper religious boundaries. Results in the geographic persistence (left), and ethnic movers only models (right) remain very similar to our baseline specifications.

Figure III.39: Causal Mechanisms: Religious Conversion



Notes: Same mediation modelling approach as above in section III.2. We include eight religious denomination dummies as mediators (Protestant, Catholic, Muslim, Traditional; each defined for male and female spouse).

IV Description of Pre-Registered Analyses and Supplementary Analyses

Tables IV.12 and IV.13 list all analyses we specified in our preregistered report (See http://bit.ly/3qYufI8). Importantly, the report specified the ancillary hypotheses (i) that both cash crops and print technologies would lead to more homogeneous political preferences among members of treated groups, (ii) that cash crops would be associated with lower levels of inter-ethnic trust than publishing, and (iii) that due to sharper boundaries, areas of historical cash crop production would see higher levels of ethnic conflict than those exposed to missionary printing. As the results below indicate (see Table IV.14, Figure IV.40, and Figure IV.41), we find no or only partial support for these additional hypotheses. Table IV.15 summarizes additional analyses that we added after pre-registration.

IV.1 Pre-registered Analyses

Table IV.12: Preregistered Analyses

Pre-Registered Analysis	Implementation
Ethnic politicization and salience, group-level	
Group-level politicization	
Group-level ethnic politicization based on a group listed as relevant in EPR and PREG under the same or a clearly synonymous name.	As described. See Figure 2.
Group-level salience	
Group-level ethnic salience based on a group listed as socially-relevant in AMAR.	As described. See Figure II.12.
Ethnic politicization, individual-level	
Ethnic vs national identity	
Ethnic salience using Afrobarometer survey responses to whether individuals identify more in ethnic or national terms	
-Geographic-matching: survey locations matched to Ethnologue polygons	As described. See Table 1.
-Ethnic-matching: based on language individuals speak at home	As described. See Table 2. As a robustness, we also tested matching on declared ethnicity (which has fewer observations) and results remain unchanged.
Bloc-voting	
Homogeneity of political preferences among members of the same ethnic group.	Implemented with vote intention. See Figure IV.40.
Mechanisms	
Mechanisms analysis: Ethnic salience among stayers vs leavers	
- Geographic-matching: sub-group analysis of leavers	As described. See Table 1, Column 5.
- Ethnic-matching: sub-group analysis of leavers	As described. See Table 2, Columns 5-6.

Table IV.13: Preregistered Analyses - Continuation

Continuation

Pre-Registered Analysis	Implementation
Ethnic boundary-making	
Inter-ethnic marriage	
Inter-ethnic marriage using data from the Demographic and Health Surveys at all 16 levels of Ethnologue's linguistic hierarchy	
- Geographic-matching: Assignment based on survey location	As described. See Figure 3
-Ethnic-matching: Assignment based on husband's ethnicity	As described. See Figure 4 (left-hand side)
-Ethnic-matching: Assignment based on wife's ethnicity	As described. See Figure II.17 .
Stayers vs leavers analysis of inter-ethnic marriage	
-Geographic-matching: sub-group analysis of stayers and leavers	As described. See Figure II.16.
-Ethnic-matching: sub-group analysis of leavers, assignment based on husband or wife's ethnicity.	As described. See Figure 4 (right-hand side) and II.17
Ethnocentric Trust	
Ethnocentric trust: based on two questions from Afrobarometer Round 3 about trust in co-ethnics and non-co-ethnics	Deviation: Given the high correlation between answers to trust coethnics and trust non-coethnics, ethnocentric trust is computed instead as a binary variable flagging those reporting high coethnic trust but low *general trust* (as oppposed to low non-coethnic trust).
-Geographic-matching	See Table IV.14, Columns (1)-(3).
-Ethnic-matching	See Table IV.14, Columns (4)-(6).
Stayers vs leavers analysis of ethnocentric trust	Not pursued given weakness in trust data described above and inconclusive results of baseline ethnocentric trust analysis reported in Table IV.14.
Ethnic Conflict	
Ethnic conflict: aggregate all clearly ethnicity-related events from ACLED to the Ethnologue polygons	Use ACLED identity militia measure. See Figure IV.41.
Addressing endogeneity	
Instrumental Variables	
Use cash crop suitability to run 2SLS-IV models to identify the causal effects of colonial cash crop extraction on ethnic identities.	As described. Lines 3-4 in Figure 2. Column (4) in Table 1. Column (5) in Table 2. Figure II.14.
Intensive Margin within exposed areas	
Intensive margin: subsetting analysis to groups with at least one publication and estimating the effect of magnitude in publication record	As described. See Tables 1 and 2, as well as Figures II.14 and II.15.
Lingusitic proximity	
Analysis of linguistically-proximate groups to test spillovers	Not implemented yet. Spillover analysis is an avenue for future research.

IV.1.1 Trust

In the pre-registered report, we conceived that another observable implication of the effects of cash crops and print technologies on boundary-making was via co-ethnic trust. In particular, we expected that the boundary policing induced by cash crops would lead to stronger intra-ethnic bonds, manifesting in higher levels of self-reported trust in co-ethnics than those from different ethnic groups. In contrast, we anticipated the more porous boundaries arising from publishing may expand bonds of trust. To test this, we follow Nunn and Wantchekon (2011) and look at three binary outcomes of interest that flag whether respondents: (i) declare high level of trust towards co-ethnics, (ii) declare high level of trust towards most people, and (iii) have a coethnic trust premium (i.e., report trusting coethnics, but not trusting people in general.) Results are reported below in Table IV.14. We observe that in both the geographic and the ethnic specifications, the cash crop treatments are associated with lower levels of co-ethnic trust (Columns (1) and (4)), but this appears to be a function of lower levels of trust in general (Columns (2) and (5)). Indeed, the cash crop treatment is associated with a positive coethnic trust premium (only significant in the ethnic specification, see Column (6)). No robust pattern appears for print technologies. These results only partially aligh with the hypothesis that historical cash crop production engendered more exclusionary identities. However, as these outcomes are only collected in Round (3) of the Afrobarometer, they should be interpreted with caution. Better data, perhaps collected from behavioral field experiments, would be valuable in further testing this hypothesis.

Table IV.14: Cash Crops, Print Technologies, and Ethnic Trust

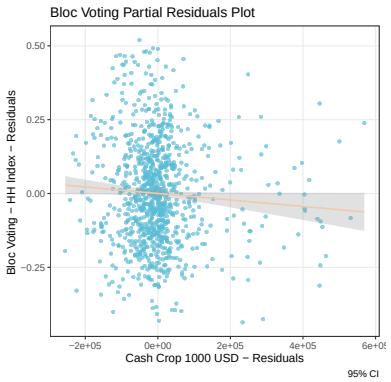
	Trust					
	Geographic-Level			Ethnic-Level		
	(1)	(2)	(3)	(4)	(5)	(6)
Cash crops USD pkm2	−0.021**	−0.021***	0.003			
	(0.010)	(0.005)	(0.010)			
Pubs pth pop (1923)	−0.010*	−0.004	0.004			
	(0.006)	(0.005)	(0.010)			
Cash crops USD pkm2				−0.061***	−0.018*	0.040**
				(0.016)	(0.010)	(0.018)
Pubs pth pop (1923)				0.002	−0.003	−0.010
				(0.008)	(0.007)	(0.012)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No	No	No
Fixed Effect	Country-Round	Country-Round	Country-Round	Town	Town	Town
Ethnic Stayer/Leaver	Both	Both	Both	Both	Both	Both
Mean dep. var.	0.265	0.178	0.916	0.265	0.178	0.916
Observations	20,561	67,269	20,134	20,131	64,614	19,720
R^2	0.124	0.072	0.031	0.272	0.229	0.196

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardized OLS estimates. Standard errors are reported in parentheses and clustered at the location level. Columns (1)-(3) use specification at the geographic level, like in Table 1in the paper, whereas Columns (4)-(6) use specification at the ethnic level, like in Table 2 of the paper. The outcome in Columns (1) and (4) is a binary variable equal to one if respondent has high level of trust towards coethnics, Columns (2) and (5) is a binary variable equal to one if respondent has high level of trust towards most people, and in Columns (2) and (5) it is equal to one if the declared level of trust towards co-ethnics is larger than the declared general level of trust towards most people.

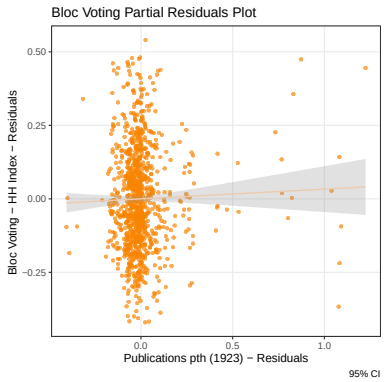
IV.1.2 Bloc - Voting

Another potential observable implication of the politicization of ethnicity induced by cash crops and printing was ethnic group members' coordination on common political preferences and strategies. To test this, we pre-registered analyzing bloc-voting across different ethnic groups. Our regressions are similar to those described in the section on Afrobarometer. The outcome of interest is the Herfindahl Index of political party choices in a hypothetical voting scenario. In Afrobarometer, respondents are asked to choose the party for which they would vote if elections were to be held tomorrow. The results are reported in the online Appendix, in Figure IV.40. We only observe a weak positive association with the publication treatments that disappears when restricting the analysis to ethnicities listed in the bibliographies. However this association is not statistically significant. Further research, maybe using surveys that are representative of each ethnicity, is required to study this question more carefully.

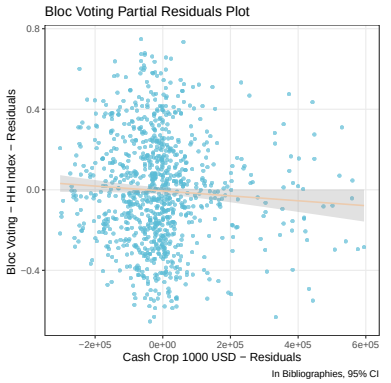
Figure IV.40: Bloc Voting in Afrobarometer



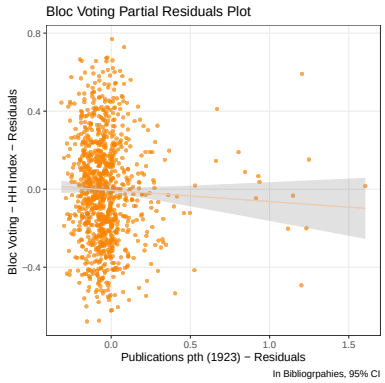
(a) Cash Crops Treatment



(b) Publications Treatment



(c) Cash Crops Treatment - In Bibliographies



(d) Publications Treatment - In Bibliographies

Notes: The figures show partial residuals correlations of the regression model described in section 4. The unit of observation is the ethnicity. The outcome of interest is the Herfindahl Index of political party choices in a hypothetical voting scenario. In Afrobarometer, respondents are asked to choose the party for which they would vote if elections were to be held tomorrow. A larger Herfindahl index indicates more homogeneity in the choice of political parties within the ethnic group. The slope of the liner fits in the plots gives OLS estimate of the treatment (cash crops in Figures IV.40a and IV.40c and print technologies in Figures IV.40b and IV.40d). The regression model includes historical and geographic controls described in the text, as well as country and survey round fixed effects. Figures IV.40a and IV.40b are from the regression with the entire sample, whereas Figures IV.40c and IV.40d are from regressions that restrict the analysis to ethnicity with languages listed in our bibliographies.

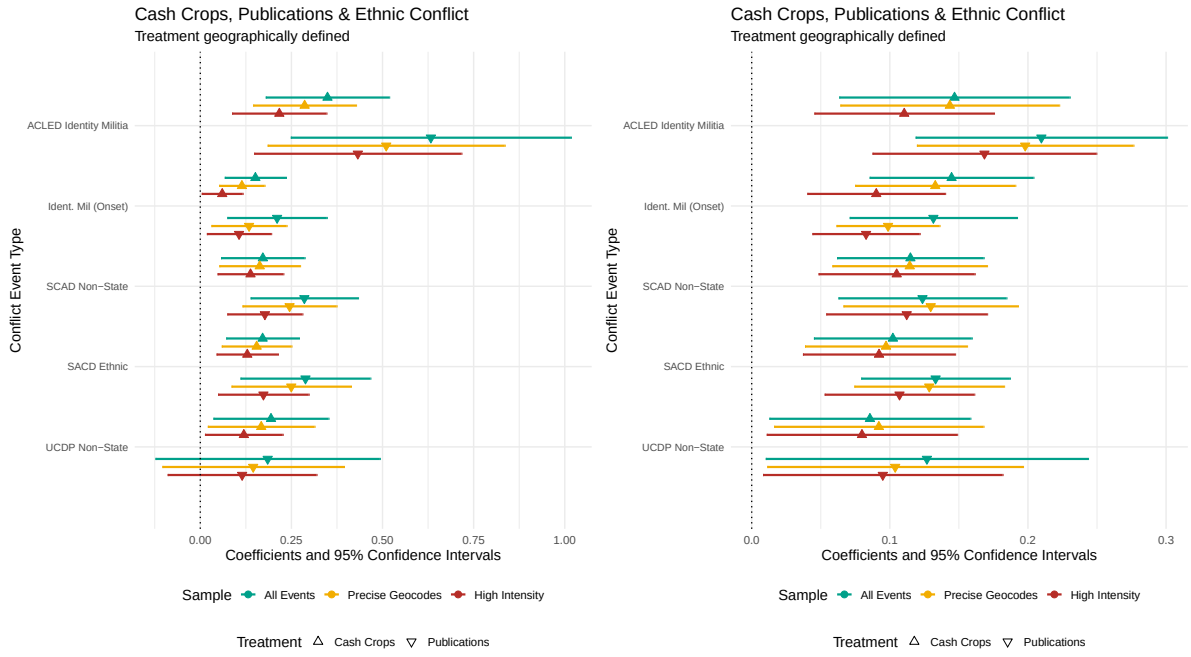
IV.2 Ethnic Conflict

A final pre-registered analysis looks at the impact of cash crops and publishing on ethnic conflict. Theoretically we build on Caselli and Coleman (2013) who argue that ethnic groups with less porous boundaries should be more prone to inter-ethnic conflict as social exclusion helps reduce leakage of spoils from violence. In line with this, we expected that groups treated with cash crops and that developed more exclusionary ethnic boundaries should have higher levels of contemporary inter-ethnic conflict. To test this, we spatially match plausibly ethnicity-related events from three prominent conflict data sets to the WLMS polygons used in the group-level analyses above. First, we count all ACLED events within WLMS polygons where one of the actors involved is

designated as an "identity militia" (Raleigh et al., 2010). We also code an onset version that only counts the first event of the respective militia. Second, we use two conflict outcomes from the SCAD database (Salehyan et al., 2012): All events that SCAD codes as motivated by "ethnic discrimination, ethnic issues" and, alternatively, all events that are classified as "extra-governmental violence", which captures local-level communal conflict. Third, we aggregate all communal/non-state violent events from the UCDP-GED database to WLMS polygons. We then rerun our baseline group-level OLS specifications with binary conflict indicators or logged event counts as dependent variables. To reduce concerns that any findings may be due to geocoding errors and/or reporting bias, we also run models in which we only aggregate precisely geocoded and/or high-intensity events to ethnic polygons (Weidmann, 2016).

The coefficients on our historical cash crop and missionary publishing dummies are positive across all 30 specifications and statistically significant in all but three cases (publication dummies in the three models with the logged count of UCDP non-state events as outcome). The effects of cash crops align with Caselli and Coleman's (2013) framework on ethnicity as a technology of exclusion which increases incentives for conflict. What accounts for the positive and significant effects of publishing? One possible explanation is that, even as print technologies opened the door to the assimilation of linguistically-related outsiders, this merely displaced conflict to higher-level ethnic cleavages—pointing to the potential violent consequences of strengthening imagined communities. Testing this mechanism versus others (such as, communal conflict as a downstream consequence of ethnic politicization or weakened state capacity due to stronger ethnic institutions) represents an important area for future research.

Figure IV.41: Cash Crops, Publications & Contemporary (Ethnic) Conflict



Notes: The figure presents results from models of geocoded conflict events aggregated to WLMS polygons. The left-hand side uses logged event counts $(+1)$ as dependent variables, whereas the right-hand side uses binary outcomes of whether any event is listed in polygon e since 1997 (ACLED) or 1989 (SCAD/UCDP). We use binary indicators for exposure to historical cash crop production and missionary publishing. All models include country fixed effects and confidence intervals are based on country-clustered standard errors. Control variables remain the same as in the group-level analysis above. We compare results across three subsets of conflict events to address concerns about geocoding errors and reporting bias: (a) All events, (b) only events with precise geocodes, and (c) only high-intensity events with precise geocodes (fatal events in ACLED/SCAD, > 5 fatalities in UCDP-GED.)

IV.3 Supplementary, Non-preregistered Analyses

Table IV.15: Non-pre-registered Analyses

Non-pre-registered Analysis	Results
Additional ethnic conflict specifications	
Two ethnic conflict outcomes from the SCAD database (Salehyan et al., 2012) as well as non-state conflict events from UCDP-GED. See discussion in IV.2	See Figure IV.41.
Alternative potential channels of ethnicization and robustness checks	
Effect of group size	See Table III.9, and Figure III.18, III.19, III.20, & III.21
Proximate effects of education, political engagement, urbanization, and wealth.	See Figures III.24, III.25, & III.26.
Pre-colonial political centralization	See Table III.9 & Figure III.18.
British colonial legacy	See Figure III.31 & III.32.
Distinguishing the mode of cash crop production	See Figure III.33 & III.34 and Table III.10
Heterogeneity by local ethnic diversity	See Figure III.35.
Controlling for effects of religion	See Figure III.38.

References

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110(3):512–529.

Albaugh, Ericka A. 2014. *State-Building and Multilingual Education in Africa*. Cambridge University Press.

Baker, CA. 1962. "Nyasaland, the history of its export trade." *The Nyasaland Journal* 15(1):7–35.

Ball, Jeremy Robert. 2003. *The Colossal Lie: The Sociedade Agrícola do Cassequel and Portuguese colonial labor policy in Angola, 1899-1977*. PhD Thesis: University of California, Los Angeles.

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

Bates, Robert H. 1983. Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa. In *State Versus Ethnic Claims: African Policy Dilemmas*, ed. Donald Rothchild and Victor A Olorunsola. Boulder, CO: Westview Press pp. 152–171.

Bederman, Sanford H. 1966. "Plantation Agriculture in Victoria Division, West Cameroon: An Historical Introduction." *Geography* 51(4):349–360.

Brown, Leslie H. 1968. "Agricultural Change in Kenya: 1945-1960." *Stanfords University Food Research Institute Studies* 8:33–50.

Cagé, Julia and Valeria Rueda. 2016. "The Long-Term Effects of the Printing Press in Sub-Saharan Africa." *American Economic Journal: Applied Economics* 8(3):1–31.

Caselli, Francesco and Wilbur John Coleman. 2013. "On the Theory of Ethnic Conflict." *Journal of the European Economic Association* 11(1):161–192.

Church, RJ Harrison. 1969. "The Firestone Rubber Plantations in Liberia." *Geography* 54(4):430–437.

Clarence-Smith, William Gervase, ed. 1996. *Cocoa Pioneer Fronts since 1800*. London: MacMillan.

Cogneau, Denis and Alexander Moradi. 2014. "Borders that Divide: Education and Religion in Ghana and Togo Since Colonial Times." *The Journal of Economic History* 74(03):694–729.

Desmet, Klaus and Stephen L Parente. 2010. "Bigger is better: market size, demand elasticity, and innovation." *International Economic Review* 51(2):319–333.

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

Fieldhouse, David Kenneth. 1978. *Unilever Overseas: The Anatomy of a Multinational, 1895-1965.* London: Croom Helm.

Frankema, Ewout, Erik Green and Ellen Hillbom. 2016. "Endogenous Processes of Colonial Settlement. The Success and Failure of European Settler Farming in Sub-Saharan Africa." *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 34(2):237–265.

Frankema, Ewout and Morten Jerven. 2014. "Writing history backwards or sideways: towards a consensus on A frican population, 1850–2010." *The Economic History Review* 67(4):907–931.

Gibbon, Peter. 2011. "Experiences of Plantation and large-scale Farming in 20th Century Africa." *Danish Institute for International Studies:* Working Paper No. 2011/20.

Guillebaud, Claude William. 1959. *An Economic Survey of the Sisal Industry of Tanganyika*. Welwyn: Tanganyika Sisal Growers' Association and James Nisbet and Co.

Hance, William A, Vincent Kotschar and Richard J Peterec. 1961. "Source Areas of Export Production in Tropical Africa." *Geographical Review* 51(4):487–499.

Hansen, Susan P. 1970. Basic Data on the Economy of Uganda. In *Overseas Business Reports*, ed. US Department of Commerce. Vol. 26 Washington DC: US Government Printing Office.

Hogendorn, Jan S. 1978. *Nigerian Groundnut Exports: Origins and Early Development*. Zaria: Ahmadu Bello University Press.

Iliffe, John. 1979. *A Modern History of Tanganyika*. Cambridge, UK: Cambridge University Press.

Isaacman, Allen and Richard Roberts. 1995. *Cotton, Colonialism, and Social History in Sub-Saharan Africa*. Portsmouth, NH: Heinemann.

Kelly, Morgan. 2019. "The Standard Errors of Persistence." *UCD Working Papers Series* DP 13783(19/3).

Klein Goldewijk, Kees, Arthur Beusen and Peter Janssen. 2010. "Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1." *The Holocene* 20(4):565–573.

Mann, Michael and Valerie Sanders. 1994. *A Bibliography of African Language Texts*. London: Hans Zell Publishers.

Murdock, George Peter. 1967. *Ethnographic Atlas: A Summary*. Pittsburgh, PA: University of Pittsburgh Press.

Nagu, Joseph Tarmu. 1986. *The Development of Smallholder Tea Production in Tanzania: An Economic Analysis of Factors Inflfluencing Green Leaf Tea Output*. PhD Thesis: University of Tennessee, Knoxville.

Ndoye, Ousseynou and David Kaimowitz. 2000. "Macro-economics, markets and the humid forests of Cameroon, 1967-1997." *Journal of Modern African Studies* pp. 225–253.

Nunn, Nathan and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101(7):3221–3252.

Posner, Daniel N. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48(4):849–863.

Posner, Daniel N. 2005. *Institutions and Ethnic Politics in Africa*. New York: Cambridge University Press.

Posner, Daniel N. 2017. "When and Why do some Social Cleavages Become Politically Salient rather than Others?" *Ethnic and Racial Studies* 40(12):2001–2019.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47(5):651–660.

Robinson, Amanda Lea. 2014. "National versus Ethnic Identification in Africa: Modernization, Colonial Legacy, and the Origins of Territorial Nationalism." *World Politics* 66(4):709–746.

Rowling, Frank and Charles-Edward Wilson. 1923. *Bibliography of African Christian Literature*. London: Conference of Missionary Societies of Great Britain and Ireland.

Salehyan, Idean, Cullen S Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull and Jennifer Williams. 2012. "Social conflict in Africa: A new database." *International Interactions* 38(4):503–511.

Smalley, Rebecca. 2013. "Plantations, contract farming and commercial farming areas in Africa: A comparative review." *Institute for Poverty Land and Agrarian Studies:* Land and Agricultural Commercialisation in Africa project Working Paper No. 055.

US Department of Agriculture. 1949. *Foreign Agriculture Report No. 38*. Washington, DC: US Government Printing Office.

US Department of Agriculture. 1950. *Foreign Crops and Markets No. 61*. Washington, DC: US Government Printing Office.

US Department of Commerce. 1956. *Chemical and Rubber: Industry Report*. Washington DC: Business and Defense Services Administration.

Vail, Leroy and Landeg White. 1980. *Capitalism and Colonialism in Mozambique: A Study of Quelimane District*. London: Heinemann.

Van Dongen, Irene S. 1961. "Coffee trade, coffee regions, and coffee ports in Angola." *Economic Geography* 37(4):320–346.

Weidmann, Nils B. 2016. "A closer look at reporting bias in conflict event data." *American Journal of Political Science* 60(1):206–218.