Appendix to the Paper: Cash Crops, Print Technologies and the Politicization of Ethnicity in Africa

Yannick I. Pengl* Philip Roessler† Valeria Rueda‡

All references to "Supplementary Information" relate to our extended online appendix (SI_extended.pdf) available at https://doi.org/10.7910/DVN/WQEQPN

*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

A Additional Results & Robustness **3**

A.1 Robustness of Group-level results 3

A.2 Magnitude of the Effects 5

A.3 Alternative Definition of the Publications Treatment 8

A.4 Addressing Endogeneity: DHS Inter-Marriage Models 9

A.5 Additional DHS Results 13

B Alternative Explanations & Plausibility of Mechanisms **16**

B.1 Group Size 16

B.2 Mediation Models 21

B.3 Treatment Interactions 24

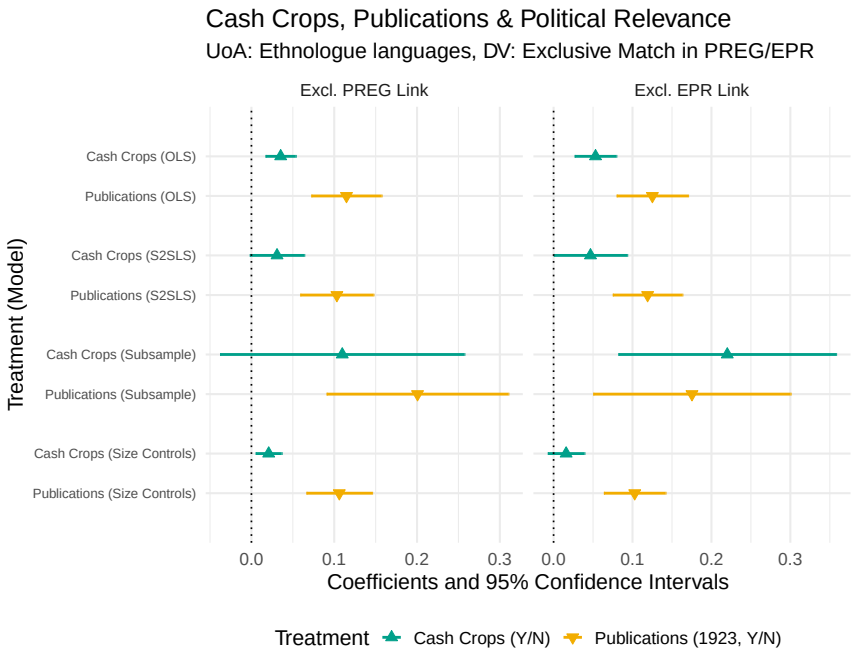
B.4 Heterogeneity by Colonizer 28

A Additional Results & Robustness

A.1 Robustness of Group-level results

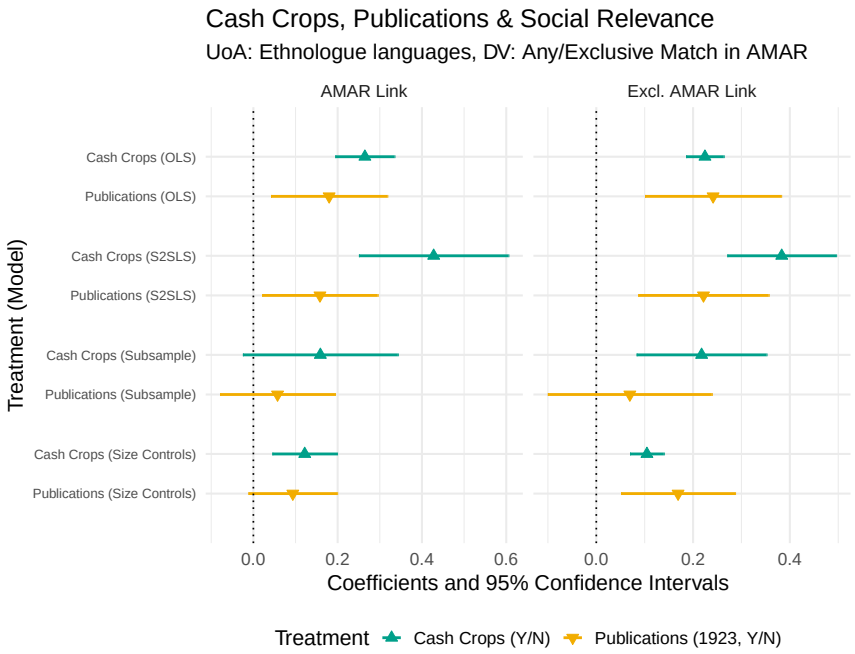
This section presents robustness of the group-level analysis to alternative definitions of the outcome variable.

Figure A.1: 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 A.2: 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.

A.2 Magnitude of the Effects

This section presents results that allow comparing treatment effect magnitudes to other covariates. Table A.1 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 A.3 to A.5 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 A.1: 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 A.2: 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 A.3: 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 A.4: 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 A.5: 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

A.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 A.6), which is our most demanding specification (as it includes town-level fixed effects).

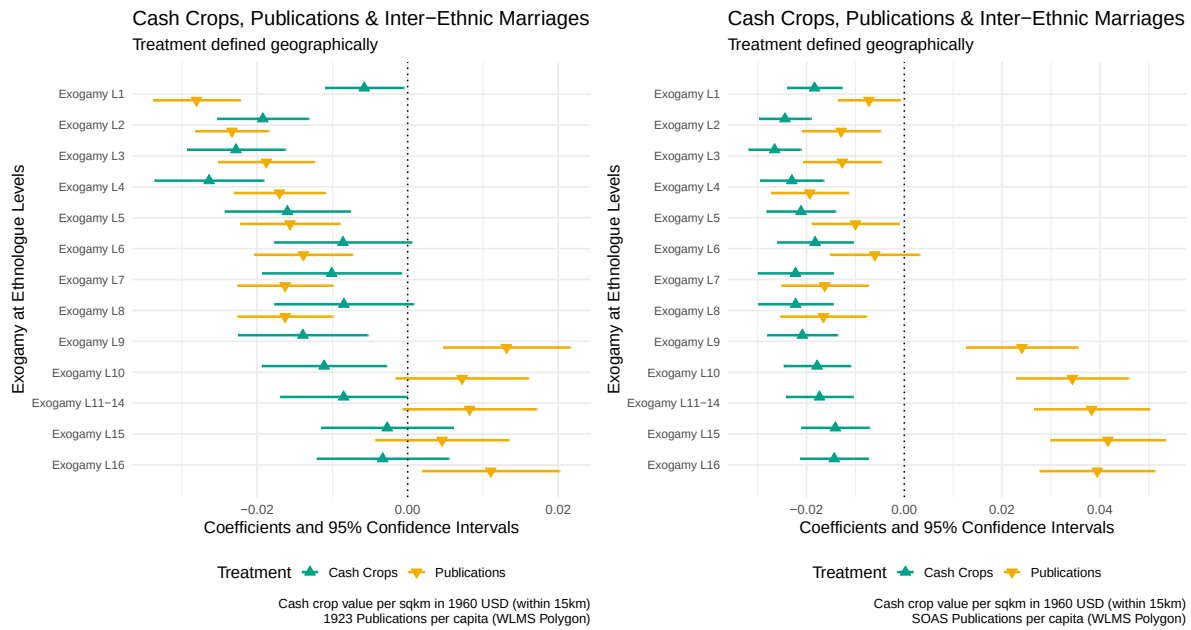
In the DHS, we also test for the robustness of the cash crop and publications effects. Figure A.3 shows two alternative specifications. First, we define cash crops and all control variables at the ethnic polygon level rather than the more local DHS enumeration areas. Second, we use our alternative publishing 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 A.6: 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 A.3: 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.)

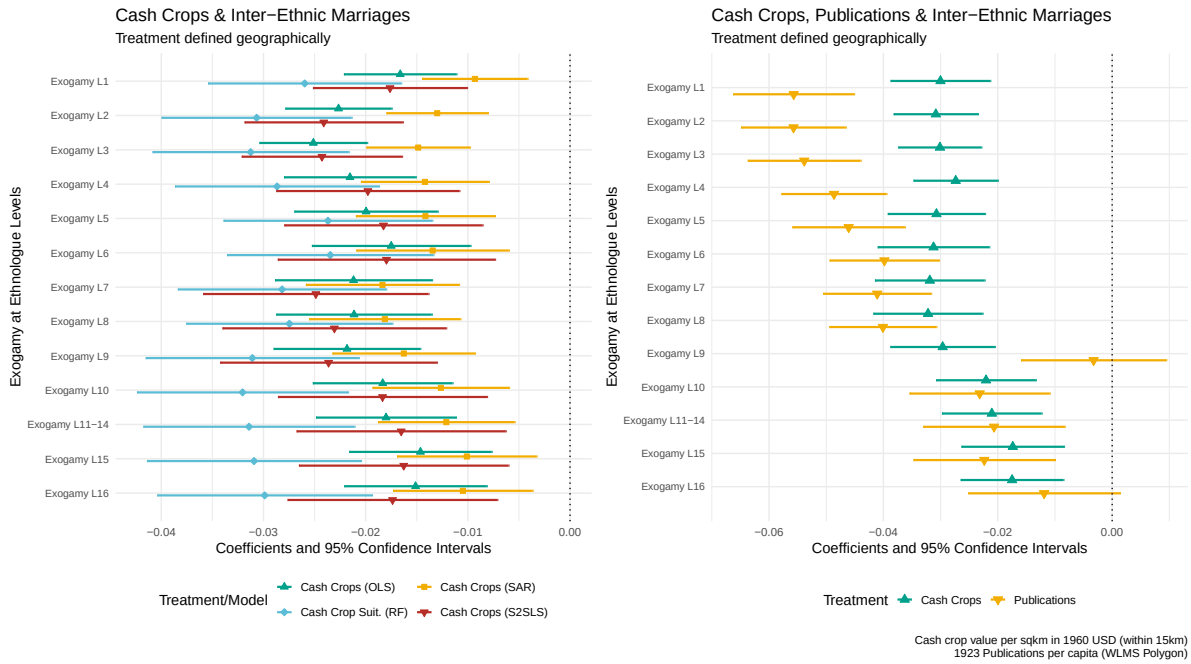
A.4 Addressing Endogeneity: DHS Inter-Marriage Models

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 A.4 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).

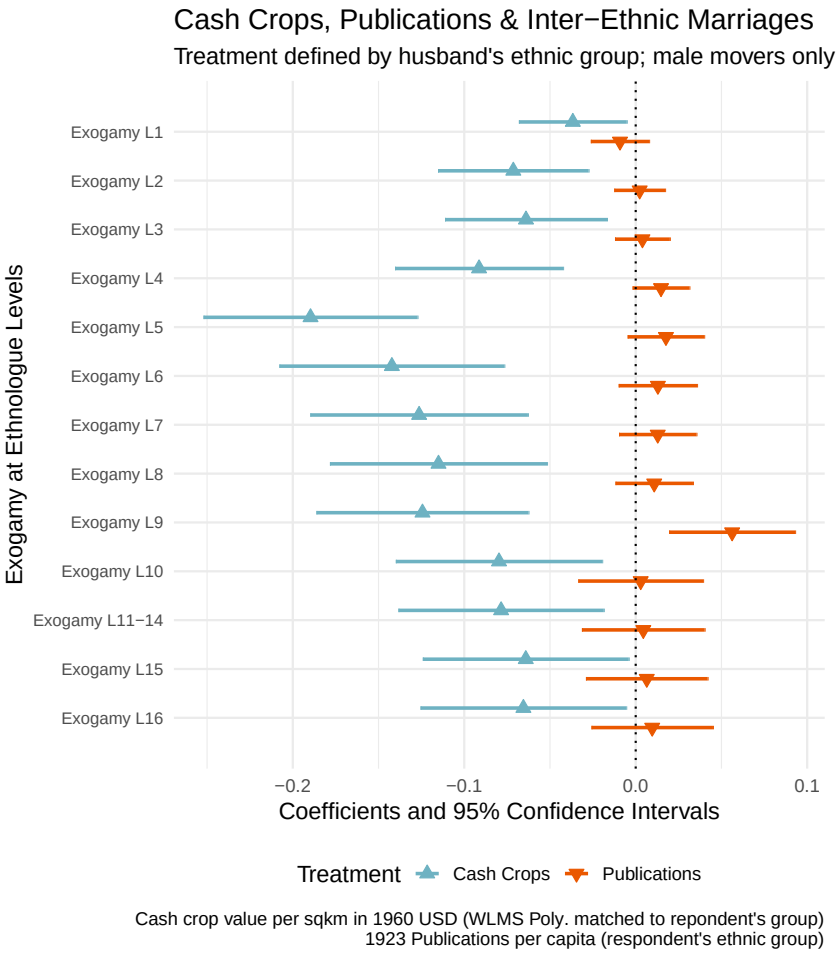
Second, we address the potentially endogenous assignment of missionary language standardization to large and already mobilized ethno-linguistic groups by running intensive margin only analyses. The right-hand panel in Figure A.4 restricts the sample to DHS couples residing in WLMS polygons with at least 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 A.5 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 A.4: 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 A.5: Cultural persistence – Intensive margin models



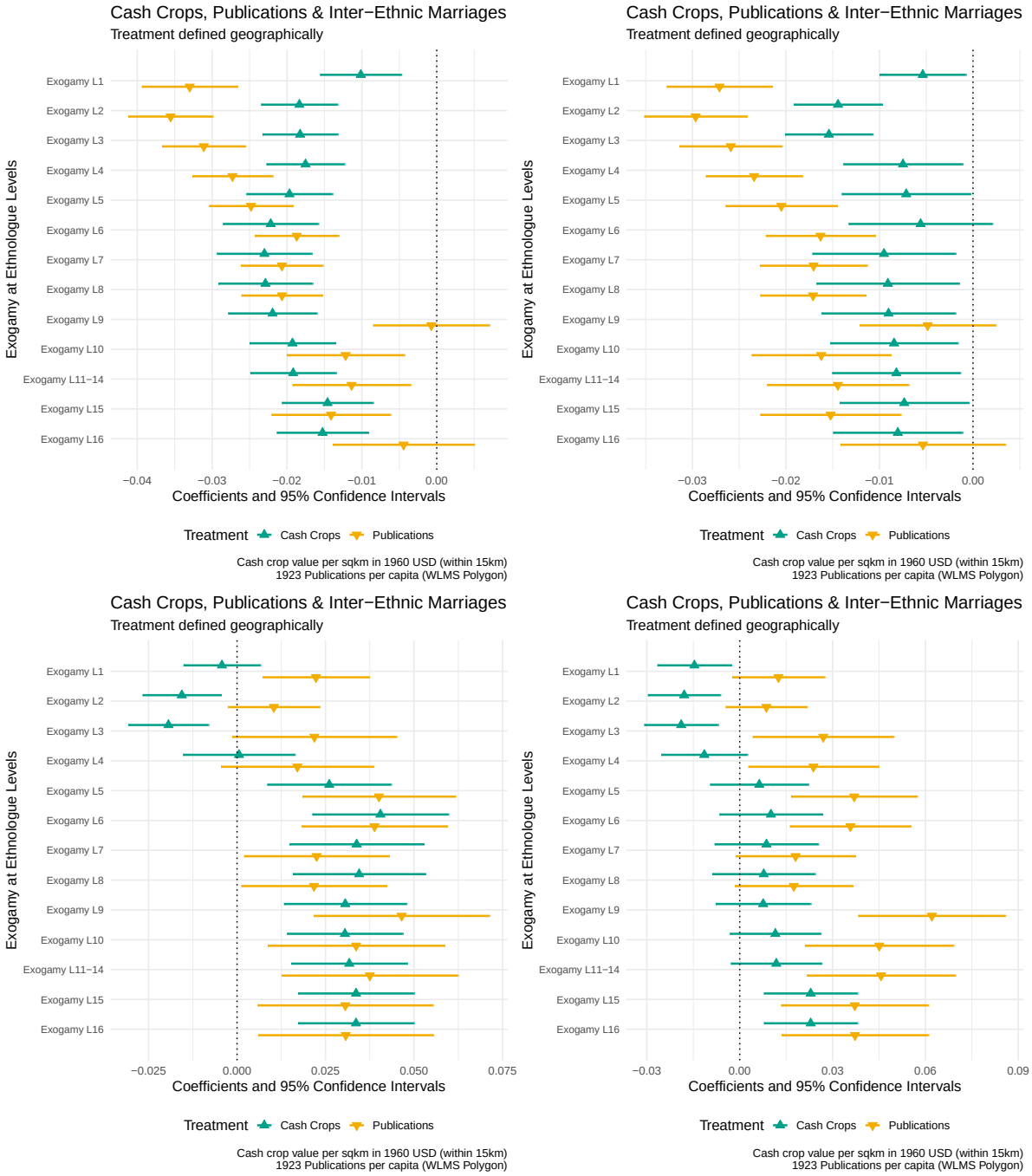
Notes: The figure reports findings from intensive margin versions of our cultural persistence models with location fixed effects. The sample now only includes couples in which the male spouse is an ethnic mover and self-reports to be from an ethnic group with at least one vernacular publication.

A.5 Additional DHS Results

Figure A.6 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.

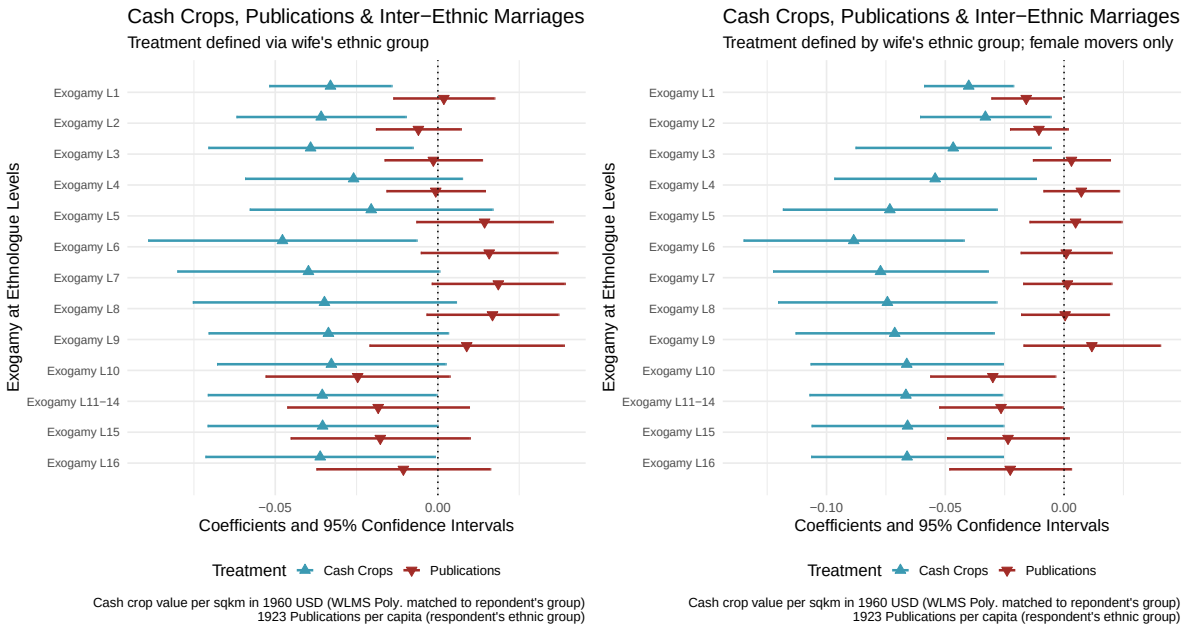
Figure A.7 reports models that assign both treatments based on the wife's instead of the husband's self-reported ethnic group. The cash crop coefficients are comparable to the male movers analysis but the publishing effects are smaller and tend to lose statistical significance. The ethnic persistence effect of vernacular publishing on ethnic boundaries thus seems to be mostly driven by male respondents' marital choices.

Figure A.6: 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 A.7: 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.)

B Alternative Explanations & Plausibility of Mechanisms

B.1 Group Size

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 B.7, Figure B.8)

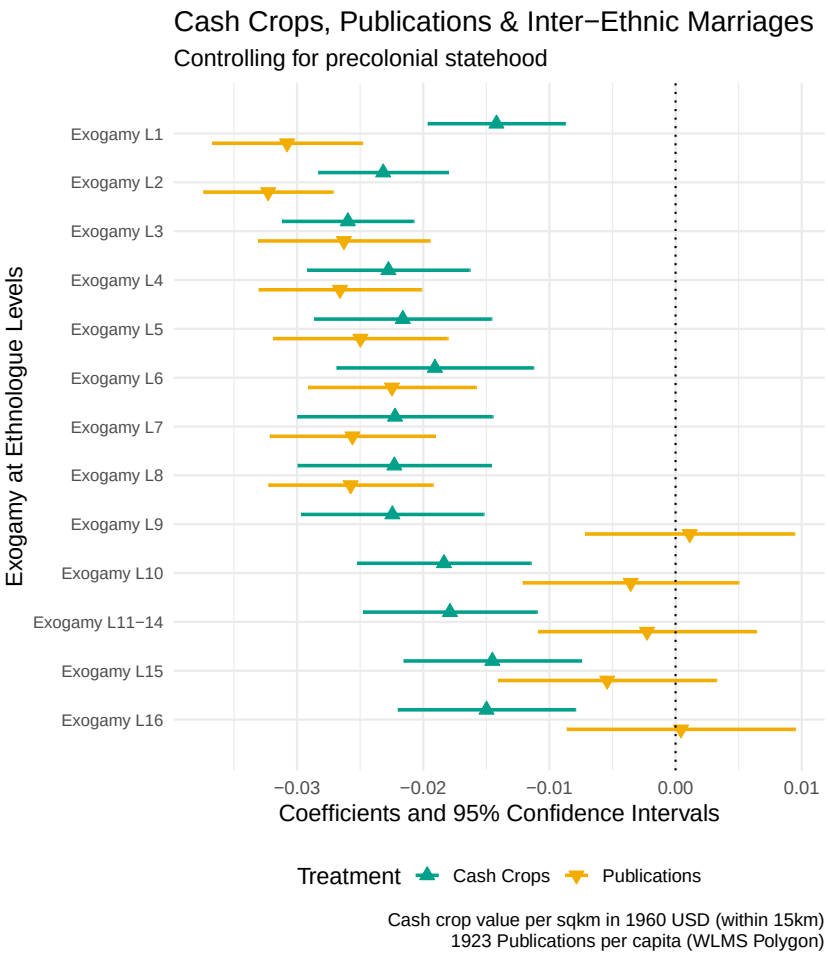
Fourth, we add logged ethnic polygon size to our ethnic leaver models to account for potentially disprortionate measurement error in ethnic leaver status and our geographically matched cash crop treatment among respondents from comparatively small groups. Figure B.9 shows that in our Afrobarometer sample, below-median sized groups have indeed higher shares of ethnic leavers than larger groups. Figures B.10 and B.11 show that our findings in the ethnic leaver specifications remain unaffected when accounting for ethnically matched polygon size.

Table B.7: 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

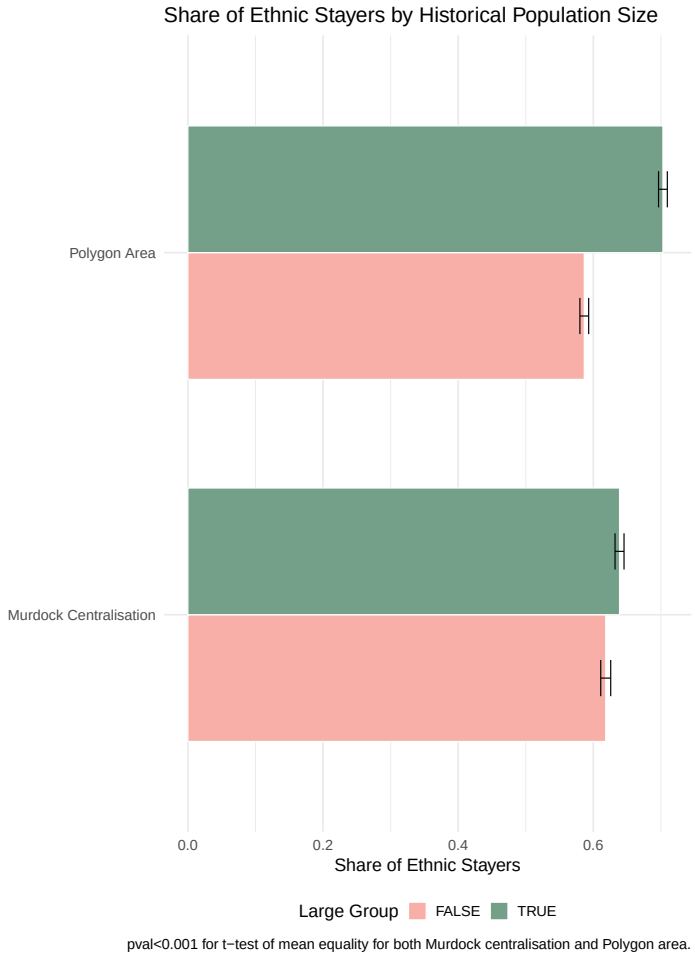
Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments, including Murdock centralisation, are defined at the location level (Columns (1)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

Figure B.8: DHS-Controlling for Pre-Colonial State Centralization



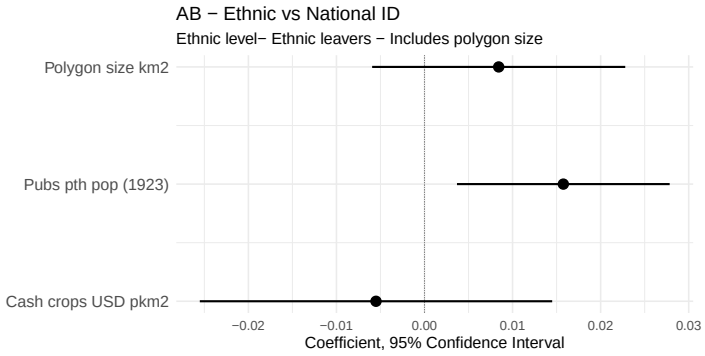
Notes: The figure replicates the geographic persistence analyses of inter-ethnic marriages from the main text. We add a precolonial statehood dummy based on Murdock (1967).

Figure B.9: Average share of ethnic leavers by group size



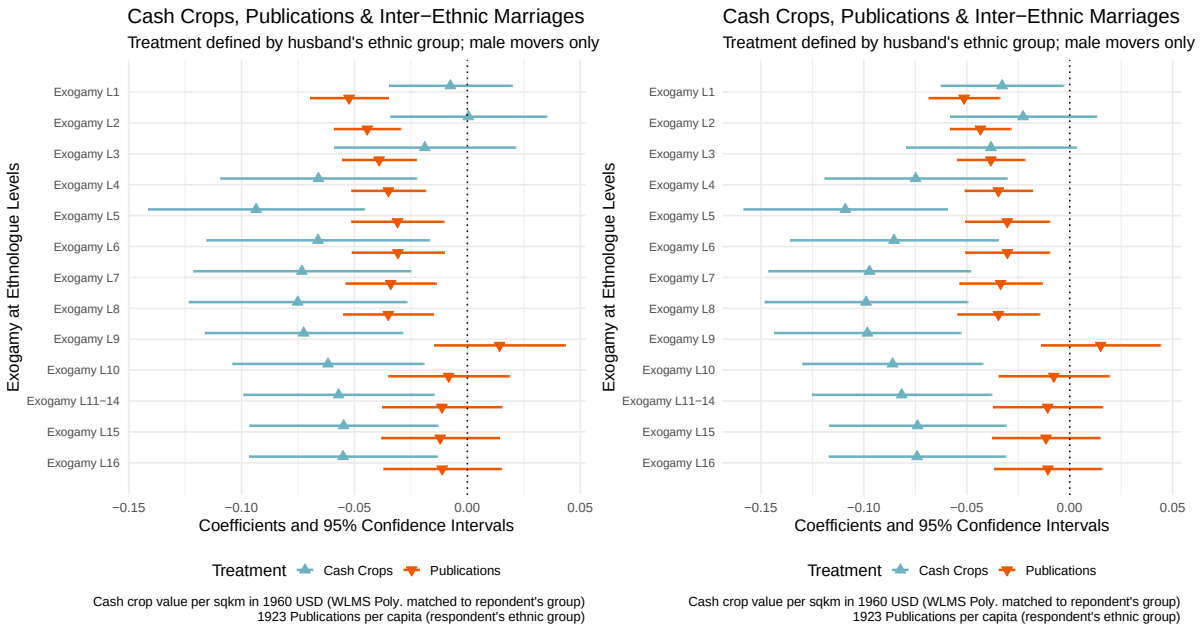
Notes: The figure shows average group size and ethnic polygon size among ethnic leavers and ethnic stayers in the Afrobarometer

Figure B.10: 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 B.11: 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.

B.2 Mediation Models

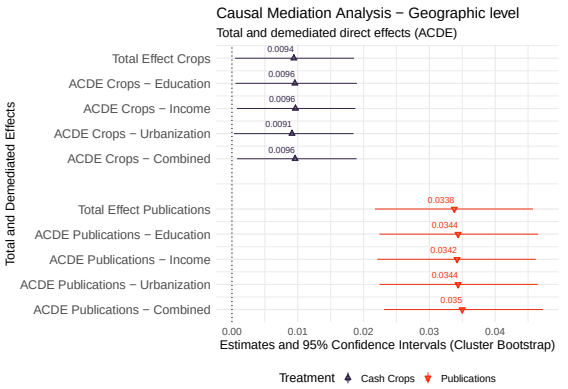
This section investigates potential mediating variables through which our historical treatments may affect contemporary ethnicity outcomes. We follow Acharya, Blackwell and Sen (2016) and estimate average controlled direct effects (ACDE), i.e. the remaining effect after accounting for specific mediators.

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).[2] The results are reported in Figure B.12. 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 B.13 and B.14.

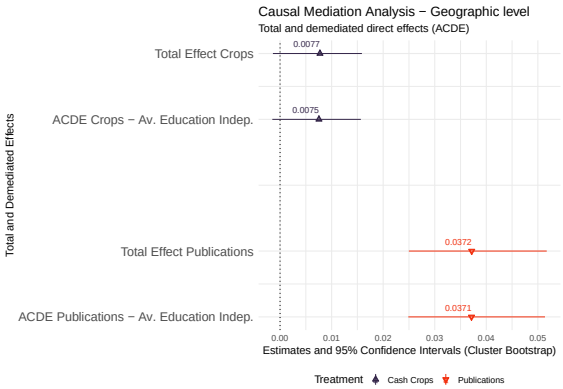
The results suggest that modernization does not explain much of our effects, and in some cases, its effect goes in the opposite direction (lesser ethnic salience and more porous boundaries, see figures B.12a, B.12d, and B.13). Early group-level or location advantages in education explain a small share of the publications treatment for ethnic salience (up to 3%, see figures B.12b and B.12e), but a much larger effect on inter-ethnic marriages (15-43%, see Figure B.14). 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 B.12c and B.12f).

[2] 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.

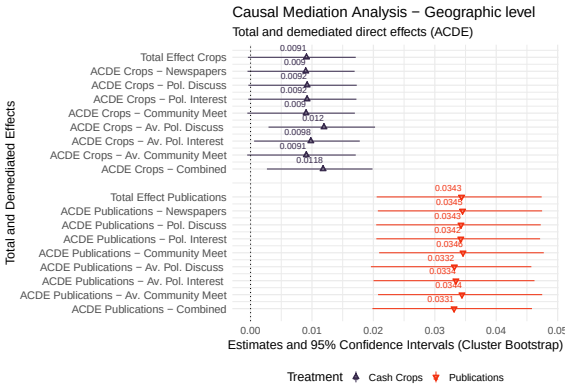
Figure B.12: AB Causal Mechanisms



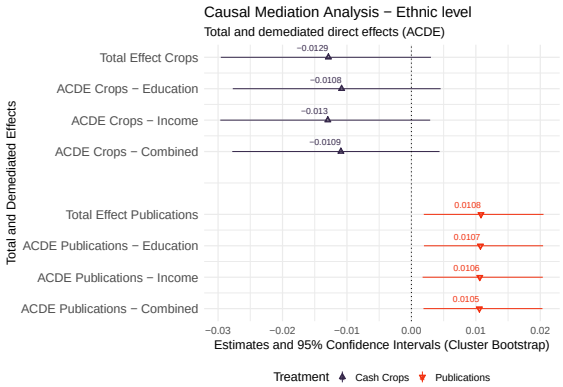
(a) Geographic Level - Modernization



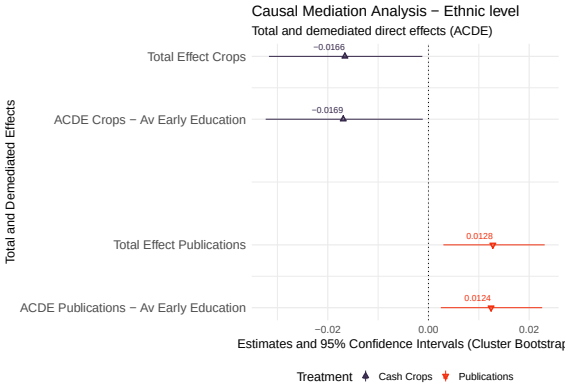
(b) Geographic Level - Early Elite



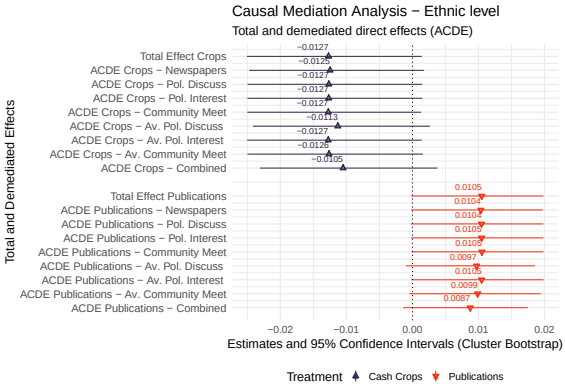
(c) Geographic Level - Public Sphere



(d) Ethnic Level - Modernization



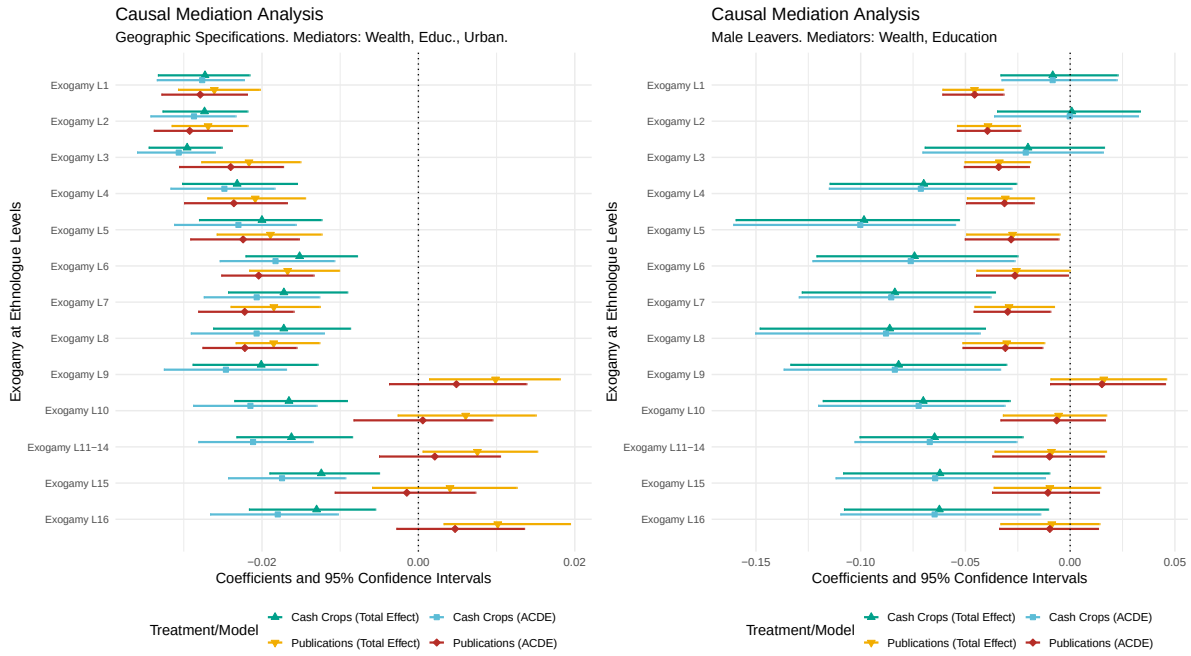
(e) Ethnic Level - Early Elite



(f) Ethnic level - Public Sphere

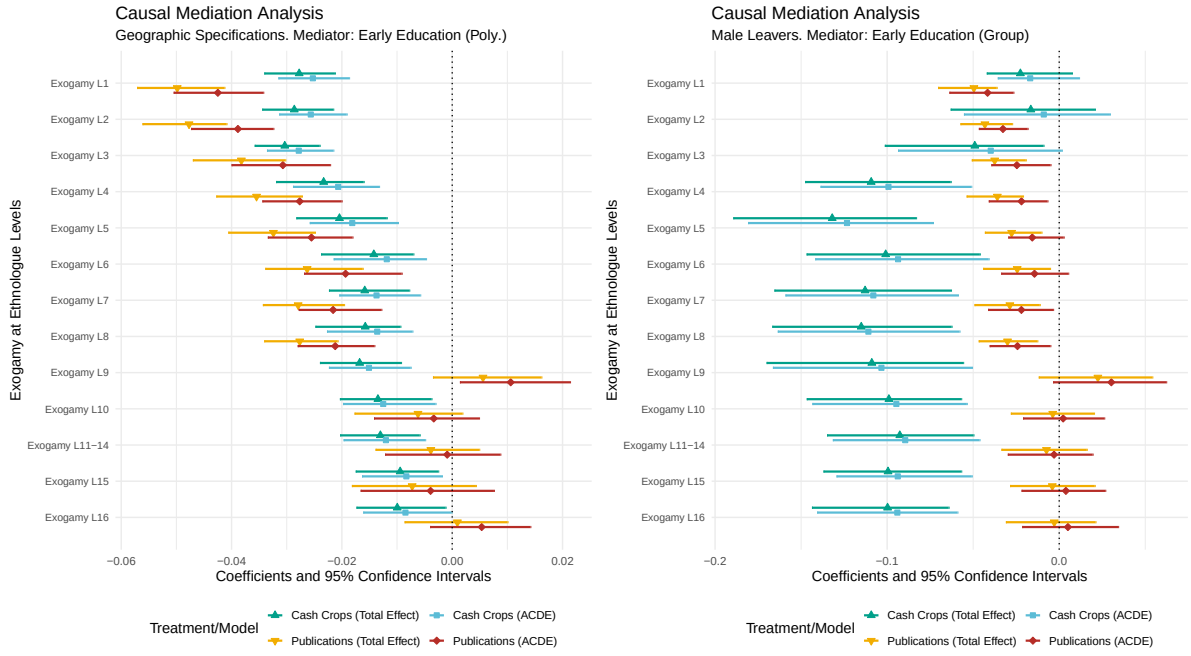
Notes: The figures show the ACDE estimated via sequential g estimation (Acharya, Blackwell and Sen, 2016). Standard errors are estimated through non-parametric bootstrapping with 150 iterations, clustered at the location level. The outcome is a binary variable equal to one if respondent declares a stronger 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 B.2.

Figure B.13: 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 B.14: 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.

B.3 Treatment Interactions

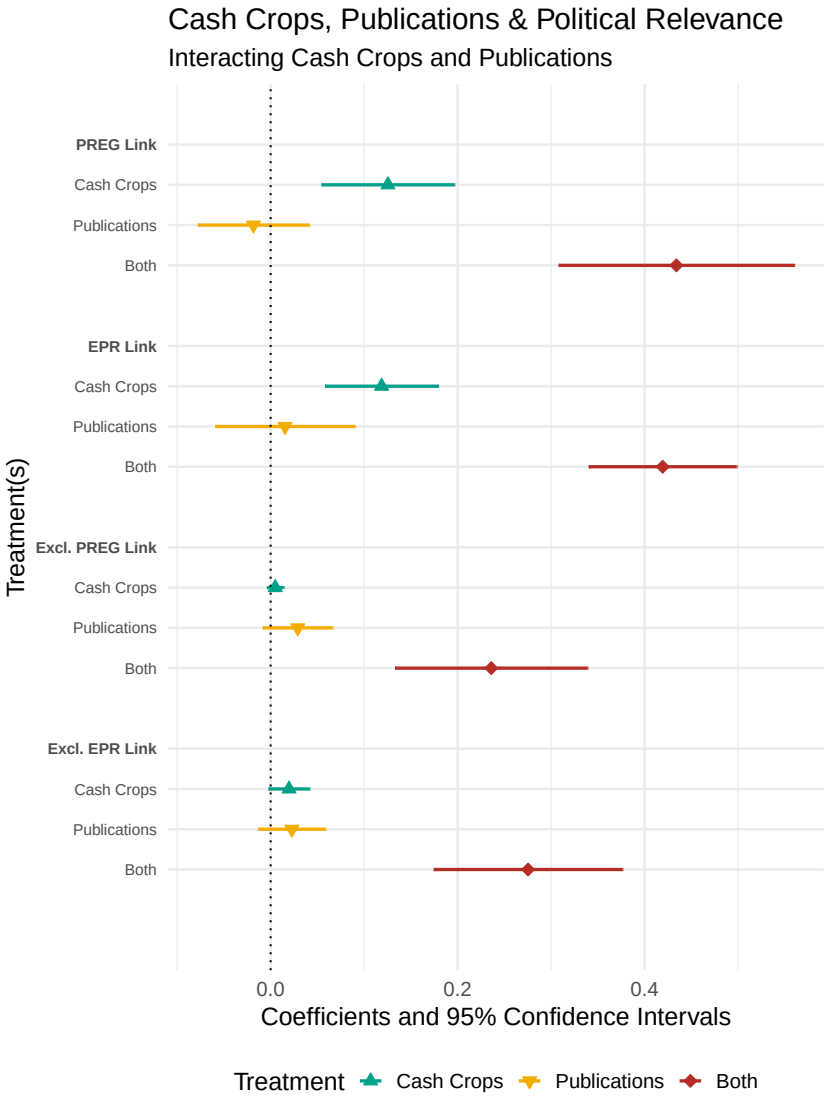
This section investigates potential interaction effects between our treatments of interest, cash crops and publications, across specifications. As our theory predicts similar effects of cash crops and publishing on group-level relevance and individual identity salience, we expect either additive or mutually 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 B.15 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 B.16). 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.

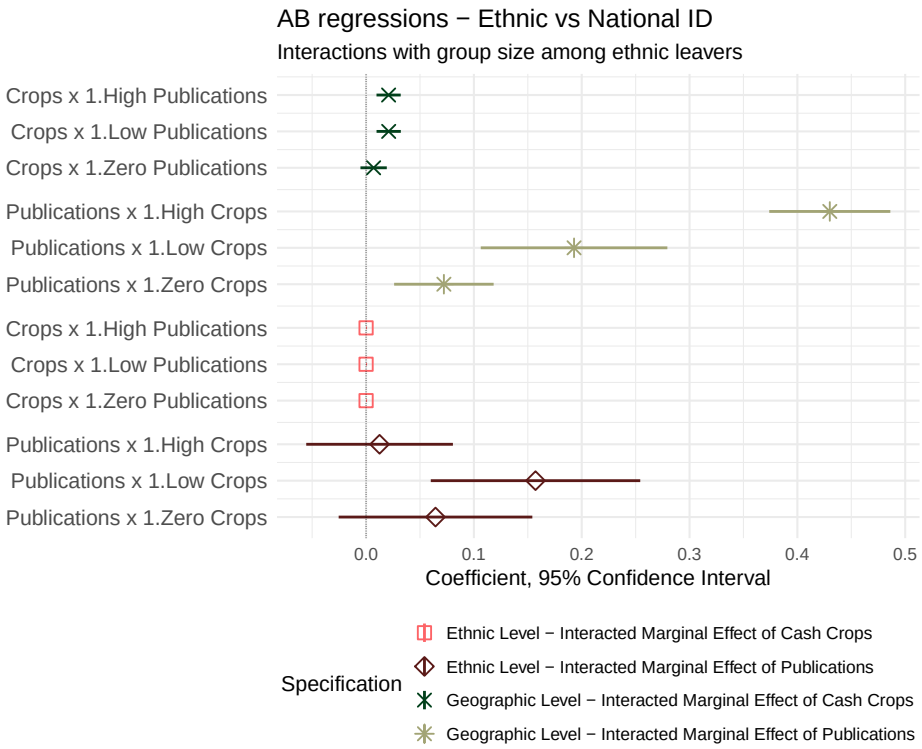
In the DHS marriage analyses Figure B.17), high levels of vernacular publishing tend to dampen the effect of cash crops (left-hand panel). In the absence of any historical cash crop production, missionary publishing is associated with less inter-marriage on levels 1-8 of the Ethnologue language tree 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 B.15: Cash Crops, Print Technologies, and Political Relevance - Interactions



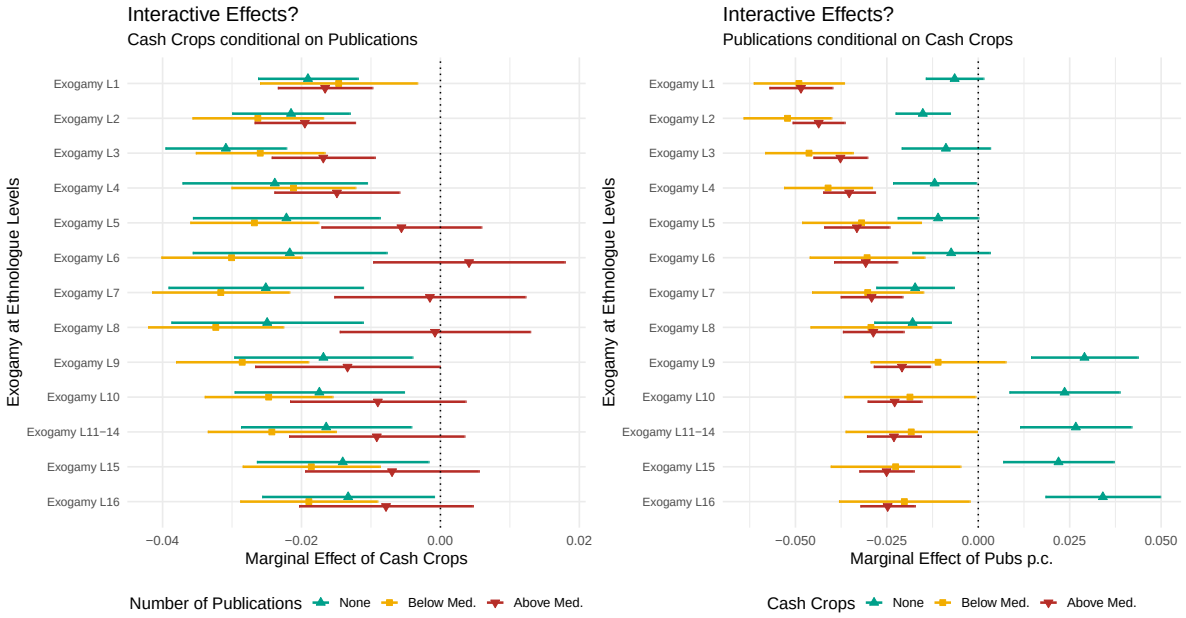
Notes: This figure shows the estimated effects when the two treatments of interest are interacted in the group-level specifications. The regressions are run for the four possible definitions of group-level politicization (exclusive or non-exclusive link in either EPR or PREG). The treatments are binary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

Figure B.16: 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 B.17: Treatment Interactions



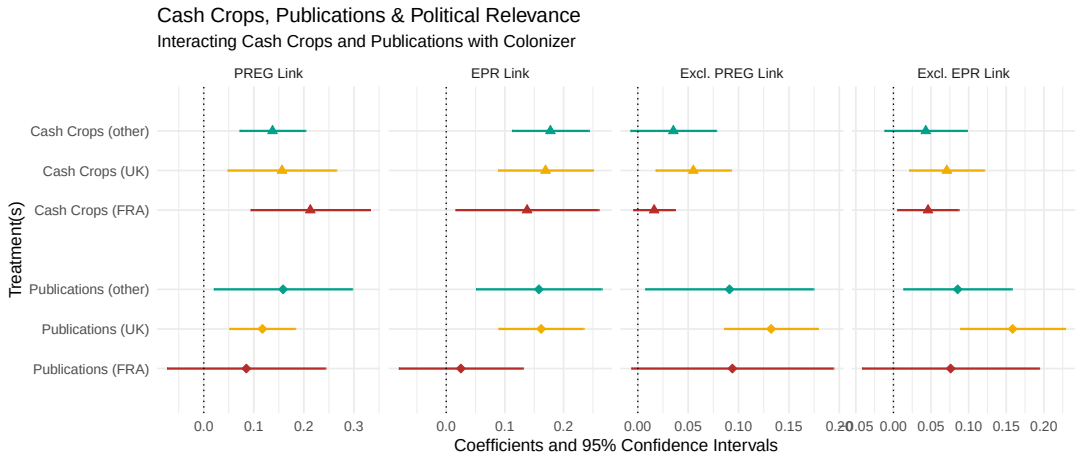
Notes: The figure replicated the geographic models of our DHS-based exogamy analyses but now interacts both historical treatment variables with each other. The left-hand panel shows results from models that interact the continuous cash crop treatment with zero, low, or high levels of publications. The 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.

B.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 B.18) and individual identity salience (Figure B.19 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 B.20). 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).

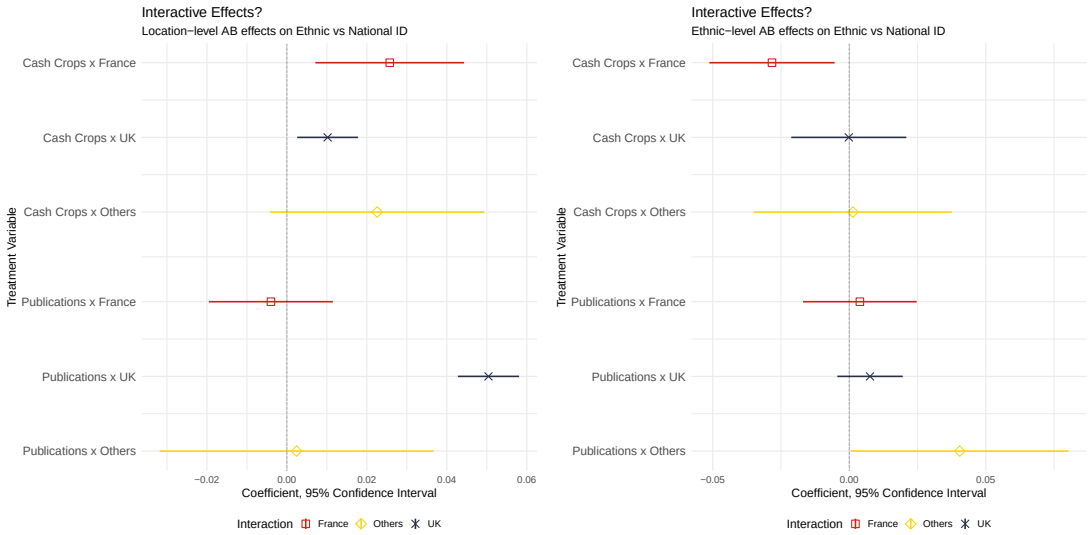
In French colonies, the publication effects on group-level relevance and inter-ethnic marriages are weaker than elsewhere (Figures B.18 and B.20), while the effect on Afrobarometer identity salience disappears completely (Figure B.19). One potential explanation is that French colonial governments put 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 B.18: Group-level Politicization and Imperial Identity



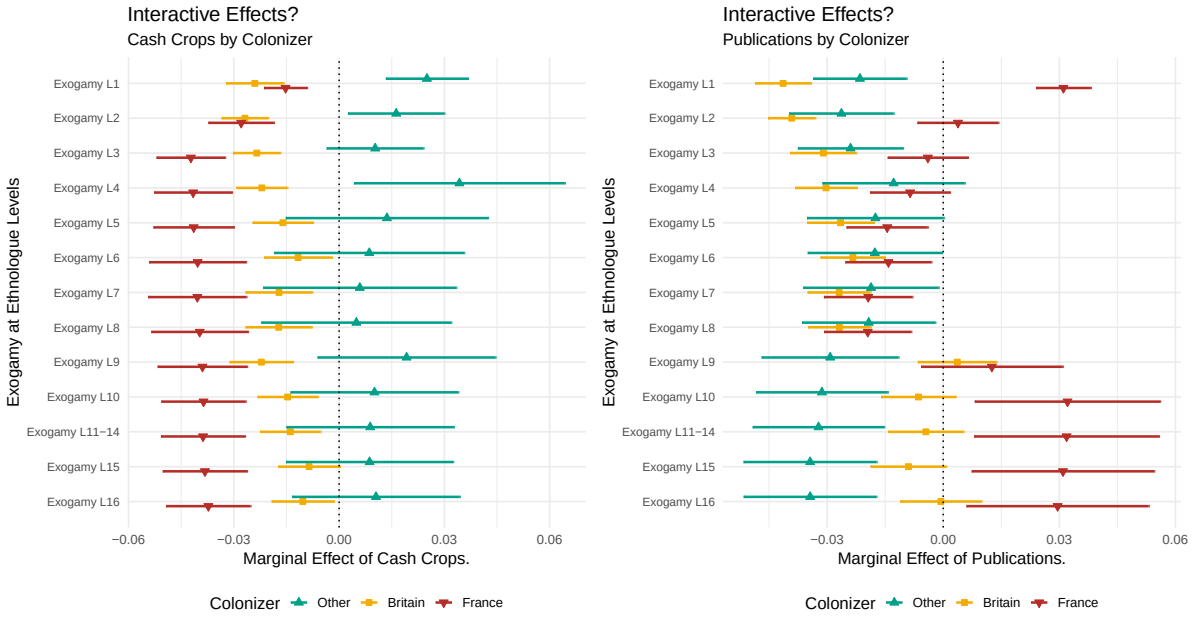
Notes: The figure replicates the group-level specifications with the PREG and EPR-based political relevance outcomes.Our historical treatments are now interacted with binary indicators for the respective country's colonizer (Britain, France, other).

Figure B.19: 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 B.20: 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).

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.

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.

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.

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.

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.

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.

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.