



UNIVERSITY OF
GOTHENBURG

THE LINK BETWEEN ETHNIC FRACTIONALIZATION AND CORRUPTION REVISED

Ethnic Voting in Africa

HÅKAN BERNHARDSSON

WORKING PAPER SERIES 2019:13

QOG THE QUALITY OF GOVERNMENT INSTITUTE
Department of Political Science
University of Gothenburg
Box 711, SE 405 30 GÖTEBORG
December 2019
ISSN 1653-8919
© 2019 by Håkan Bernhardsson. All rights reserved.

The Link between Ethnic Fractionalization and Corruption Revised: Ethnic Voting in Africa
Håkan Bernhardsson
QoG Working Paper Series 2019:13
December 2019
ISSN 1653-8919

ABSTRACT

This paper¹ revisits the relationship between ethnic fractionalization and corruption. Earlier literature argues that ethnic fractionalization leads to corruption via mechanisms involving ethnic in-group favoritism. In this study, an alternative theory suggests that the causal relationship runs in the other direction: when the political system is corrupt and fails to deliver security, voters will fall back on ethnic institutions. This creates the stronger patterns of ethnic identity and ethnic voting that we see in countries considered to be ethnically fractionalized. Conducting three analyses: a regression and an instrumental variable design on the country level, and an individual level analysis on party preferences from the Afrobarometer dataset, the thesis finds support for the alternative theory.

Keywords: ethnic fractionalization, corruption, ethnic voting

Håkan Bernhardsson
Department of Political Science
University of Gothenburg
hakber@gmail.com

¹ This is a revised version of my master's thesis with the same name, finished in May 2019. The empirical material has not changed, but large parts of the theory and presentation have improved since.

Introduction

Previous studies of the link between ethnic fractionalization and corruption have provided much evidence for a correlation between ethnic fractionalization and corruption (Mauro, 1995, 1998; La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1999; Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg, 2003; Alesina & Ferrara, 2005; Glaeser & Saks, 2006; Dincer, 2008). The causal direction between the two variables is generally proposed to go from ethnic fractionalization to corruption, at least when it is explicitly mentioned. However, the previous literature has not provided much empirical evidence for that ethnic fractionalization causes corruption. A handful of papers lay out a theory regarding the causal direction, which is in most cases built on the notion that an inherent ethnic in-group favoritism will lead citizens to prefer their co-ethnics and therefore, conduct in corrupt practices.

This paper will investigate the notion that an inherent ethnic in-group favoritism leads to corruption with the hypothesis of a reverse causal direction between the two variables. The empirics that are presented in the paper concern the degree of ethnic voting in Africa, based on the degree of corruption in countries and the degree of perception of corruption among individuals. The theoretic motivation for that corruption leads to ethnic in-group favoritism is to a great extent built on the work of North, Wallis, and Weingast (2009), and the paper suggests that patron-client networks take over state functions such as upholding the safety of individuals and property in corrupt societies. These patron-client networks rely on the cultural habits of its members to uphold elite privileges without the use of physical force. Enough large absence of state institutions is hypothesized to make ethnic “natural states” implode under the patron-client practices to the extent that it leads to ethnic fractionalization. Conversely, the “open-access society”, in which corruption is absent, will promote values of equality and promote blind justice, regardless of elite status or cultural background among citizens.

The results of the analyses conducted in the paper point towards that corruption cause ethnic in-group favoritism, rather than the other way around. Such a causal direction contradicts the notion that ethnic fractionalization leads to corruption because of an inherent ethnic in-group favoritism.

This paper will proceed as follows: First I will go through the previous research, and especially the theoretic motivations in several papers that suggest causal relationships where ethnic fractionalization

is the independent variable. I will then suggest an alternative theoretical framework. I will then present a research strategy, present results and discuss the implications of the findings.

Previous research

Corruption generally has negative consequences for human development, as it both reduces economic growth and the quality of social services, which means that corruption is negatively correlated with, for example, life expectancy, educational attainment, the standard of living and literacy. The absence of corruption is also a component in Quality of Government, which is positively correlated with environmental sustainability, economic equality, and other measures (Holmberg, Rothstein, & Nasiritousi, 2009).

Ethnic fractionalization has been suggested as one of many causes of corruption in previous literature (Mauro, 1995, 1998; La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1999; Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg, 2003; Alesina & Ferrara, 2005; Glaeser & Saks, 2006; Dincer, 2008). If ethnic fractionalization leads to corruption, it should also lead to several negative outcomes that are correlated with corruption, which has been suggested by other scholars. Studies from the USA show that cities that are ethnically fractionalized prefer lower taxes above public goods provisions (Alesina, Baqir, & Easterly, 1999). Ethnic fractionalization is also negatively correlated with economic growth, quality of policies, and quality of institutions at a country-level (Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg, 2003).

Some of the most cited papers that suggest a causal direction in which ethnic fractionalization per se leads to corruption are Mauro (1998), La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1999), Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003), which are papers that lack thorough and explicit theory about the mechanisms of ethnic fractionalization. In the next section follows a walkthrough of some papers that motivate a theory behind why ethnic fractionalization can affect both corruption and other related variables, as an orientation to the current state of theory, and criticism of the strengths and weaknesses of each respective theory.

Alesina and Ferrara (2005) propose three reasons behind the causal direction between ethnic fractionalization and corruption. The rest of the theoretical motivations will be categorized after their similarity with Alesina and Ferrara (2005), as all cited papers bear similarity to that work. The first

explanation is ethnic in-group favoritism, meaning the aggregation of individual preferences of attributing positive utility to the well-being of one's group. Putnam (2007) carries on this idea, using ethnic categories as a base for social identities, where the distance between social identities should lead to distrust and vice versa. Using this logic, Putnam argues that ethnic fractionalization leads to lower levels of social trust, lower confidence in local government, or lower likelihood of giving to charities, and he supports his claims with data from the USA. The theory does, however, have contradictory empirical findings, as it also shows that the ethnic in-group levels of trust will decrease in ethnic groups in ethnically diverse communities, while the differences between ethnic in-group trust and trust towards other ethnicities are uncorrelated with fractionalization (Putnam, 2007).

Dincer (2008) refers to that in-group favoritism within ethnicities can be a source for corruption. In his short motivation for the in-group favoritism among co-ethnics, he refers to the works the anthropologist Van den Berghe (1987) and Vanhanen (1999) for further reading. However, Van den Berghe (1987) suggests that the type of social networks that can be subject to in-group favoritism consist of a few hundred members, meaning that a theory built on his work regarding ethnic in-group favoritism should be flawed when applied to ethnicities that can amount to up to millions of members. Vanhanen (1999, 2012a, 2012b) suggests that "we can trace the roots of ethnic conflict and violence to human nature" (Vanhanen, 2012a), and defines what he calls ethnic nepotism, which he argues is an evolutionary drive for favoring one's ethnicity. He supports his theory by establishing the correlation between ethnic fractionalization and ethnic interest conflicts in countries (Vanhanen, 2012a). However, ethnic fractionalization should be a necessary precondition for ethnic interest conflicts, and there should be little need of studying the extent of ethnic interest conflicts in homogeneous countries. Van den Berghe (1987) also argues that in most cases when genetic differences can be observed between people, this is the result of long-distance migration. Ethnic interest conflicts, such as conflicts between tribes, should instead at least historically have occurred in environments in which it is almost impossible to decide ethnicity based on appearance. It should also be noted that many ethnic conflicts are built upon religion and/or language, rather than "race".

The second explanation by Alesina and Ferrara (2005) is that ethnicity affects the strategies that individuals use during market transactions. In places with market failures, ethnic affiliation is used as a reputation mechanism, which evens out information asymmetry, with the side effect of losses

attributed to generalizations. Ethnicities will mainly be necessary when legal contracts cannot be enforced due to weak institutions, as the stakes of economic cooperation are not as high where there are means to resolve legal disputes. This implies that ethnicities play a particularly important role when the legal power of the state is weak, which means that state institutions play an essential role in explaining ethnic conflicts. This points to a causal direction from weak state institutions to ethnic in-group favoritism.

Glaeser and Saks (2006) propose that if corruption is introduced, to begin with, it will persist in ethnically fractionalized societies, as voters will not be interested in removing the leaders in charge of the corruption as long as they provide resources to the voter's ethnic group. Their model at best implies that ethnic fractionalization can be preserved or increased in already corrupt societies, but not why ethnic fractionalization leads to corruption, to begin with. They investigate their theory using an OLS regression, but there is no obvious reason as to why there should exist a linear relationship between ethnic fractionalization and corruption if the causality is solely built on that other variables are a prerequisite for corruption.

The third explanation by Alesina and Ferrara (2005) relates to that the cost of production increases with ethnic diversity as a result of difficulties in communication over lingual or cultural lines. This theoretic motivation, however, loses much of its explanatory power when applied to other types of differences than languages, such as race or religion, as they should technically not affect communication so much.

One of the most thoroughly defined theories regarding the causal relationship between ethnic fractionalization and corruption is Cerqueti, Coppier, and Piga (2012), who have created an advanced model to investigate details of the relationship. In sum, they propose a principal-agent environment consisting of entrepreneurs, bureaucrats, and controllers. Entrepreneurs rely on bureaucrats to run their businesses, and the bureaucrats are controlled by the controllers who can give them fines if they do not follow the laws. In the model, the controllers are assumed not to report the activity of bureaucrats belonging to the same ethnicity. Another assumption of the model is that a higher fractionalization will increase the monitoring costs due to communications barriers which should, in turn, reduce the monitoring level, and thereby increase the level of corruption. There will, therefore, exist an optimal monitoring level for the state, where the monitoring costs and the losses of corruption are balanced. All actors in the model are assumed to act rationally except the controllers, that do not

maximize utility because of assumed characteristics of ethnicity. Moreover, the thought of non-corrupt controllers is an exciting assumption as the other type of bureaucrats in the model are assumed to be as corrupt as possible. However, in this model, the controllers even have a corrupt function. The principal-agent assumptions in the modeling of corruption have also been criticized for a mischaracterization of systemic corruption (Persson, Rothstein, & Teorell, 2013). The internal contradictions among the assumptions, as well as the issue with external validity to systemic corruption make the model flawed. In summary, much of the previous research suggests that ethnic fractionalization leads to corruption and other negative outcomes, but few scholars have a thorough theoretical foundation for their claims. Those who have an explicit theory regarding the effects of ethnic fractionalization, generally base their theories on that people are inherently subjects of ethnic in-group favoritism. The possibility for a reverse correlation or impact from ethnic fractionalization through a spurious relationship has not been discussed in the presented literature.

Theory

The proposed theory outlined in this paper suggests that ethnic identities are neither given from nature nor the Tower of Babel, but rather through political interaction within groups with shared interests in security. Another theoretic viewpoint that will be challenged is that people will favor their co-ethnics under all circumstances. A contending explanation of the correlation between ethnic fractionalization and corruption will be suggested, based on the work by North, Wallis, and Weingast (2009), who have studied the evolution from what they call natural states to open-access societies.

According to the theory, in very small societies, consisting of a group of families that all know each other, the person or group of people who are best at coercing others by violence or threats of violence will gain power by upholding the security of others. When a society gets so big that one cannot possibly know each member of the society on a personal basis, the coercion by the ruling elites, which is a requirement for the absence of violence, cannot build on individual relations. In societies larger than around 1000 members, “individual relationships cannot be based solely on personal knowledge and trust; they must be reinforced by the web of interests created by the social order” (North, Wallis, & Weingast, 2009).

Some of the earliest human civilizations have been theocracies, that extract rent from society to priest-politicians. The social consequences of organized religion also help the authorities maintain social control, but religion is not completely necessary to maintain the interests of elites (North,

Wallis, & Weingast, 2009). One's culture will contain a set of values regarding how to interact with others, and breaking the conventions will result in punishment from the surroundings. As the social order is built as a patron-client-network, with the elites responsible for the security of a community in the top, there are great incentives for cooperation within the network, as one's position in society depends upon the goodwill in the personal relationship with people higher up in the hierarchy. Natural states limit the ability to create organizations that are separated from personality and identity (North, Wallis, & Weingast, 2009). Therefore, ethnicity and cultural upbringing, rather than a somewhat independent civil society, becomes a more evident base for political mobilization in the natural state.

In the open-access society, the police and the courts will uphold impersonal security, meaning it is independent of personal relations and personally motivated actions from the ruling elites. Such a society will lead to the promotion of beliefs about inclusion, equality and shared growth, which will be norms based on the reality of perceived fairness of opportunity and impartial distribution of welfare (North, Wallis, & Weingast, 2009). In contrast, the police and courts in a natural state will likely depend upon the personal characteristics and properties of the rulers in charge of those functions, and act directly in their interests.

In countries with dysfunctional state institutions, the shared beliefs within a culture will lead its members to embrace the means of upholding their security — namely to submit to a ruling elite that relies on cultural rules to secure the community from violence. The theory bears much resemblance to the thoughts of Marx and Gramsci, who argue that religion and culture are used by the elites to maintain class society, with the addition of North, Wallis, and Weingast (2009) that the absence of violence makes such an order much more tolerable for the people.

I suggest that ethnicities can be a base for a natural state, as ethnicity is often built on at least one of the two foundations culture and religion. I also suggest that ethnicity is more of an artifact from earlier times in countries that can be characterized as open-access orders.

A consequence of the connection between the ethnicity and natural states that depend upon personalized patron-client networks is that such networks should implode under their weight when they become so large that too many people are involved, and too many personal relationships need to be maintained to extract rents to the elites. Under these circumstances, the ruling elites might either implement a more efficient bureaucracy in the rent extraction, meaning less corruption or the patron-

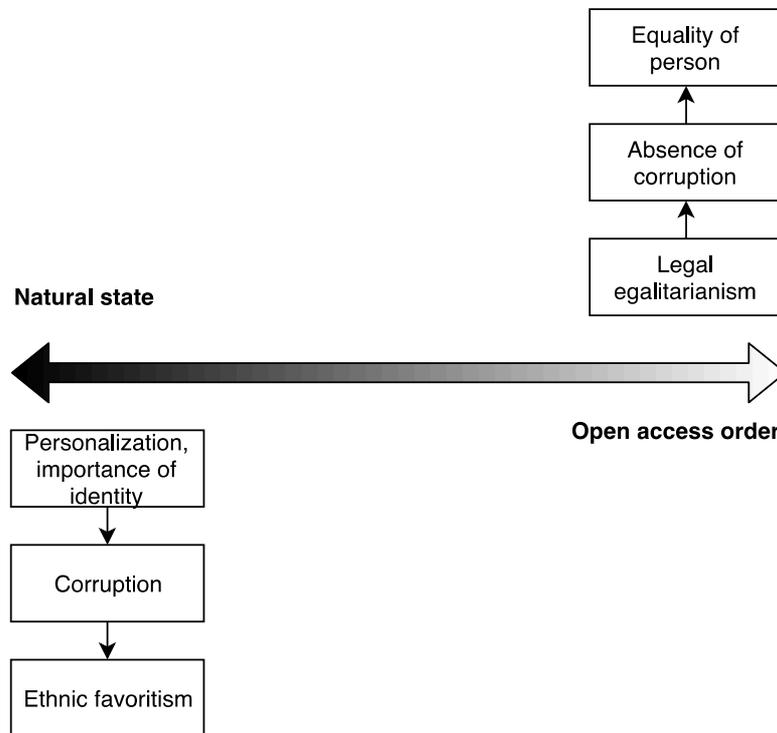
client networks can divide into new hierarchies that can either choose to co-operate or struggle with each other. Such a mechanism could explain a causal direction going from corruption to ethnic fractionalization.

If people are used to that the institutions that govern their behavior only apply to particular parts of the population, and that universal welfare is impossible due to corruption, it is not so surprising that voters will vote for particularized welfare, meaning transfers from tax to particular groups, as a quest for a fair provisions of public goods. The most practical way of arranging such clientelist practices will be through the transfer of public goods to ethnic groups, as they already have cohesion and identity that can be mobilized in elections. In other words, in a very corrupt country, the only collective action that could at least guarantee some welfare to one's collective group would be to vote for that public goods should be distributed in a particularistic manner.

Orjuela (2014) argues that the sentiment of "it is our time to eat" might drive ethnic minorities towards voting according to their ethnic interests if they already are discriminated against politically. The plain fear of being disfavored by what other ethnic groups would do with their power could turn voters belonging to the majority groups toward ethnic parties even though the voters are not primarily motivated by ethno-nationalistic tendencies, but just a fear of repression and a fear of losing material privileges. In summary, the natural state that is characterized by a dependence upon personal relationships will be built on networks of favors and co-favors, meaning corruption and a need for ethnic in-group favoritism. Meanwhile, the open-access order will be characterized by the absence of personal relationships as the base for public interaction, meaning an absence of corruption and values of equality between ethnicities among the population. This is modeled in figure 1.

Based on the proposed theory, I will investigate the hypothesis: *Corruption will lead to ethnic in-group favoritism.*

FIGURE 1, VISUALIZATION OF PROPOSED THEORETICAL MECHANISMS



Research strategy

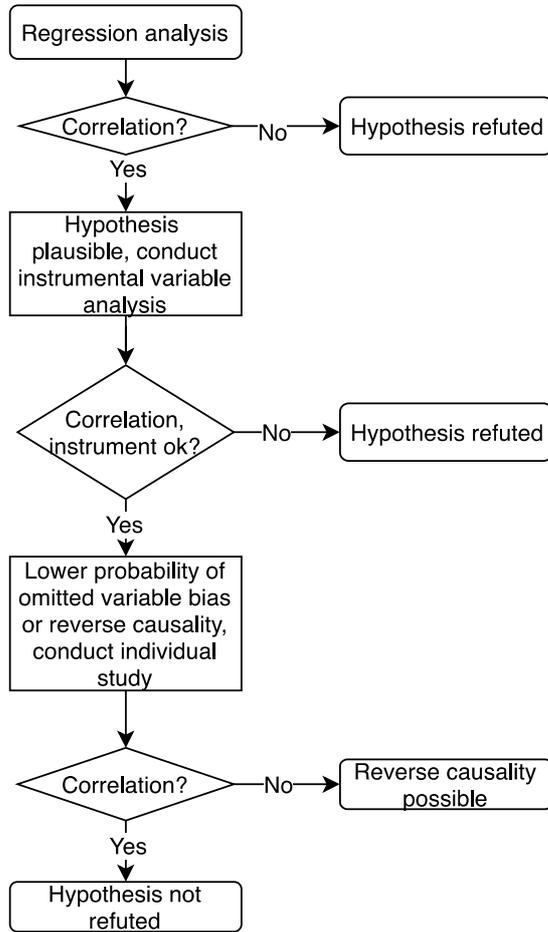
This study will be conducted at two levels. I will first study the country-level effects of corruption on ethnic in-group favoritism, measured as ethnic voting. The first aim is to establish whether or not we can observe a correlation between corruption and ethnic voting, using regression analysis. Without establishing a correlation, it should be very improbable with a causal relationship.

The second aim of the country-level analysis is to test the hypothesis that corruption per se leads to ethnic voting and that the phenomenon is neither explained by a spurious relationship, nor by reverse causality. As this paper criticizes much of the previous theory that is built upon that inherent ethnic grievances and ethnic in-group favoritism lead to corruption, I will also employ an instrumental variable regression, as the causality and causal direction are highly relevant. The purpose of the instrumental variable regression is to get as close to an experimental setting as possible, by isolating the effects of corruption on ethnic voting to the greatest possible extent.

The answer to whether individuals who perceive a high degree of corruption will vote for ethnic parties to a greater extent, is yet another piece of evidence to the question of the causal mechanism between corruption and ethnic in-group favoritism. The suggested theory proposes that natural states will lead to more corruption, which leads individuals to understand that universal welfare is impossible and that their ethnic group is necessary for their welfare and security. From that suggestion follows that corruption should lead the individual to embrace their ethnicity, which includes voting for an ethnic party. If we can measure a correlation between the perception of corruption and ethnic voting on the individual level, the findings will suggest that the theory is plausible in this regard.

How should this research design respond to a reality in line with the antithesis to the hypothesis suggested in this paper, that ethnic fractionalization leads to ethnic voting, which leads to corruption? The correlation should still be established. The reverse causality should be caught by the instrumental variable regression, but there is always a risk of a false positive. In the individual study, there should however be no obvious reason for the perception of corruption to be correlated with ethnic voting, as the corruption caused by ethnic voting should be equally noticeable. A summary of the research strategy and how it will respond to the underlying data be in figure 2.

FIGURE 2 VISUALIZATION OF THE STEPS INVOLVED IN THE RESEARCH STRATEGY



The combined results of an individual level study and two country-level studies should give us a fairly good view of how the structural effects of corruption cause ethnic voting on a macro-level as well as on a micro-level.

Data

The primary sources of data in this study are the Afrobarometer and the Quality of Government dataset (Bratton, Mattes, & Gyimah-Boadi, 2015; Teorell et al., 2019). The Afrobarometer rounds span over the years 2005, 2008, 2013 and 2016. A summary of the sources of the variables can be found in the appendix.

Operationalization

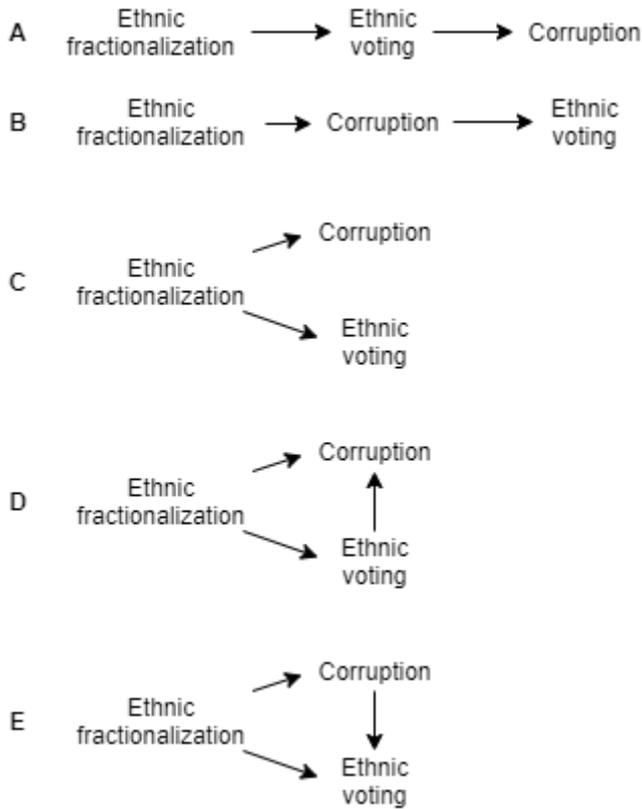
In this study, ethnic in-group favoritism will be operationalized as ethnic voting in parliamentary elections. The motivation for the usage is that the choice of a party that specifically champions one's ethnic group is a form of ethnic in-group favoritism, as it can both indicate the voter's intentions of favoring his or her group for office, or it can indicate how the politician or the political party favorites particular ethnicities.

The definition of ethnic voting will use the ethnic fractionalization within a party at the individual level, which makes it a classification based on which ethnic group a majority of the voters for a party belongs to, to use the phrasing by (Chandra, 2011). The definition of ethnic voting in each country will be: $(\sum_{i=1}^{ethnic\ groups} p_i \sum_{j=1}^{parties} q_{ij}^2) \cdot (\sum_{i=1}^{parties} r_i \sum_{j=1}^{ethnic\ groups} s_{ij}^2)$, where p_i stands for the fraction of ethnic group i to the entire population, q_{ij} stands for the size of party j in ethnic group i , r_i stands for the fraction of party r to the entire population, while s_{ij} means the size of ethnic group j in party i .

This method was originally used in a master's thesis by El Koubi (2016) who has also attached an example of how the calculation is done in practice in her appendix. The method is an interaction by two of Chandra's (2011) four methods of deciding the degree that a party is an ethnic party, namely a classification based on how a majority of an ethnic group votes multiplied by which ethnic group a majority of the voters for a party belongs to, for all parties and ethnic groups in a country. The motivation for this measure is, to begin with, a practical issue of the difficulty in deciding both the "de jure"-content of party manifestos from an entire continent, but it can also be advantageous use "de facto"-popularity in ethnic groups, as all ethnic voting patterns will not be conducted on formally ethnic parties. A continuous variable for the degree of ethnic voting for a party also has the advantage of bringing more possible nuance to the measures and results than a simple dummy variable approach.

The definitions of the other variables used in this study can be found in the appendix.

FIGURE 3 VISUALIZATION OF THE POSSIBLE CAUSAL RELATIONS INVOLVED



The mutual relationships between ethnic fractionalization, ethnic voting, and corruption

In figure 3, I list all possible ways in which ethnic fractionalization can be an independent variable that is consistent with that ethnic voting and corruption can be correlated. The modeled causal relations do not contain feedback loops, but feedback loops that do not feedback to ethnic fractionalization are simply combinations of two or more models in the figure.

Model A, which is perhaps the most employed in the earlier literature, will both be caught by the instrumental variable study and the individual study, as their purpose is to decide the causal direction involved. The same goes for model D, which contains the same direction of causality between ethnic fractionalization and ethnic voting. Model C should result in the instrument being correlated with the error term in the instrumental variable study, and should not result in a significant relationship in the individual study. Model E should lead us to find a significant correlation between ethnic fractionalization and ethnic voting, also when corruption is introduced in the model if the regression can measure such an effect.

The remaining model is model B, in which ethnic fractionalization leads people to favorite their co-ethnics in other aspects than through politics, which should, in turn, lead to corruption. Such a causal relationship should probably not be captured by the model. But consider that in many non-urban cases, ethnic groups in fractionalized countries live separated from each other. If ethnic fractionalization leads directly to corruption, without being mediated through politics, it would also need to work through ethnic in-group favoritism. But why should the local policeman or bureaucrat engage in corrupt practices by favoriting his or her co-ethnic locally, when everyone belongs to the same group? As an example, if I am a Muslim living in the Northern parts of Nigeria, with the largest majority of Muslims, I will likely not meet many Christians, and there would be little room for the police or bureaucrats to treat me better than they treat Christians. I would, however, statistically most likely vote for the APC, alike most other Muslim-dominated parts of Nigeria. It is hard to see how the local level corruption in somewhat ethnic homogeneous entities can be created by ethnic fractionalization in the national parliament.

The least complicated explanation for such a phenomenon should be that the country-wide ethnic fractionalization creates corruption at the state level, which then trickles back down to the local level and also affects local communities, which then leads to ethnic voting, but the explanation that ethnic fractionalization and corruption are unrelated has a higher degree of parsimony over such a mechanism.

Case selection and scope

The study will be conducted by a statistical evaluation of 31 African countries. Several reasons motivate the use of ethnic voting in Africa. To begin with, the variance is greater, both in terms of ethnic fractionalization and corruption, as compared with for example many European states. Second, the data from the Afrobarometer include better measures of ethnicity, but also the perception of corruption, as compared with other large surveys such as the European Values Survey, World Values Survey, Lapop, and other similar sources. It should be noted that the scope of this study is ethnic in-group favoritism in electoral democracies, via the study of ethnic voting. The study of ethnic in-group favoritism in countries without electoral democracy would require other methods

The spread of ethnic voting, corruption, and ethnic fractionalization is large in the population. Botswana, with a Bayesian Corruption Indicator score of 37, has a lower perception of corruption than for example South Korea and a comparable level with many EU countries, such as Poland, Lithuania,

Cyprus, and Malta. Guinea with a score of 69, ranks higher than countries such as Syria, Congo, and Iraq. The dataset also includes countries such as Uganda, where the largest ethnic group makes up only 16.5 percent of the population, with another eight ethnic groups in the span of 3-10 percent. However, the dataset also includes Lesotho, with almost no ethnic differences within the population.

Posner (2005) suggests that his findings regarding ethnic voting in Zambia should be possible to generalize to environments such as Los Angeles as well. I would be cautious when trying to generalize the results and aim at being consistent with how the ethnic groups are defined. The large variance in the observed variables does, however, indicate a degree of generalizability, at least within the spectrum we have in the data.

Analyses

The first analysis will study the correlation between corruption and ethnic voting, by the use of regression analysis on a panel dataset consisting of the included countries and years in the Afrobarometer dataset. As we can see in the appendix, the data has some issues with heteroskedasticity, meaning that a robust regression will be applied, rather than an OLS regression. Normally, a random-effects or linear-effects population-averaged would probably be applied for a short, unbalanced dataset such as this. However, ethnic fractionalization and ethnic polarization are constant over the period in the dataset, meaning that the variables will be omitted in a linear fixed-effects model, where the internal change in each country is measured, which is also one of the two components in a random-effects model. I will, therefore, study the between-effects exclusively, meaning the study of the difference between countries. The time-component will, therefore, be averaged out, and just the country cases will be compared in the regression.

Two key variables that will be introduced as control variables are ethnic fractionalization and ethnic polarization. Ethnic fractionalization is defined as the probability that two randomly selected people in a country will belong to different ethnic groups, i.e. $1 - \sum_{i=1}^{ethnic\ groups} \pi_i^2$, where π_i denotes the relative size of ethnic group i . A country with only one ethnic group would receive the value of 0, while a hypothetical case in which every citizen of a country belongs to different ethnicities, the fractionalization would be 1. Ethnic polarization is, on the other hand, a measure of the distance from the current distribution of relative sizes between ethnic groups in a country, to a completely bipolar distribution, which should represent the highest level of polarization. The formula for ethnic

polarization is, therefore $1 - \sum_{i=1}^{ethnic\ groups} \left(\frac{1-\pi_i}{2}\right)^2 \pi_i$, where π_i has the same meaning (Montalvo & Reynal-Querol, 2005).

GDP per capita is used as a control variable, as the level of economic development is likely to affect voting behavior. The size of the population is used as a control variable, as it is correlated with the perception of corruption, although the causality is debated (Knack & Azfar, 2003). The level of democracy is used to control for that autocratic regimes could either promote or prevent ethnic politization. Years of democracy is used as new democracies lack the party system stability of older democracies, which could affect the outcome in both directions. The level of education is used both as a proxy for human development and because it is not implausible that the level of education can promote tolerance. The age distribution in a country is one indicator of human development and can affect the stability of election results over time.

Apart from the regression, I will also conduct an instrumental variable study, to further gain knowledge regarding the causal direction of the relationship between corruption and ethnic fractionalization, an instrumental variable regression will be presented. This is done by replacing corruption in the model with a measure that is correlated with corruption without being correlated with ethnic fractionalization or the error term. If we find that the instrumental variable is correlated with ethnic voting as well, we will have an indication of that the causes of corruption that should be uncorrelated with ethnic fractionalization will also lead to more ethnic voting. If the instrumental variable is uncorrelated with the error term, it is also an indication against an omitted variable bias. In the first stage of the instrumental variable regression, the components that will constitute the instrumental variable are used in a linear regression to establish a linear model in which the new components will estimate corruption.

An essential general issue with all these components is that the model assumes that the instrumental variable is unrelated to unmeasured causes of the dependent variable (Sovey & Green, 2011). In this case, it is in particular essential to make sure that the instrument is not affected by the degree of ethnic fractionalization. Therefore, the assumption that ethnic fractionalization leads to corruption means that an instrument that is caused by corruption could be affected by the existence of ethnic fractionalization. I will therefore specifically look for instruments that cause corruption, and that are neither correlated with ethnic fractionalization nor with ethnic voting. Press freedom is used as I theorize that a free press has better preconditions of detecting corruption and demand responsibility.

Similarly, civil society interest groups will help to hold the political class accountable. Uneven economic development is chosen as it is a documented source of corruption (Uslaner, Svendsen, Svendsen, Svendsen, & Elgar, 2019).

In the appendix, there is a thorough checklist of the properties of the instrumental variables, as suggested by Sovey and Green (2011). In sum, the analysis of the instrument points towards that it is appropriate to use in this study. The exclusion criteria, meaning that there should be no risk of that the components of the instrumental variable affect the dependent variable other than through the instrumental variable, is likely met. This is confirmed by a value of the F-test above the rule-of-thumb of 10, meaning that the instrument is not likely to be correlated with the error terms, so the theoretical motivations of the construction of the instrument are not refuted.

Results

TABLE 1, ROBUST REGRESSION WITH ETHNIC VOTING AS DEPENDENT VARIABLE

Ethnic voting	1	2	3	4	5	6
Corruption (0-100)				.003	.005**	.005*
Ethnic Fractionalization (0-1)	-.005		.050	-.055		.009
Ethnic Polarization (0-1)		.113	.130		.214**	.217*
GDP/cap (log)	.008	-.001	-.004	.011	-.002	-.003
Population (log)	-.003	.003	.002	-.008	-.003	-.003
Democracy (-10 - 10)	.009	.009	.009	.011	.013	.013
Population aged <14 (log)	-.277	-.233	-.260	-.343*	-.291*	-.293*
Education level (log)	-.127	-.089	-.090	-.180*	-.091	-.088
Years of democracy (log)	-.028	-.025	-.025	-.026	-.021	-.021
R-squared	.217	.324	.333	.324	.556	.551
Adj, R-squared	-.021	.098	.067	.078	.379	.338
No, Observations	94	87	87	94	87	87
No, Groups	31	29	29	31	29	29

p<0.05, **p<0.01, ***p<0.001

TABLE 2, INSTRUMENTAL VARIABLE 2SLS REGRESSION, ROBUST TO HETEROSKEDASTICITY, WITH ETHNIC VOTING AS DEPENDENT VARIABLE AND AN INSTRUMENT FOR CORRUPTION

Model	1		2		3	
	1st stage	2nd stage	1st stage	2nd stage	1st stage	2nd stage
Ethnic voting						
Corruption (0-100)		.006		.010		.011*
Ethnic Fractionalization (0-1)	9.698	-.206			11.412	-.211
Ethnic Polarization (0-1)			-8.534	.280**	-7.348	-.314**
GDP/cap (log)	-1.394	-.012	-4.800*	-.049	-4.917	-.039
Population (log)	1.428	-.007	1.628	.004	1.128	-.314
Democracy (-10 - 10)	-.877	.009	-1.090	.018*	-1.129	.024*
Population aged <14 (log)	8.989	-.186	24.545*	-.386	25.156	-.321**
Education level (log)	-3.143	-.061	-.842	.004	.494	-.012
Years of democracy (log)	2.161	.000	3.553	-.020*	4.356	-.036
Press freedom (1-100)	.643***		.592***		.713***	
Civil society interest groups (0-10)	2.359**		3.092***		2.835***	
Uneven economic development (0-10)	-2.017		-2.253		-3.0230	
1st stage F-test	11.253		10.120		11.495	
Kleibergen-Paap	13.176		11.743		12.465	
	(p=0.004)		(p=0.008)		(p=0.006)	
Hansen J	1.224		3.289		3.875	
	(p=0.542)		(p=0.1931)		(p=0.1441)	
R-squared between		0.253		0.370		0.429
No. Observations		87		80		80
No. Groups		29		27		27

p<0.05, **p<0.01, ***p<0.001

Individual study

This part of the study will focus on the individual mechanisms of ethnic voting, by investigating whether individuals who perceive a high degree of corruption in society will also vote for ethnic parties to a greater extent. A multilevel, mixed-effects linear regression will be applied, to answer how

much the change in perception of corruption will affect the level of ethnic fractionalization in the party of choice. This analysis uses the Afrobarometer round 6, but just respondents that identify with parties and have stated ethnicity, which are 21711. The dependent variable, the ethnic fractionalization of each party in the survey has been calculated with the help of that the survey collects data on the respondent's ethnicity. If we know the relative size of every ethnic group among the voters for a party, we can also calculate the ethnic fractionalization of the voters for the party. The independent variable, a corruption index, has been calculated by combining answers to questions regarding the perception of corruption in various institutions. I will also control for political attitudes by introducing an index regarding the attitude to tax-funded services, that has been created in the same manner, as well as an index regarding how authoritarian the respondent is. This is a simplified measure of left-right and GAL-TAN-attitudes based on answers in the Afrobarometer. The country-level control variables have been retrieved from the Quality of Government dataset, and have the purpose of letting us better measure the specific individual effect. A full description of the variables can be found in the appendix.

Results

TABLE 3, MULTILEVEL, MIXED-EFFECTS LINEAR REGRESSION WITH ETHNIC FRACTIONALIZATION IN THE PARTY OF CHOICE AS DEPENDENT VARIABLE

Fixed effects	
Corruption perception (0-1)	-.040***
More state (0-1)	-.001
More authoritarian (0-1)	.015***
Random	
Democracy	2.80e-14
SE:	0.000
Country level corruption	3.22e-12
SE:	0.000
GDP/cap	1.10e-22
SE:	0.000
Level 1 cons	.740
SE:	.023
Level 2 cons	.016
SE:	0.000
Log-likelihood	15759.834
Number of obs	21711
Number of groups	30

*p<0.05, **p<0.01, ***p<0.001

The results in table 3 show a significant decrease in multi-ethnicity in the party choice if a voter perceives a high degree of corruption, i.e., that voters who perceive more corruption are more prone to ethnic voting. As ethnic fractionalization is a somewhat complicated measure, to begin with, as explained earlier in this chapter, the perceived outcome of the decrease in ethnic fractionalization will depend on how large the fractionalization is, to begin with. It is worth noting that the effect is not of an enormous magnitude. A move over the entire spectrum of perception of corruption leads just to a .04 decrease in the fractionalization of the political party of one's choice. The effect is nonetheless significant at the p<0.001 level.

The random effects, which measure the differences between each country and the population mean are also small, while the effects of an authoritarian attitude is a third in size compared to the effects

of perception of corruption. However, authoritarian values have the opposite effect as compared with the perception of corruption.

If we interpret the results, they point towards that people who perceive a higher degree of corruption will also vote for more ethnically homogeneous parties. This means that they politically favor their ethnic groups, unlike people who do not perceive such high levels of corruption, that vote for more ethnically diverse parties.

If we combine the findings at the country level and the individual level, we can establish that a correlation exists between corruption and ethnic in-group favoritism, measured as ethnic voting. We can establish that the causal direction is unlikely the reversed of what is proposed and that the relationship is unlikely to be caused by an omitted variable. We have also concluded that there exists an individual mechanism in which voters who perceive more corruption will also vote for ethnic parties to a greater extent.

Concluding discussion

The study finds that ethnic in-group favoritism is a result of corruption, rather than a result of just ethnic fractionalization. The findings challenge the previous literature theory, that has primarily suggested that ethnic fractionalization causes corruption through a mechanism of ethnic in-group favoritism, and instead points towards that institutions may affect the degree or characteristics of ethnic identification among people. The findings of this study are in line with Ahlerup and Olsson (2012) who suggest that the roots of contemporary ethnicities can be found in the competition for public goods and that state experience has a homogenizing influence on culture and ethnic identity. The findings are also not contradicted by Easterly (2001), who finds that institutional strength has an absorbing effect on ethnic conflict. Hroch (1993) has observed how the nationalistic idea grows stronger as the state is put into crisis, as people tend to over-value the protective comfort of their national group during such conditions.

If we assume that the many previous studies that have concluded a correlation between ethnic fractionalization and corruption are consistent with the mechanisms in the real world, there should be at least three plausible explanations for such a relationship. The first type of causal relationship, that ethnic fractionalization leads to corruption has already received most of the attention of this paper, and I have suggested that such a relationship is implausible based on the results of this study. The

second causal relationship, a spurious relationship in which a third variable causes both ethnic fractionalization and corruption can be plausible. My most likely candidate for such a variable would be related to the strength of state institutions, such as state capacity. The relationship between ethnic in-group favoritism and corruption in non-democracies such as USSR or SFR Yugoslavia, which were two countries with a relatively high degree of corruption but in comparison low levels of ethnic conflicts during the socialist rule as compared to later levels, are a weak spot in my theory that evolves much around corruption. The two countries did, however, have a fair share of state capacity despite corruption, which could be a contending explanation to corruption. A spurious relationship that does not revolve around state institutions should either be missing among the control variables, or its effect should be very strong but indirect on both variables, while it is also consistent with the 2SLS study and the individual study in this paper. Such a relationship is not impossible, but implausible.

The third explanation is reverse causality, in which corruption causes ethnic fractionalization. An implicit consequence of my theory is that the natural states that rely on culture or ethnicity still rely on personalization and patron-client networks. This means that they can likely not scale up without negative consequences. If they grow in size, they will eventually become too big for a personalized rule, leading to a more impersonal bureaucracy that should decrease corruption. Or they should divide into new natural states, that can either choose to cooperate or engage in conflict for the benefit of the elites in the state. If we would extrapolate this argument, an increase in corruption leads to a more natural state-like rule, which should fractionalize larger collective identities over time. Vice versa, North, Wallis, and Weingast (2009) conclude that open-access societies lead to more inclusive collective identities. Hobsbawm (2012) has also studied the effects of state institutions on ethnic characteristics and suggests that state institutions such as conscription will help to build a large enough base for a collective identity². Based on their theories, together with my application of them and the results from the analyses in this paper, a causal direction from corruption to ethnic fractionalization seems to be the most likely candidate among the discussed explanations of the correlation.

The roots of the natural state in the ability of an elite actor to uphold security and resolve conflicts bear much resemblance with the roots of capitalism in historical materialism. Marx (1867) suggests

² Only half the French people spoke at least some French at the time of the French revolution, and somewhere between 12-13 percent of the population spoke what could be considered the French language, but the share increased as institutions such as conscription were introduced.

that the mechanisms involved in the accumulation of capital are the driving forces of historical development. The first accumulation of capital is, however, the so-called primitive accumulation, in which one actor uses force to steal goods from others. I.e., the actor with the greatest violence potential at an early stage becomes the first capitalist and the first actor to create extractive institutions that depend on the surplus-value of other actor's labor. Olson (1993) has proposed similar roots of the state in the theory of how the roving bandit, who strikes different villages at different times becomes a stationary bandit after some time, i.e. someone who collects taxes in the same manner as present-day racketeering.

The connection between extractive institutions, the surplus-value addressed in Marx' Capital and ethnic fractionalization is perhaps not so far-fetched if we study the case of Papua New Guinea. As no plants or animals fit for domestication existed on the island before a migration of farmers that was a prerequisite for the neolith revolution in the country, large parts of the jungles that were unfit for farming were hunter-gatherer societies until recent centuries (Golson & Hughes, 1980)³. Papua New Guinea also happens to be the home of 856 known languages, or 12 percent of all the languages in the world, with just 0.11 percent of the world's population. Even if hunter-gatherer communities are also dependant upon the absence of violence, and even if tribal warfare has been observed in as primitive societies like those of gorillas, the results of violent clashes in pre-neolithic societies will never result in extractive institutions, as there is no surplus-value to extract from enemy tribes, due to the low productivity. The lack of surplus value will make the practices of the stationary bandit impossible, meaning that the pre-conditions for elites imposing their culture upon citizens will be absent. Combined with the findings of Ahlerup and Olsson (2012) regarding the roots of ethnicity in the competition for public goods, it seems as if the material conditions for a large ethnic fractionalization during pre-industrial times can have been relative overpopulation in relation to the production in the agriculture, provided that such practices existed.

The relevance of this study on the use of ethnic fractionalization in political science and related fields has already been discussed, but the study could have some applications on the outside world. The

³ Olsson and Hibbs Jr (2005) conclude that: *"These native New Guineans [are described as] "walking encyclopedias" with detailed knowledge of every imaginable use that could be made of hundreds of plants and animals. This profound knowledge of the natural environment, gained through thousands of years of observation, has also been recorded among other primitive peoples. [...] The notion that native New Guineans or Aborigines perhaps might have "missed" some crops or animals that could have been successfully domesticated therefore seems highly unlikely."*

extent of migration is perhaps currently greater than at any point earlier in history, which will inevitably lead to that different ethnicities live side by side more than before. From the basis of earlier studies and my contribution, an issue of societal relevance could receive different answers or explanations if studied closer: Will the migration, and the ethnic fractionalization that entails, lead to a greater degree of corruption or generally a lower human development in the receiving country? Based on the previous knowledge of ethnic fractionalization, the answer should probably be yes, but I would rather suggest no - it should rather depend upon the institutions in the receiving country. I hypothesize that ethnic identity politics could be more common among immigrant groups that have experienced more corruption in their home countries, than in the rest of the population, but that this difference should decrease over time, as the incentives to take part in a natural state will disappear when the surrounding open-access society provides better security and better opportunities.

In summary, this study includes both a theory and results that support an alternative explanation of the correlation between ethnic fractionalization and corruption. The connection between state institutions and ethnicity could provide to be an interesting field for further studies, which could reshape the understanding of both the origins of the state and the meaning of ethnic homogeneity or pluralism in society.

REFERENCES

- Ahlerup, Pelle and Ola Olsson (2012). “The roots of ethnic diversity”. In: *Journal of Economic Growth* 17.2, pp. 71–102.
- Alesina, Alberto, Reza Baqir, and William Russell Easterly (1999). “Public goods and ethnic divisions”. In: *The Quarterly Journal of Economics* 114.4, pp. 1243–1284.
- Alesina, Alberto and Eliana La Ferrara (2005). “Ethnic diversity and economic performance”. In: *Journal of economic literature* 43.3, pp. 762–800.
- Alesina, Alberto et al. (2003). “Fractionalization”. In: *Journal of Economic growth* 8.2, pp. 155–194.
- Borders, Reporters Without (2018). *Press freedom index. Version 2018*. eng. url: <https://rsf.org/en/ranking>.
- Bratton, W., W. Mattes, and E. Gyimah-Boadi (2015). *Afrobarometer (round 6)*. url: <http://www.afrobarometer.org/>.
- Cerqueti, Roy, Raffaella Coppier, and Gustavo Piga (2012). “Corruption, growth and ethnic fractionalization: a theoretical model”. In: *Journal of Economics* 106.2, pp. 153–181.
- Chandra, Kanchan (2011). “What is an ethnic party?” In: *Party Politics* 17.2, pp. 151–169.
- De Jong, Eelke and Christian Bogmans (2011). “Does corruption discourage international trade?” In: *European Journal of Political Economy* 27.2, pp. 385–398.
- Dincer, Oguzhan C. (2008). “Ethnic and religious diversity and corruption”. eng. In: *Economics Letters* 99.1, pp. 98–102. issn: 0165-1765.
- Easterly, William Russell (2001). “Can Institutions Resolve Ethnic Conflict? *”. eng. In: *Economic Development and Cultural Change* 49.4, pp. 687–706. issn: 00130079.
- El Koubi, Erin Nicole (2016). “Ethnic Political Parties and Civil Conflict”. In: *LSU Master's Theses* 1203. url: https://digitalcommons.lsu.edu/gradschool%5C_theses/1203/.
- Esarey, Justin and Leslie Schwindt-Bayer (2019). “Estimating Causal Relationships Between Women’s Representation in Government and Corruption”. In: *Comparative Political Studies*. doi:

10.1177/0010414019830744.

eprint:

<https://doi.org/10.1177/0010414019830744>. url: <https://doi.org/10.1177/0010414019830744>.

Foundation, Mo Ibrahim (2017). *The Ibrahim index of african governance*. eng. url: <http://%20mo.ibrahim.foundation/iiag/>.

Glaeser, Edward L. and Raven E. Saks (2006). “Corruption in America”. eng. In: *Journal of Public Economics* 90.6, pp. 1053–1072. issn: 0047-2727.

Golson, Jack and PJ Hughes (1980). “The appearance of plant and animal domestication in New Guinea”. In: *Journal de la Société des Océanistes* 36.69, pp. 294–303.

Gupta, Sanjeev, Hamid Davoodi, and Rosa Alonso-Terme (2002). “Does corruption affect income inequality and poverty?” eng. In: *Economics of Governance* 3.1, pp. 23–45. issn: 1435-6104.

Haken, N. et al. (2015). *Failed states index 2006-2018*. url: <http://fundforpeace.org/fsi/>.

Hobsbawm, Eric J (2012). *Nations and nationalism since 1780: Programme, myth, reality*. Cambridge university press.

Holmberg, Soren, Bo Rothstein, and Naghme Nasiritousi (2009). “Quality of Government: What You Get”. eng. In: *Annual Review of Political Science* 12, pp. 135–161. issn: 1094-2939. url: <http://search.proquest.com/docview/743014576/>.

Hroch, Miroslav (1993). “From national movement to the fully-formed nation: the nation building process in Europe”. In: *New Left Review* 198. issn: 0028-6060.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010). “The worldwide governance indicators: a summary of methodology, data and analytical issues”. eng. In: *World Bank Policy Research Working Paper* 5430.

Knack, Stephen and Omar Azfar (2003). “Trade intensity, country size and corruption”. eng. In: *Economics of Governance* 4.1, pp. 1–18. issn: 1435-6104.

La Porta, R et al. (1999). “The quality of government”. In: *Journal of Law, Economics, and Organization* 15.1, pp. 222–279. issn: 8756-6222.

Marshall, M. G., K. Jagers, and T. R. Gurr (2018). “Polity iv project, political regime characteristics and transitions, 1800-2017”. eng. In: *Center for Systemic Peace*.

Marx, Karl (1867). *Chapter Twenty-Six: The Secret of Primitive Accumulation*. Capital: Volume One.

Mauro, Paolo (1995). “Corruption and growth”. In: *The quarterly journal of economics* 110.3, pp. 681–712.

— (1998). “Corruption and the composition of government expenditure”. eng. In: *Journal of Public Economics* 69.2, pp. 263–279. issn: 0047-2727.

Montalvo, Jos G and Marta Reynal-Querol (2005). “Ethnic Polarization, Potential Conflict, and Civil Wars”. eng. In: *American Economic Review* 95.3, pp. 796–816. issn: 0002-8282.

North, Douglass C, John Joseph Wallis, and Barry R Weingast (2009). “Violence and the rise of open-access orders”. In: *Journal of democracy* 20.1, pp. 55–68.

Olson, Mancur (1993). “Dictatorship, democracy, and development”. eng. In: *The American Political Science Review* 87.3. issn: 00030554. url: <http://search.proquest.com/docview/214434459/>.

Olsson, Ola and Douglas A Hibbs Jr (2005). “Biogeography and long-run economic development”. In: *European Economic Review* 49.4, pp. 909–938.

Orjuela, Camilla (2014). “Corruption and identity politics in divided societies”. eng. In: *Third World Quarterly, 2014, Vol. 35, Iss. 5, pp. 753-769* 35.5, pp. 753–769.

Persson, Anna, Bo Rothstein, and Jan Teorell (2013). “Why Anticorruption Reforms Fail - Systemic Corruption as a Collective Action Problem”. eng. In: *Governance-An International Journal Of Policy Administration And Institutions, 2013, Vol. 26, Iss. 3, pp. 449-471* 26.3, pp. 449–471.

Posner, Daniel N. (2005). *Institutions and Ethnic Politics in Africa*. Cambridge: Cambridge University Press. isbn: 9780511808661.

Putnam, Robert D. (2007). “E Pluribus Unum : Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture”. In: *Scandinavian Political Studies* 30.2, pp. 137–174. issn: 0080-6757.

- Sovey, Allison J. and Donald P. Green (2011). "Instrumental Variables Estimation in Political Science: A Readers' Guide". eng. In: *American Journal of Political Science* 55.1, pp. 188–200. issn: 00925853. url: <http://search.proquest.com/docview/822736815/>.
- Standaert, Samuel (2015). "Divining the level of corruption: A Bayesian state-space approach". eng. In: *Journal of Comparative Economics* 43.3, pp. 782–803. issn: 0147-5967.
- Stiftung, Bertelsmann (2018). *Transformation index of the bertelsmann stiftung 2018*. url: <http://www.bti-project.org/en/index/overview>.
- Teorell, Jan et al. (2019). *The Quality of Government OECD Dataset*. eng. url: <http://www.qog.pol.gu.se>.
- Uslaner, Eric et al. (May 2019). "Corruption, inequality, and trust". In:
- Van den Berghe, P. L. (1987). *The ethnic phenomenon*. ABC-CLIO.
- Vanhanen, Tatu (1999). "Domestic ethnic conflict and ethnic nepotism: A comparative analysis". In: *Journal of Peace Research* 36.1, pp. 55–73.
- (2012a). "Ethnic Conflict and Violence in Heterogeneous Societies". eng. In: *The Journal of Social, Political, and Economic Studies* 37.1, pp. 38–66. issn: 0278839X. url: <http://search.proquest.com/docview/1009278501/>.
- (2012b). *Ethnic Conflicts*. Ulster Institute for Social Research.

APPENDIX

Country-level studies

TABLE 4, DESCRIPTIVE STATISTICS OF COUNTRY-LEVEL STUDIES

Name	Description	Obs.	Mean	Min	Max	St. dev	Source	Dataset name
Ethnic voting	A calculated score from the Afrobarometer. See El Koubi (2016) for methodology	95	.152	.029	.664	.103	Teorell et al. 2019. Original source: Standaert 2015	bci_bc
Corruption	Bayesian corruption indicator. A composite measure of the perception of corruption, made from over 20 surveys with 80 questions.	95	55.58	31.82	68.8	9.13	Teorell et al. 2019. Original source: Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg 2003.	bci_bc
Ethnic Polarization	Index of ethnic polarization as described in the method chapter	87	.521	.017	.897	.199	Montalvo and Reynal-Querol 2005	
Ethnic Fractionalization	Index of ethnic fractionalization, involving an index of racial and lingual characteristics.	94	.687	.255	.930	.179	Teorell et al. 2019. Original source: Kaufmann, Kraay, and Mastruzzi 2010	al_ethnic
Population aged <14 (log)	Total population between the ages 0 to 14 as a percentage of the total population.	95	3.712	2.940	3.91	.171	Teorell et al. 2019. Original source: Kaufmann, Kraay, and Mastruzzi 2010	wdi_pop14
Population (log)	Total population.	95	16.32	12.20	19.0	1.33	Teorell et al. 2019. Original source: Kaufmann, Kraay, and Mastruzzi 2010	log_pop
GDP/cap (log)	GDP per capita based on purchasing power parity.	95	7.899	6.475	9.95	.872	Teorell et al. 2019. Original source: Kaufmann, Kraay, and Mastruzzi 2010	log_gdp
Democracy	Polity measure ranging from -10 (strongly autocratic) to +10 (strongly democratic).	95	5.158	-4	10	3.58	Teorell et al. 2019. Original source: Marshall, Jaggers, and Gurr 2018	p_polity2
Education level (log)	One of the three sub-categories in the human development index. Consists of eight indicators from five different sources.	95	3.865	3.277	4.44	.265	Teorell et al. 2019. Original source: Foundation 2017	iiag_edu
Years of democracy (log)	Calculated uninterrupted years of democracy, defined as a value above 0 in the polity2-scale.	95	2.240	0	4.26	1.22	Teorell et al. 2019. Original source: Borders 2018	
Press freedom	Measures the press freedom in the country, and the efforts made by the government to see to that press freedom is respected.	94	25.01	1	54	10.3	Teorell et al. 2019. Original source: Borders 2018	rsf_pfi
Civil society interest groups	To what extent is there a network of cooperative associations or interest groups to mediate between society and the political system? Data available every second year 2005-2017, so the 2007 value is used for Afrobarometer round 3, and the 2015 value is used for Afrobarometer round 6.	88	5.614	2	10	1.57	Teorell et al. 2019. Original source: Stiftung 2018	bti_ig

Uneven economic development	Includes measures related to the GINI coefficient, the income share of the top and bottom 10%, slum population, etc.	92	7.674	3.5	9.7	.983	Teorell et al. 2019. Original source: Haken et al. 2015	ffp_used
------------------------------------	--	----	-------	-----	-----	------	---	----------

Individual study

Below are to begin with the definitions of the variables used in the study. Then follows descriptive statistics of the variables. All data is taken from Afrobarometer round 6 unless specified otherwise Bratton, Mattes, and Gyimah-Boadi 2015.

Partifrak: ethnic fractionalization in the party of choice, as calculated from the Afrobarometer data. See chapter 3 for a closer description of the measure.

Corruption perception: index about perception of corruption composed by answers to the following questions in Afrobarometer:

Q53A: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: The President* and Officials in his Office? *or prime minister, depending on the most powerful leadership role

Q53B: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Members of Parliament?

Q53C: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Government Officials?

Q53D: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Local government councilors?

Q53E: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Police?

Q53F: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Tax Officials (e.g. Ministry of Finance officials or Local Government tax collectors)

Q53G: How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: Judges and Magistrates?

More state: index about the economic relation between the individual and the state, composed by answers to the following questions:

Q26E: For each of the following actions, please tell me whether you think it is something a good citizen in a democracy should always do, never do, or do only if they choose: Pay taxes they owe to government 1=Never do 2=Do only if they choose 3= Always do

Q44: Which of the following statements is closest to your view? Choose Statement 1 or Statement 2. Statement 1: Citizens must pay their taxes to the government in order for our country to develop. Statement 2: The government can find enough resources for development from other sources without having to tax the people. 1=Agree very strongly with Statement 1, 2=Agree with Statement 1, 3=Agree with Statement 2, 4=Agree very strongly with Statement 2, 5=Agree with neither

Q65C: If the government decided to make people pay more taxes or user fees to increase spending on public health care, would you support this decision or oppose it? 1=Strongly oppose, 2=Somewhat oppose, 3=Neither support nor oppose, 4=Somewhat support, 5=Strongly support, 6=It depends (e.g., on size of the increase)

More authoritarian: composed by the following variables:

Q16: Which of the following statements is closest to your view? Choose Statement 1 or Statement 2. Statement 1: Government should be able to ban any organization that goes against its policies. Statement 2: We should be able to join any organization, whether or not the government approves of it. 1=Agree very strongly with Statement 1, 2=Agree with Statement 1, 3=Agree with Statement 2, 4=Agree very strongly with Statement 2, 5=Agree with neither

Q30: Which of these three statements is closest to your own opinion? Statement 1: Democracy is preferable to any other kind of government. Statement 2: In some circumstances, a non-democratic government can be preferable. Statement 3: For someone like me, it doesn't matter what kind of government we have. 1=Statement 3: Doesn't matter, 2=Statement 2: Sometimes non-democratic preferable, 3=Statement 1: Democracy preferable

Q33: Which of the following statements is closest to your view? Choose Statement 1 or Statement 2.
 Statement 1: Political parties create division and confusion; it is therefore unnecessary to have many political parties in [ENTER COUNTRY]. Statement 2: Many political parties are needed to make sure that [ENTER NATIONALITY] have real choices in who governs them. 1=Agree very strongly with Statement 1, 2=Agree with Statement 1, 3=Agree with Statement 2, 4=Agree very strongly with Statement 2, 5=Agree with neither

Level of democracy: Polity measure ranging from -10 (strongly autocratic) to +10 (strongly democratic). `p_polity2` in Teorell et al. (2019). Original source: Marshall, Jaggers, and Gurr 2018

GDP per capita (log): GDP per capita based on purchasing power parity. Log-transform applied, because of skewness of distribution. `wdi_gdppppcur` in Teorell et al. (2019). Original source: Kaufmann, Kraay, and Mastruzzi 2010

Control of Corruption: country-level measure of perception of corruption. `wbgi_cce` in Teorell et al. (2019). Original source: Kaufmann, Kraay, and Mastruzzi 2010

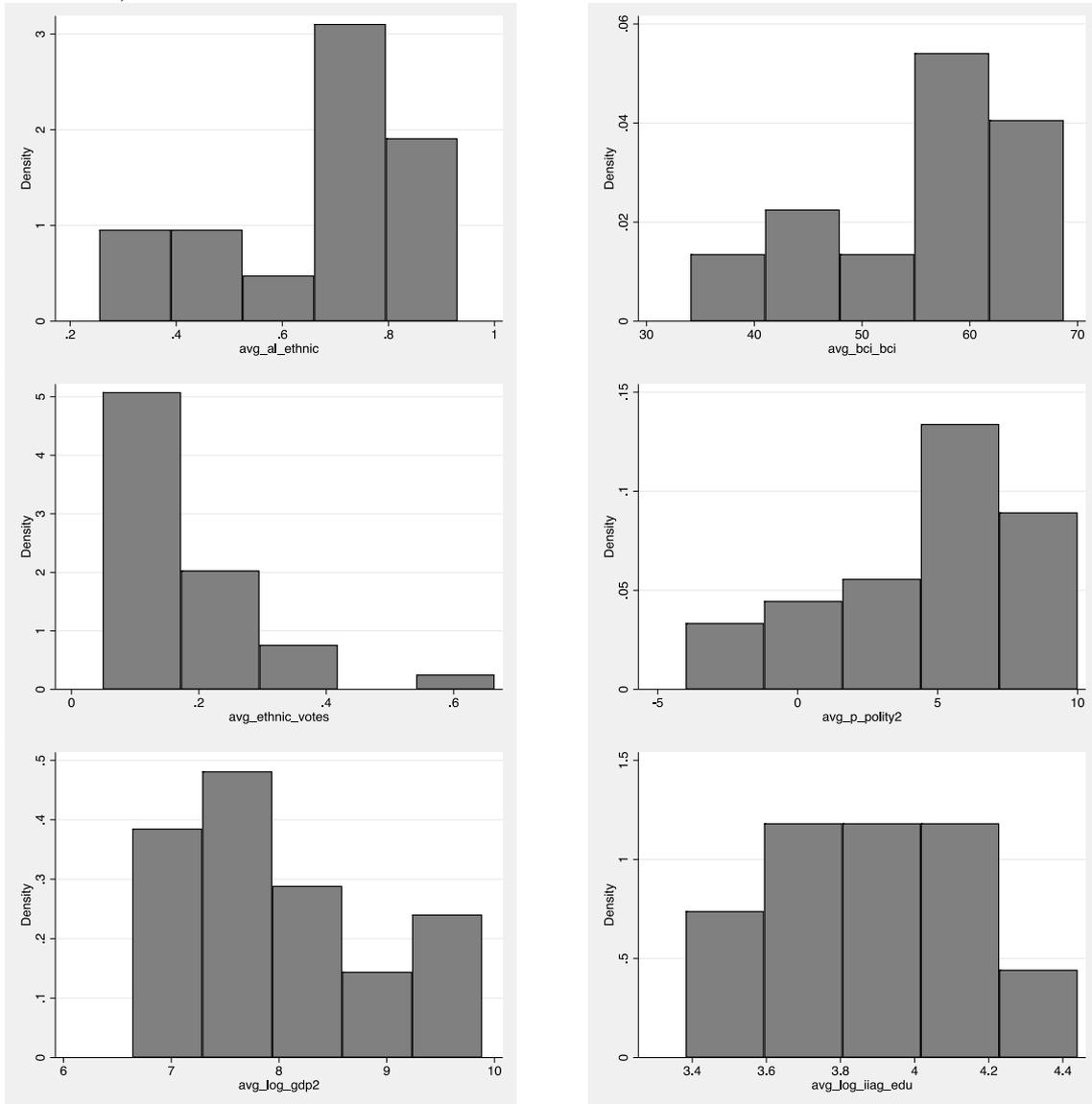
In table 5 we can see the descriptive statistics of the variables

TABLE 5, DESCRIPTIVE STATISTICS OF INDIVIDUAL-LEVEL ANALYSIS

Variable name	num	mean	min	max	sd
Party fractionalization	32,472	.7203501	0	.9489706	.2140973
Perception of corruption	24,829	.4783423	0	1	.2267099
More state	31,928	.3507315	0	1	.1019617
More authoritarian	28,643	.3060754	0	1	.2395673
Democracy	32,472	5.141907	-4	10	3.513579
Corruption	31,28	-.5723053	-1.30416	.9495435	.5039418
GDP/cap (log)	32,472	6689.129	796.9608	60028.93	10606.51

HISTOGRAMS

FIGURE 4, HISTOGRAMS OF THE USED VARIABLES IN THE COUNTRY-LEVEL REGRESSION



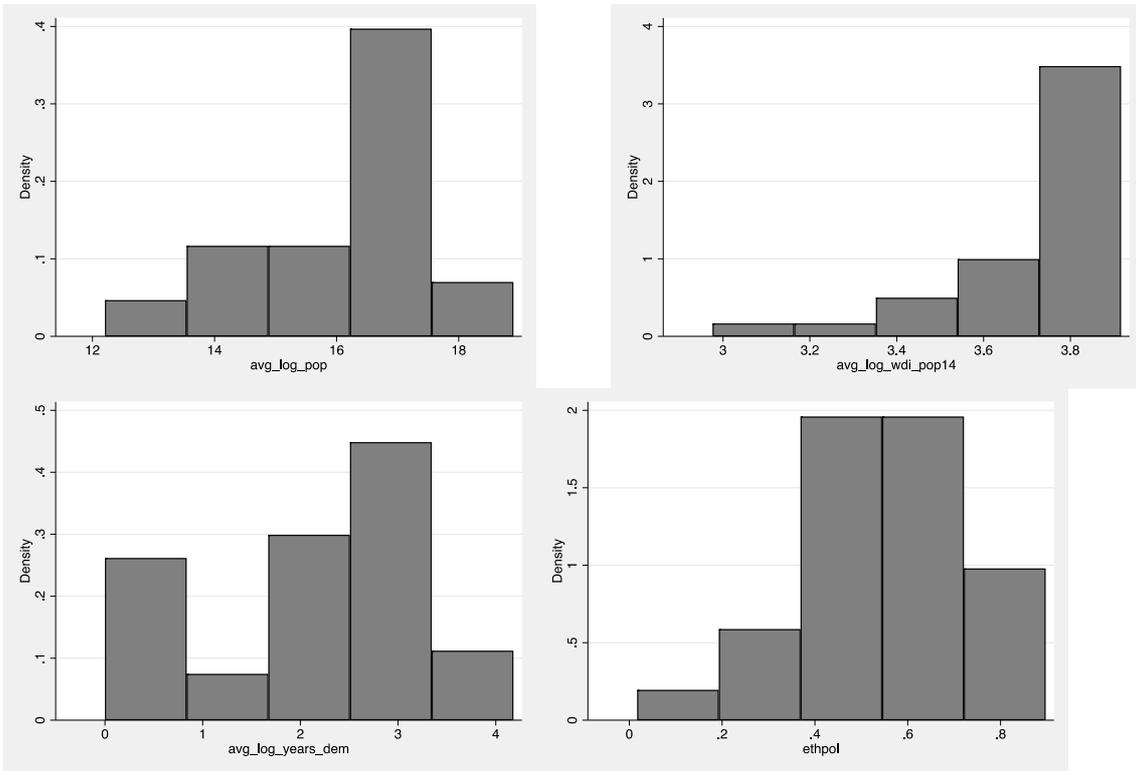
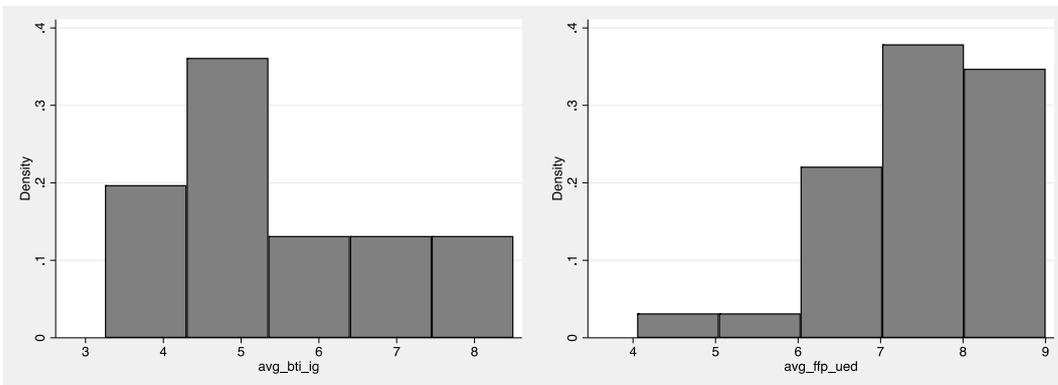


FIGURE 5: HISTOGRAMS OF THE USED VARIABLES IN THE INSTRUMENT VARIABLE STUDY



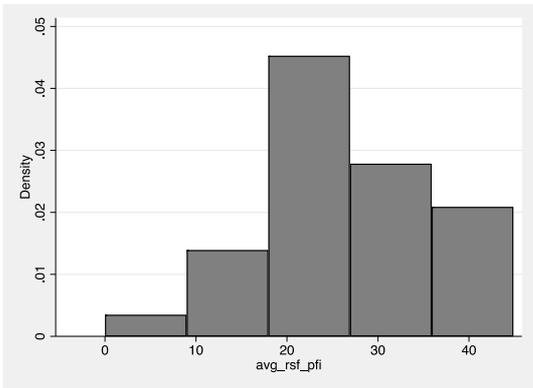
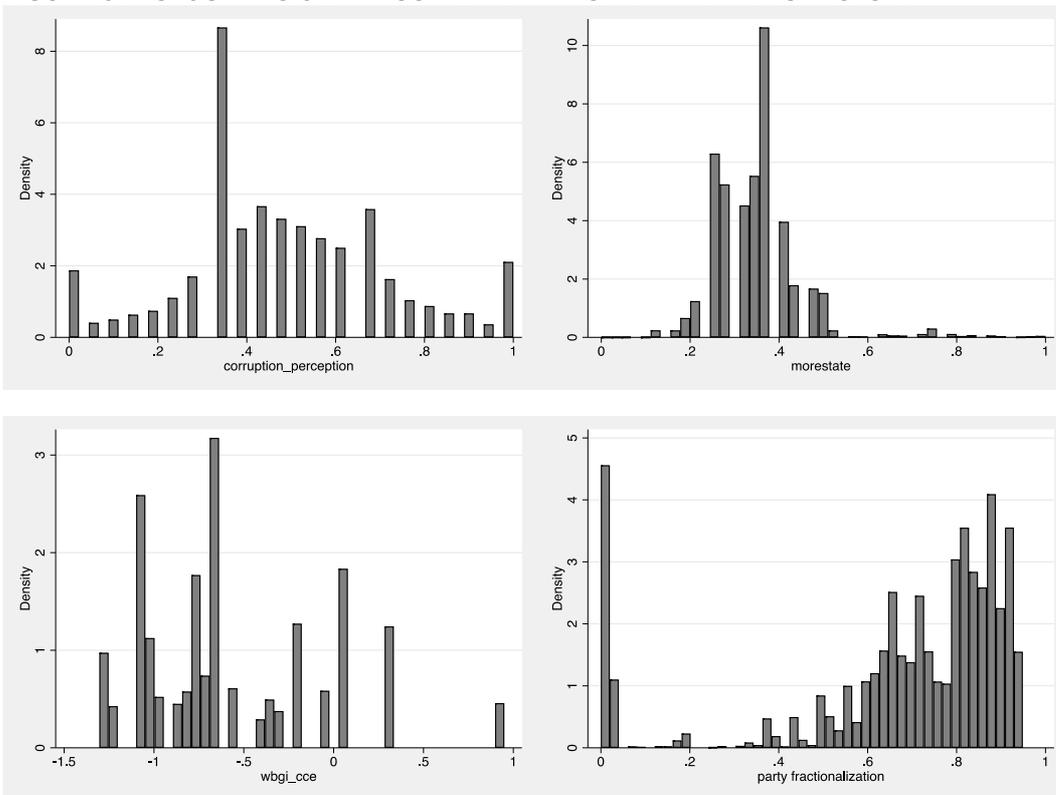
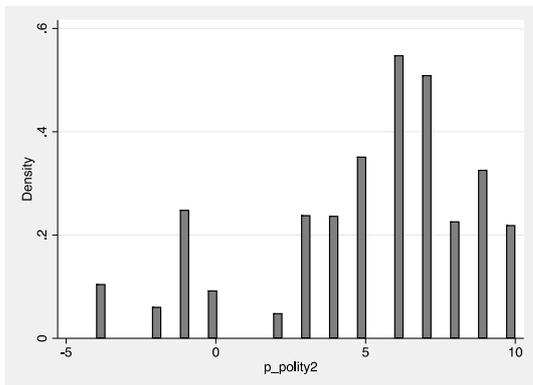


FIGURE 6: HISTOGRAMS OF THE USED VARIABLES IN THE INDIVIDUAL STUDY





REGRESSION DIAGNOSTICS

Multicollinearity

In table 6, we can observe that the level of democracy and the years of democracy are above the rule-of-thumb value of 5, which is likely as the years of democracy is built based on the continuity of the democracy variable. After the removal of years of democracy, we observe no values above 5, as seen in table 7. In table 8, we can observe the results of the regression when the variable years of democracy is omitted. The significance generally increases, and the effect becomes slightly stronger.

TABLE 6, MULTICOLLINEARITY SCORES

Variable	VIF	1/VIF
Democracy	7.88	0.126868
Years of democracy (log)	7.62	0.131229
Population aged <14 (log)	4.71	0.212234
GDP/cap (log)	3.83	0.260910
Education Level (log)	2.94	0.339569
Corruption	2.26	0.443317
Population (log)	1.63	0.612461
Ethnic fractionalization	1.62	0.617845
Ethnic polarization	1.47	0.679140

Mean VIF 3.77

TABLE 7, MULTICOLLINEARITY SCORES AFTER REMOVAL OF YEARS OF DEMOCRACY

Variable	VIF	1/VIF
Population aged <14 (log)	4.67	0.214109
GDP/cap (log)	3.48	0.287638
Education Level (log)	2.92	0.342560
Corruption	2.23	0.449107
Ethnic fractionalization	1.59	0.627504

Population (log)	1.55	0.643733
Ethnic polarization	1.44	0.693241
Democracy	1.31	0.762704
Mean VIF	2.40	

TABLE 8, ROBUST REGRESSION WITHOUT YEARS OF DEMOCRACY

	4	5	6
Ethnic voting			
Corruption (0-100)	.005*	.006**	.006**
Ethnic Fractionalization (0-1)	-0.050		-0.005
Ethnic Polarization (0-1)		.213**	.212*
GDP/cap (log)	0.003	-0.008	-0.007
Population (log)	-0.010	-0.005	-0.005
Democracy (-10 - 10)	0.006	.008*	.008*
Population aged <14 (log)	-.399*	-.296*	-.295*
Education level (log)	-.225**	-0.091	-0.106
R-squared	0.435	0.569	0.567
Adj. R-squared	0.263	0.426	0.394
No. Observations	94	87	87
No. Groups	31	29	29

Heteroskedasticity during OLS regression

In table 9 follows the regression before correcting for heteroskedasticity by use of robust regression. Table 10 and 11 indicates that the data has problems with heteroskedasticity, which leads to the use of robust regression.

TABLE 9, OLS REGRESSION WITH ETHNIC VOTING AS DEPENDENT VARIABLE

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ethnic voting						
Corruption (0-100)				0.008	.009*	.010*
Ethnic Fractionalization (0-1)	-0.172		-0.152	-0.209		-0.183
Ethnic Polarization (0-1)		.161	.136		0.253	0.226
GDP/cap (log)	-.025	-.049	-.037	-0.008	-0.033	-0.018
Population (log)	.004	.010	.016	-0.010	-0.003	0.003
Democracy (-10 - 10)	.001	.006	.008	0.007	0.015	0.018
Population aged <14 (log)	-.128	-.219	-.158	-0.224	-0.336	-0.267
Education level (log)	-.091	-.059	-.086	-0.087	-0.034	-0.065
Years of democracy (log)	.015	.002	-.004	0.007	-0.012	-0.021
R-squared	.125	.151	.185	0.265	0.357	0.405
Adj. R-squared	-.141	-.131	-.141	-0.003	0.100	0.124
No. Observations	94	87	87	94	87	87
No. Groups	31	29	29	31	29	29

TABLE 10: CAMERON & TRIVEDI'S DECOMPOSITION OF IM-TEST

Heteroskedasticity	29.00	28	0.4125
Skewness	8.00	9	0.5341
Kurtosis	4.32	1	0.0377
Total	41.32	38	0.3277

TABLE 11: BREUSCH-PAGAN / COOK-WEISBERG TEST FOR HETEROSKEDASTICITY

chi2(1)	=	5.57
Prob >chi2	=	0.0183

ETHNIC VOTING SCORES

TABLE 12, CALCULATED ETHNIC VOTING SCORES FROM AFROBAROMETER 3-6

Round				
Country	3	4	5	6
Algeria				0.1743349
Benin	0.1341196	0.1577043	0.160517	0.172181
Botswana	0.0518673	0.0535502	0.0520464	0.0451351
Burkina Faso		0.2361726	0.1623485	0.1213898
Burundi			0.6642144	
Cameroon			0.0945853	0.0795054
Cape Verde	0.2319027	0.1787482		0.108141
Cote d'Ivoire			0.5151948	0.2578199
Gabon				0.1460714
Ghana	0.2187126	0.2081557	0.2350929	0.2393313
Guinea			0.3597646	0.3657575
Kenya	0.1084234	0.1954245	0.1492057	0.2112517
Lesotho	0.0999087	0.0671755	0.0451729	0.0406236
Liberia		0.0602237	0.0710765	0.0397666
Madagascar	0.0998412	0.1110948	0.0738843	0.080418
Malawi	0.128453	0.1561377	0.1151922	0.1369964
Mali	0.0322842	0.0415885	0.0536527	0.0681547
Mauritius			0.3188097	0.2700564
Morocco				0.1585895
Mozambique	0.2036528	0.156498	0.1201208	0.1711715
Namibia	0.2777414	0.2922999	0.2903441	0.2713821
Niger			0.1595229	0.2193528
Nigeria	0.1329415	0.0938054	0.1310653	0.143357
Sao Tome and Principe				0.3327368
Senegal	0.1903325	0.139003	0.1161529	0.1339343
Sierra Leone			0.2800359	0.27438
South Africa	0.136182	0.1080908	0.1369537	0.1079517
Tanzania	0.1127852	0.0623846	0.0287226	0.0407626
Togo			0.1142364	0.0938591
Togo				
Uganda	0.093046	0.0671997	0.0589902	0.0842157
Zambia	0.1015217	0.0982467	0.1070491	0.1108115
Zimbabwe		0.1335044	0.0812928	0.109229

INSTRUMENTAL VARIABLE CHECKLIST

Below is a checklist for appropriate use of instrumental variables in political science, created by Sovey and Green (2011), that will be applied on the choice of instrumental variables in this study.

What is the estimand? The estimand is perception of corruption, which is not an exact measure of the “true” extent of corruption. Press freedom and civil society interest groups could have a gate-keeping function against corruption, if corrupt practices have a larger chance of being detected.

It has been rather hard to compare the composition of this instrumental variable to that of others, just because most accessible studies that use corruption as an instrumental variable use ethnic fractionalization as an instrument (Gupta, Davoodi, & Alonso-Terme, 2002; Mauro, 1998; De Jong & Bogmans, 2011; Esarey & Schwindt-Bayer, 2019).

Is the instrumental variable independent of the potential outcomes? There could be reason to question the homogeneity of the causal effect. Press freedom and the extent of civil society interest group could all be related to practices of grand corruption. However, news media, social media or action groups would probably have little interest in the average persons wrongdoings, i.e. petty corruption. Also Uslaner (2008) suggests that, although petty corruption might be correlated with inequality, inequality is a much better suited explanation for grand corruption.

This might however relate to another fundamental issue of the research design, namely that a more specific measure of grand corruption could be better suited as a means to explain the causal link between perception of corruption and ethnic voting.

Explain why it is plausible to believe that the instrumental variable is unrelated to unmeasured causes of the dependent variable.

The two issues I have aimed to address here are both the case of a reverse causality, i.e. that ethnic voting causes corruption. The second issue I have aimed to avoid are that the independent variables could be systematically related to unobserved causes of the dependent variable. I have identified the

greatest risk here to be if variables related to ethnic divisions, such as ethnic fractionalization or ethnic polarization, are both causing ethnic voting and corruption.

For both reasons, I have regarded variables caused by corruption as unfitting, as they hypothetically could be affected by a corruption that is in turn affected by either ethnic voting or e.g. ethnic fractionalization.

Press freedom has a potential weakness, which is that countries with a higher degree of corruption could also be more prone to censorship and state-owned journalism. A decrease of press freedom in a country could also be a reaction to journalistic investigations of corrupt political elites, that can, in turn, be theorized to be the consequence of ethnic fractionalization. However, if we consider that many European countries have had a relatively high degree of press freedom for two hundred years, it seems as if press freedom could be somewhat resilient to changes in levels of corruption.

The relationship between corruption and civil society interest groups could be that a decrease in corruption is correlated with other measures, such as GDP per capita or welfare, that could be preconditions for civic participation in interest groups. Another explanation could be that corrupt countries are more autocratic, and therefore persecute interest groups to a higher degree. However, the measure is not of the freedom of interest groups, but of the mere existence of them. The emergence of e.g. the labor movement in Europe during the 19th century also points towards that a high degree of spare time and money are not necessary preconditions for the formation of civil society interest groups.

For the last measure, uneven economic development, a plausible other explanation could be that ethnic fractionalization leads to corruption, that leads to uneven economic development. Here, I would, however, suggest a historical materialist approach, i.e., the development of productive forces and the production relations are the main determinants of the uneven distribution of growth. These materialist causes are in turn blind to ethnicities.

Suppose an instrumental variable is deemed exogenous because it is random or near random. Are the exclusion restrictions valid? It is plausible to believe that the instrumental variable could have direct effect on the outcome, but most plausible explanations would rather point towards that the instrumental variable should increase the value of the dependant variable, and such an effect has not been measured.

A plausible path from press freedom to ethnic voting is a greater extent of the ethnic press, that could lead to ethnic, political mobilization. In this case, press freedom should be positively correlated with ethnic voting, but the measures are very uncorrelated. A possible explanation of this is that press freedom could lead to a higher degree of journalistic professionalization, as the pressure to be regime-friendly should decrease. This should, in turn, mean more impartial news coverage.

A plausible path from civil society interest groups to ethnic voting would be if a higher degree of civil society would lead to a greater extent of ethnic interest groups, that could increase ethnic political mobilization. If this were the case, these measures would be positively correlated, but there is almost no correlation at all between the two in the used dataset.

Are the instruments weak? As we will see in the results section, the F-value is slightly above 10, where instrumental variables with an F-value could be considered as weak as a rule of thumb Sovey and Green 2011. Also, the Hansen J statistic and Kleibergen-Paap LM statistic are within the suggested bounds.

Does the instrumental variable have a monotonic effect on the treatment? Press freedom has a theoretical maximum. It is plausible that the effect of increased press freedom will decrease with higher values of the variable, if we hypothesize information about corruption as a form of disease, where newspaper readers are vectors that infect their surroundings with relevant information.

It is also plausible that both inequality and civil society interest groups have threshold effects, where e.g. some inequality might be seen as justified, which can obstruct forms of corruption, or that a handful civil society interest groups are manageable for a corrupt regime.

Are the observations subject to spillover effects? I do not exclude that the instruments can be subject to spillover effects, but these will probably be more subtle and indirect.