



GÖTEBORGS  
UNIVERSITET

DEPARTMENT OF POLITICAL SCIENCE

# THE LINK BETWEEN ETHNIC FRACTIONALIZATION AND CORRUPTION REVISED

Ethnic voting in Africa

**Håkan Bernhardsson**

---

Master's Thesis:	30 higher education credits
Programme:	Master's Programme in Political Science
Date:	2019-05-27
Supervisor:	Carl Dahlström
Words:	13616



## **Abstract**

This thesis revisits the relationship between ethnic fractionalization and corruption in Africa. An earlier literature argues that ethnic fractionalization leads to corruption via mechanisms involving ethnic favoritism. In this study, an alternative theory suggests that the casual 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 – an OLS 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

# Contents

<b>List of Figures</b>	<b>2</b>
<b>List of Tables</b>	<b>3</b>
<b>1 Introduction</b>	<b>5</b>
<b>2 Previous research</b>	<b>8</b>
<b>3 Theory</b>	<b>15</b>
<b>4 Research strategy</b>	<b>19</b>
4.1 Design . . . . .	19
4.2 Data . . . . .	20
4.3 Case selection/scope . . . . .	20
4.4 Ethnicity, ethnic voting, fractionalization and polarization . .	21
<b>5 Analyses</b>	<b>23</b>
5.1 Country-level correlation between corruption and ethnic voting	23
5.1.1 Results . . . . .	24
5.2 Instrument variable regression on country-level . . . . .	26
5.2.1 Results . . . . .	27
5.3 Individual perception of corruption and ethnic voting . . . . .	29
5.3.1 Results . . . . .	30
5.4 Summary . . . . .	31
<b>6 Concluding discussion</b>	<b>32</b>
<b>References</b>	<b>36</b>
<b>A Appendices</b>	<b>I</b>
A.1 Descriptive statistics and sources . . . . .	II
A.1.1 Country-level studies . . . . .	II
A.1.2 Individual level study . . . . .	IV
A.2 Histograms . . . . .	VII

A.3	Regression diagnostics . . . . .	IX
	A.3.1 Multicollinearity . . . . .	IX
	A.3.2 Heteroskedasticity . . . . .	XI
A.4	Robustness checks . . . . .	XII
	A.4.1 OLS regression . . . . .	XII
	A.4.2 Instrumental variable regression . . . . .	XV
A.5	Ethnic voting scores . . . . .	XVII
A.6	Instrumental variable checklist . . . . .	XIX

# List of Figures

3.1	Visual representation of the proposed causal model. . . . .	18
5.1	Added variable plot for corruption . . . . .	26
A.1	Histograms of the used variables in the country-level regression	VIII
A.2	Histograms of the used variables in the instrument variable . .	IX
A.3	Histograms of the used variables in the individual study . . . .	X

# List of Tables

5.1	Regression with ethnic voting as dependent variable . . . . .	25
5.2	Instrumental variable 2SLS regression with ethnic voting as dependent variable and an instrument for corruption . . . . .	28
5.3	Multilevel, mixed-effects linear regression of with ethnic fractionalization in the party of choice as dependent variable . . . . .	30
A.1	Descriptive statistics of country-level analyses . . . . .	II
A.2	Descriptive statistics of individual-level analysis . . . . .	VI
A.3	Multicollinearity scores . . . . .	X
A.4	Multicollinearity scores after removal of years of democracy . . . . .	XI
A.5	Cameron & Trivedi's decomposition of IM-test . . . . .	XI
A.6	Breusch-Pagan / Cook-Weisberg test for heteroskedasticity . . . . .	XI
A.7	Robustness check after removal of observations with Cook's $D > (4/29)$ . . . . .	XII
A.8	Robustness check after removal of observations with Cook's $D > (4/87)$ . . . . .	XIII
A.9	Robustness check after correction for heteroskedasticity . . . . .	XIV
A.10	Instrumental robustness check after correction for heteroskedasticity, with new instrument . . . . .	XVI
A.11	Calculated ethnic voting scores from Afrobarometer 3-6 . . . . .	XVIII

# 1 | Introduction

The causal relationship between ethnic fractionalization and corruption in contemporary political science generally points to that the ethnic fractionalization cause outcomes such as lower welfare provisions, less economic growth and a higher degree of corruption (Easterly, 1997; Mauro, 1998; La Porta, Lopez-de Silanes, Shleifer, & Vishny, 1999; Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg, 2003). I.e., ethnic institutions are still relevant, and the state generally functions better in ethnic homogenous countries. Multi-ethnic countries should, therefore, have a higher degree of corruption and lower degree of both welfare provisions and economic growth.

Most available evidence points towards a correlation between ethnic fractionalization and corruption. However, previous scholars have devoted little attention to explain why the causal direction leads from ethnic fractionalization to corruption, or the possibility of a spurious relationship between ethnic fractionalization and corruption. It is plausible that a pre-existing level of corruption could lead people to embrace ethnic institutions above dysfunctional state institutions, where ethnic institutions can either be ethnicity as a social institution or more formal organizations, such as tribal councils, religious congregations or recurrent sit-downs between elders. This thesis will investigate the hypothesis of a causal direction from corruption to ethnic politicization, where ethnic politicization means that ethnic differences constitute dividing lines in political competition. Such a causal direction implies that ethnic fractionalization should not lead to corruption via politics, which would therefore contradict a large part of the conclusions by Alessina et al.

A prerequisite for ethnic politicization should be the prevalence of strong ethnic identities. Earlier works, such as Hobsbawm (2012) have proposed that ethnic identities are not necessarily static, but rather constructed and reconstructed by state institutions and politics. Posner (2005) has studied how the battle for power has led to the political articulation of either language or tribe, whichever has the most substantial chance of winning the majority. Laclau and Mouffe (2001) have suggested that articulation within the political discourse is the primary tool in the creation of political hegemony in which the primary conflict lines, the contending groups and collective identities are defined. Therefore, it is not unlikely that politics



can have an enhancing effect on ethnic identification, a process that could be particularly potent when state institutions do not function impartially and when elections become a competition for scarce public goods.

The extent of ethnic voting differs between multi-ethnic countries. Two countries such as Guinea and Zambia, with similar degrees of ethnic fractionalization, have very different levels of ethnic voting, according to Afrobarometer data (Bratton, Mattes, & Gyimah-Boadi, 2015). The variance in ethnic voting between multi-ethnic countries implies that ethnic fractionalization is not the only important factor in explaining the politicization of ethnicity, measured as ethnic voting.

There are several possible motivations for why corruption can be a contributing factor to ethnic voting. Orjuela (2014) has proposed that corruption can exacerbate grievances along ethnic lines, as corruption has just the property of depending on particularized trust, meaning a positive difference between in-group and out-group trust. Corruption should decrease the levels of generalized trust (Rothstein, 2013), which combined with a state that is unable to enforce contracts between people creates a need for reliance on other institutions, where ethnic institutions should be a useful choice, in order to, for example, conduct market transactions (Alesina & Ferrara, 2005).

In summary, there is reason to believe that corruption, rather than ethnic fractionalization per se, causes ethnic politicization, which also would imply that ethnic fractionalization should not lead to corruption via politics. The research question that follows from the stated problem is: what is the causal direction in the relationship between corruption and ethnic fractionalization, and how can the causal relationship be theoretically motivated?

This thesis will conduct three analyses with the aim to investigate the correlation, the causal direction and the individual mechanisms involved in the relationship between corruption and ethnic voting. The correlation will be investigated via an OLS regression. The causal direction will be investigated by means of an instrumental variable regression, where an instrument for corruption will be constructed by measures of press freedom, the involvement of civil society in the political process, uneven economic development and internet access. The involved individual mechanism will be investigated by means of a multilevel mixed-effects linear regression. The analyses will find support for the suggested correlation, causal direction and individual mechanism involved.

The paper proceeds as follows: in section 2, I will summarize the current research on ethnic voting and the interaction between ethnicity and state institutions. In section 3, I will propose a theoretical explanation of the relationship between corruption and ethnic voting. In section 4, I will present a research strategy to approach the issue at hand. In section 5, I will test my theory with the help of three analyses based on Afrobarometer data. In section 6, I will discuss the implications of the results, with a concluding

discussion.

## 2 | 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).

The theorized effects of ethnic fractionalization are for example a higher degree of corruption and less social services as well (Alesina et al., 2003). 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 et al., 2003).

If ethnicity is seen as a constant characteristic, it is intuitive that the direction of causality is that ethnic fractionalization is the root cause of corruption, given the correlation between the variables. However, Ahlerup and Olsson (2012) propose that the roots of ethnicities are found in the competition for resources, which could imply that ethnic structures change as material factors evolve. Hobsbawm (2012) studies the state's involvement in the formation of nationalities, showing other examples of how national identities are reshaped over time. 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.

Just as state institutions and politics can contribute to the formation of nationalities and articulation of ethnic favoritism, Hroch (1993) suggests that, in the absence of functioning state institutions, one will have to rely on other institutions. Under acute stress, this will make people to "*over-value the protective comfort of their own national group.*"

Some of the most cited papers that suggest a causal direction in which ethnic fractionalization per se leads to corruption are Mauro (1998), Easterly

(1997), La Porta et al. (1999), Alesina et al. (2003), and Posner (2004), which are papers that lack thorough and explicit theory about the mechanisms of ethnic fractionalization. Here follows the walkthrough of some papers that motivate a theory behind why ethnic fractionalization can affect both corruption and other related institutions, 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 first is related to the aggregation of individual preferences of attributing positive utility to the well-being of one's group. I.e., people, in general, tend to prefer others from their ethnicity, which leads to differences at a systemic level. The second explanation 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. The third explanation relates to that the cost of production increases with ethnic diversity as a result of difficulties in communication over lingual or cultural lines.

The logic behind the first reason about prioritizing one's group could be explained by Milgram (1970), who argues that the overload of interactions between people in modern societies means that one will prioritize social responsibility to those who share many of the same characteristics as oneself. Putnam (2007) carries on this idea, theorizing that social distance is a function of one's social identity, which can be deconstructed and constructed again. In the absence of other forms of social identities, ethnic categories could be the primary source for social identities, leading to a higher social distance between ethnicities rather than between other categories. 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 contradictive empirical findings, as the study also shows that levels of trust will decrease within ethnic groups in ethnically diverse communities (Putnam, 2007). Putnam finds that ethnic diversity leads to a lower level of generalized trust, not only between races in the USA but also within the races of the respondents. Moreover, ethnocentric trust, meaning in-group trust within one's race minus out-group trust towards other races, is uncorrelated with ethnic fractionalization, which indicates that ethnic fractionalization per se does not lead to higher friction between ethnic groups, but that it instead leads to generally low levels of social trust.

Social trust is negatively correlated with corruption, but Rothstein (2013)

argues for a causality where corruption leads to lower social trust. When public officials - "*people whom the law requires to act in the service of the public*" - can not be trusted to follow the law, why would people, in general, trust others to uphold the social contract? If we assume that corruption unidirectionally causes a lower degree of social trust rather than the other way around, then we can overall rule out the causality: ethnic fractionalization  $\Rightarrow$  lower social trust  $\Rightarrow$  corruption. It could instead be so that Putnam has missed a spurious relationship in which corruption both leads to the low levels of trust and the ethnic fractionalization in the case of the USA.

Uslaner (2013) does propose a form of feedback between trust and corruption, and this relationship is a central component in his inequality trap, where inequality  $\Rightarrow$  low social trust  $\Rightarrow$  corruption  $\Rightarrow$  inequality. However, Uslaner is specific in that corruption at least requires one form of trust, an "honor among thieves", i.e., in-group trust. It seems unlikely that ethnic fractionalization contributes to the inequality trap, as ethnic fractionalization is uncorrelated with ethnocentric trust (Putnam, 2007).

Economic inequality per se seems to be a cause of clientelism, as suggested by Stokes (2011). (Orjuela, 2014) also suggests that ethnic differences can provide a moral justification for corruption and clientelism and an argument for ethnic politicization, in the sense that "*it is our turn to eat*".

It should be hard to generalize empirical results on the link between ethnic fractionalization and other measures from studies in the USA. In the USA, ethnic fractionalization is not only a measure of fractionalization per se but also a proxy of the distribution between the white majority population and the Asian, Black and Hispanic minority populations (as well as other categories). However, we know that the USA is still a racially hierarchical country which is manifested in that the median income in Asian households is 119 percent of the income of white households, while Hispanic households earn 74 percent and black households earn 59 percent of the income of white households (U.S. Bureau of Labor Statistics, 2018).

The results of the effect of ethnic fractionalization in the studies of, for example Alesina et al. (1999), that suggest that ethnic fractionalization leads to lower public goods provisions, or Putnam (2007), that suggest that ethnic fractionalization leads to lower social trust, are consistent after controlling for socioeconomic factors, such as income and education. However, their studies could still miss an effect in which corruption is an underlying variable, meaning that corruption could lead to both a higher degree of ethnic fractionalization in the affected regions and a lower extent of social trust. At least two possible explanations of this spurious relationship are likely in the case of the USA. If communities that have a higher degree of public goods provisions to schools and welfare services are attractive to live in, this could form a selection bias in which Blacks and Hispanics are referred to communities with lower public goods provisions and more corruption. Another possible explanation is that the same factors that lead to that Blacks and

Hispanics earn less than Whites and Asians (a possible explanation being structural racism), will also lead to a lower degree of public goods provisions and a smaller likelihood of combatting corruption, as the public might not be as interested in the welfare of the minority groups.

It is also worth to note that differences that are used as explanations for corruption in other countries, such as religion, are not studied so thoroughly in the case of the USA. The social distance between races in the USA or even the distance between Catholics and Lutherans in the USA should theoretically be larger than the distance between clans in Somalia or tribes in many other African countries if we look upon ethnic practices. However, while the smaller ethnic differences are thoroughly studied in the African case and suggested as an explanation for conflict and corruption, they are largely ignored in the USA.

The second explanation by Alesina and Ferrara (2005), that ethnicities affect market strategies by affecting the information bias in market transactions, has an interesting element, being that ethnicities play a particularly important role when the legal power of the state is weak. However, this implies that state institutions play an essential role in explaining ethnic conflicts, which points to a causal direction from weak state institutions to corruption.

The third explanation by Alesina and Ferrara (2005), that ethnic differences lead to higher transaction costs due to communication problems, is plausible in cases of differences in language but loses much of its explanatory power when applied to other types of differences, such as race or religion, as they should technically not affect communication so much.

Glaeser and Saks (2006) but concludes a direction of causality from ethnic fractionalization via politics to corruption: *"If an area is torn apart by ethnic divisions and leaders tend to allocate resources towards backers of their own ethnicity, then members of one ethnic group might continue to support a leader of their ethnic group, even if he is known to be corrupt."* Their proposed logic is that when corruption is introduced, to begin with, it will persist in ethnically fractionalized societies, as voters will not be interested in removing those in charge of the corrupt situation. They do, however later, therefore, examine the effect of ethnic heterogeneity overall on corruption in the USA, using linear regression. However, their theorized reason why ethnic fractionalization would lead to corruption will at best imply that ethnic fractionalization makes corruption worse in already corrupt communities because of ethnic voting, meaning that a linear model based on the theory has problems with internal validity. Moreover, they do theorize that ethnic voting can be a result of corruption or at least clientelism, even though they do not investigate that causal direction.

Dincer (2008) refers to the anthropologist Van den Berghe (1987) ethny concept, as a justification of in-group favoritism within ethnicities. Berghe developed a sociologic theory in which social networks of extended family

and relatives constitute a group which is a target for in-group favoritism. However, van den Berghe himself suggests that this type of social network consists of a few hundred members, and would be flawed when applied to ethnicities that amount to millions of members.

Another exception to the absence of theoretical motivations is Cerqueti, Coppier, and Piga (2012), who create a model explaining the relationship. In sum, they propose a principal-agent environment consisting of entrepreneurs, bureaucrats, and controllers. Entrepreneurs rely on bureaucrats in order 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 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 in relation to systemic corruption makes the model flawed.

As Dinesen and Sønderskov (2015) conclude, even though the empirical relationship between ethnic diversity and social trust is widely explored, the theoretical explanation of the causal direction seldom explored in-depth. Apart from the three suggestions by Alesina and Ferrara (2005), what other explanations of the effects of ethnic fractionalization seem credible?

An “evolutionary-biological” approach has been proposed by several authors before. People will want their genes to survive, which promotes a form of ethnic altruism, according to (Rushton, 2005). The same ethnic altruism, a genetic need for once genes to reach superiority, can explain ethnic conflicts and even genocides, according to the theory. Vanhanen (2012a) writes that “*we can trace the roots of ethnic conflict and violence to human nature*”, and defines what he calls ethnic nepotism, which he argues is an evolutionary drive for favoring one’s ethnicity. He backs his argument by pointing to a significant relationship between ethnic fractionalization and the degree of ethnic interest conflicts in a country. However, it should be evident that at least a small extent of ethnic fractionalization is needed in order for ethnic interest conflicts to be possible at all, and the distance from this insight to the proposed theory is vast. One criticism of the evolutionary-

biological approach is Van den Berghe (1987), who 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 is almost impossible to decide ethnicity based on appearance. Another argument against the evolutionary-biological approach can perhaps be found in the social constructivist discussion about race and racism, which is too long to be investigated as a whole, but where many authors conclude that racial properties are constructed by society rather than derived from human nature (Smedley & Smedley, 2005). Moreover, Vanhanen (2012a) does not respond to the many exceptions to ethnic nepotism, meaning all developed multi-ethnic countries with a deficient degree of ethnic conflict.

Vanhanen (2012b) theorizes a reason behind why ethnic conflicts do not always result from ethnic fractionalization, as "*The authoritarian communist regime in multiethnic Yugoslavia had been strong enough to prevent eruptions of ethnic violence.*" Historical examples, such as the Saddam Hussein regime in Iraq point towards that authoritarianism can even be combined with genocide.

Posner (2005) constructs the issue of ethnic voting in a public choice framework, where the political actors structure the rules of political competition so that citizens embrace the most advantageous social identity. When the institutional frameworks, such as the electoral system, changes, the political conflict lines have shifted in Zambia from tribe to language and back to tribe. Posner also argues that ethnicity is currently salient in Zambian politics "precisely because of the widespread expectation that people in positions of power will favor members of their ethnic groups", which implies that ethnic voting will follow from the expectation of clientelism. As a consequence of his proposed relationship between ethnic voting and clientelism, if the possibility of clientelism is removed, the incentives of ethnic voting should be decreased.

Another option remains: that corruption both leads to both a higher degree of ethnic fractionalization and lower to social trust. This relationship is, of course, impossible if we regard ethnicities as static or even almost static. Some evidence suggest that ethnicity might well change over time, such as Ahlerup and Olsson (2012), who propose that the roots of ethnic diversity can be found in the competition for public goods. Modern state experience, i.e., mechanisms for the distribution of public goods will have a reducing effect on ethnic diversity.

Contemporary national self-identification can also change over time. In, for example, Montenegro, the share that identifies as Montenegrins have fallen from 91 percent in the 1948 census to 45 percent in the 2011 census, and the share that identifies as Yugoslavs had fallen from 5 percent in 1981 to 0 percent in 2011 (Montstat, 2011). The identification in Taiwan points to similar results, where the share of people who consider themselves Chinese



instead of Taiwanese have been falling over time (Chen, 2017). The general trends in the world are not as drastic, and states will probably not reshape ethnic or national identities in any dramatic ways. More interesting than studying ethnic fractionalization per se is perhaps to study the politicization of ethnic identity. Easterly (2001) points towards how ethnic conflict can be seen as a function of both ethnic diversity and institutional strength, where institutional strength will have an absorbing effect on ethnic conflict. Violent ethnic conflict is perhaps the ultimate politicization of ethnic identity. However, this politicization also occurs in peaceful democracies in the shape of ethnic voting.

In summary, the literature that has been reviewed is characterized by a lack of theory behind why ethnic fractionalization should lead to corruption. It is not evident in the previous research why individuals should prefer others from their own ethnicity and why this preference must be aggregated and politicized. Many studies suggest a correlation between ethnic fractionalization and lower levels of social trust. However, the lower levels of trust apply both in-group and out-group, which makes an increase in corruption along ethnic lines implausible via that mechanism. We have also studied studies in favour for that corruption lead to both low levels of trust and inequality, and studies that conclude that inequality is a good base for clientelist practices. These theories will be part of the underpinning of the theoretic framework in this thesis, as explained in the next chapter.

## 3 | Theory

I will begin by targeting three assumptions behind the idea that ethnic fractionalization leads to corruption.

The first assumption is that ethnicities are something pre-determined, like an artefact from the Tower of Babel or a static property given by nature. This means that ethnicities are somewhat independent of state institutions. The historical roots of contemporary nationalities are however, often preceded by the formation of state institutions, as Hobsbawm (2012) argues. Hroch (1993) also explains the rise of nationalism after the breakdown of Communist rule, with that familiar ties crumble, leading to a general anxiety and insecurity *“in which the national idea takes over the role of collective integration In conditions of acute stress, people characteristically tend to over-value the protective comfort of their own national group”*. In summary, states seem to create new national identities, and the absence of states can lead to that old identities take over the role of upholding security from the states.

If we continue this thought, corruption is a form of antithesis to state power, as states rely on the maintenance of rules through force. Corruption is then only possible when rules cannot be enforced properly. Therefore, I hypothesize that strong state power, such as Hobsbawm (2012) described for example France after the revolution, will construct new ethnic identities. Meanwhile, ethnic fractionalization and ethnic conflict should be a result of a long-term lack of state capacity. Moreover, in the absence of present-day state capacity, the ancient institutions – ethnicities – take its place.

The second assumption is that people will favor co-ethnics over members of other ethnic communities. However, is it as relevant in modern societies, and in that case, how would it manifest? We can imagine a society where all forms of state institutions and judiciary treat people completely impartial, and where the economic institutions are non-extractive and promote fair competition. In such a society, ethnic favoritism would both be absent in the public sector and a non-competitive behavior in market competition. Therefore, there would be no room to treat ethnicities different in a way that would impact their workplace situation, economic status or their rights before the law. There would be no reason for ethnic politicization in those areas.

The distribution of capital between ethnicities could still be an argument for ethnic politicization in this hypothetical society, but an as broad political appeal as possible, reaching across ethnic borders, should be the preferred strategy for electoral victories, which should be an argument for the party as a rational actor to treat the distribution of capital as a non-ethnic issue.

One possible explanation to why people prefer people from their own ethnicity, that has been mentioned in the previous research, is genetic or ethnic altruism, meaning an evolutionary-biology rational explanation to ethnic favoritism. However, genetic diversity seems to have a positive impact on the population's possibility to adapt to changing environments, meaning that genetic altruism should not be evolutionarily favorable (Lankau & Strauss, 2007). Countless other arguments could be made for why people will favor co-ethnics, such as the issue of ethnicity as a cultural trait, which could be approached from a social-constructivist perspective. In conclusion, the thesis put forward here is just that the idea of ethnic favoritism should not be considered an absolute truth.

The third assumption is that ethnic favoritism in politics is a result of processes related to non-material interest. Here, I would instead suggest rather that the explanation for ethnic voting relates to the competition with other ethnicities for public goods. The aggregation of materialist interest into ethnic favoritism can have both an instrumental and a structural explanation. In the structural variant, the public choice school comes to hand, in which voters are considered both rational and policy-oriented (Dewan & Shepsle, 2011), and will consequently vote for whatever option that gains them the most, which should in most cases be ethnic favoritism in the short term.

In my theory, I would instead suggest the following explanation. I will disregard a too literal interpretation the theories of the primitive accumulation by Marx (1867) or in the Olson (1993) theory of the roving bandit, in which the roots of the state can be found in an anarchy of power differences, where the actor with the highest violence potential will be able to rule over others. I instead choose a more Hobbesian approach, as in that people prefer the absence of violence and theft, and that the materialist explanations should instead refer to that the capital owner or bandit that succeeds in creating a system of order will also achieve legitimacy from the people under their rule. Such a rule must be able to resolve conflicts over, for example property in order to prevent violence. In the most primitive community, these types of conflicts would typically be resolved by the head of the family, or whoever has the most considerable legitimacy, for example via the classic "sit-down" between gangsters in *The Sopranos*, where Tony Soprano has autocratic power over how to resolve conflicts within his branch of the Mafia. In larger societies, such as the ethny, conflicts would need to be resolved in tribal councils, by elders, by religious leaders, or by the one with the most massive capital of violence. However, when we reach above

the ethny, conflicts will need to be resolved through state institutions.

When state institutions fail to function, one will have to rely on more primitive institutions in order to achieve security and absence of violence and theft, as suggested by Hroch (1993). In failed states, ethnic institutions might provide order, but only in-group, as the power of an ethnic or religious leader will not apply to people of other ethnicities. The limits of ethnic institutions will create distrust towards other groups. If we also assume that countries with a higher degree of corruption will have scarce welfare provisions and public goods, and the competition for the existing goods could lead to ethnic voting.

In countries with impartial governments, the importance of ethnicity in politics should be low, as the impact of elections in most cases will not affect the direct output of government. In countries with a high degree of corruption and clientelism, the dominant ethnicity of the party in government should play a much more significant role, as they can provide more goods towards people of their ethnicity.

Parties that are partial towards a particular ethnicity will act according to an ethnicity-based form of clientelism, as in *the proffering of material goods in return for electoral support* where the criterion of distribution that the patron uses is: did you (will you) support me? (Stokes, 2011). However, there is little need for an explicit exchange of material goods in connection to elections, as there will be an implicit expectation from the voters that the ethnic party will work to favor them.

Stokes (2011) also suggests that economic inequality per se seems to be a cause of clientelism. If we assume rational actors, the interests of the ruling elites to employ clientelism should be independent of ethnic demographics, but one could expect a harder case of moral justification of clientelism in ethnic homogenous contexts. Add to the equation that ethnic differences can provide a moral justification for corruption, an argument for ethnic politicization, in the sense that *“it is our turn to eat”* (Orjuela, 2014).

Countries with a high degree of corruption will also have a large gap between party policies and the actual implementation of the policies, which leads to that political parties will have little credibility on actual policy matters. As a result, voters will prefer parties with a history of patronage transactions, over parties that aim to provide public goods via universalistic policies, as they offer a greater assurance of benefit (Chandra, 2007).

Another important difference from classic clientelism and a contributing factor behind why corruption could lead to ethnic voting is the fear of being disfavored by other ethnic groups, in case the other ethnic groups take over the government. The fear of repression from other ethnic groups could lead to that also rational voters that do not notably support an ethnic-nationalist cause might vote for ethnic parties, as they otherwise would risk losing a favorable treatment to other ethnic groups.

One further argument for that ethnicity are a good target for particu-

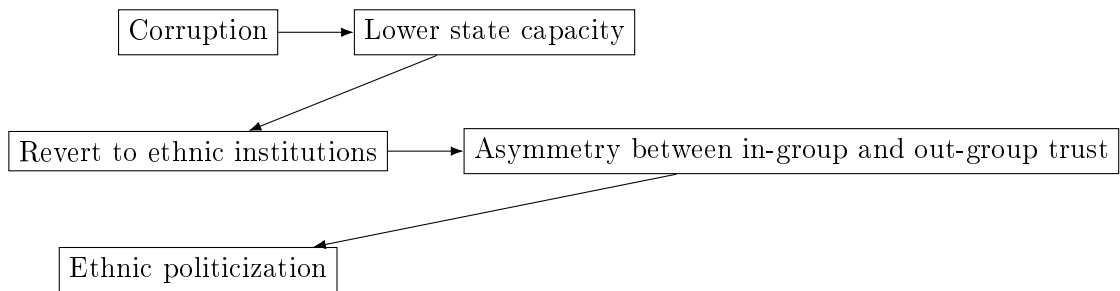


Figure 3.1: Visual representation of the proposed causal model.

laristic policies is that ethnicity is usually more static than, for example, class or the target for agrarian issues. People can change class over their lifetime or move from rural to urban areas, and many class-based parties in functioning democracies strive to reshape their electoral base by the help of universalistic policies. Meanwhile, ethnic parties work to preserve ethnic differences. Therefore, ethnicity could be a more natural category to accentuate politically.

To summarize my theory, I propose that state power will provide means of security from violence and in the long run, even reconstruct ethnicity along state borders. I also propose that the absence of state power, as manifested by the extent of corruption, will force people to rely on more primitive institutions in order to resolve conflicts – and in multi-ethnic countries, this will typically be ethnic institutions. As ethnic institutions can only resolve conflicts within one’s own ethnicity, the need for them will create an asymmetry between in-group and out-group trust. If articulated, this asymmetry will be a good basis for clientelist approaches by political entrepreneurs, meaning that ethnicity is politicized as a base for ethnic identity politics. Thus, corruption leads to ethnic voting. Figure 3.1 is a representation of the proposed causal relations.

Following from the theory, I propose the following hypothesis:

$H_1$ : The extent of corruption in a country leads to ethnic voting

## 4 | Research strategy

### 4.1 Design

This quantitative study will be conducted in three steps. First, I will use OLS regression to investigate whether a correlation exists between corruption and the politicization of ethnicity, measured as the extent of ethnic voting. This evaluation is a necessary test of the hypothesis, as a correlation indicates a relationship and a dependence between variables, direct or indirect.

A correlation does however not imply causation. An establishment of that correlation is, not proof per se for either the suggested hypothesis or theory. Does corruption lead to ethnic voting, as the hypothesis suggests, or is the causal direction reversed? Or have we observed a spurious relationship where ethnic fractionalization both leads to corruption and ethnic voting? To address this issue, I use an instrumental variable approach in order to isolate the unidirectional effects of corruption on ethnic voting. The method will work in two stages, where the purpose of the first stage is to use variables unrelated to ethnic voting and ethnic fractionalization to predict the level of corruption in the country. Next, the instrument is used instead of corruption in a regression, together with the other variables. The purpose of the instrumental variable analysis is to establish the causal direction between corruption and ethnic voting. By constructing a study that excludes the possible effects of ethnic fractionalization or omitted variables on ethnic voting, we can study the effect of corruption per se, and draw more accurate conclusions regarding the mechanisms of the relationship, including the direction of causality.

Last, I will study the relationship on an individual level, to bring insight to my proposed mechanisms, by studying how an individual's perception of corruption affects its extent of ethnic voting. In the theory section, I suggested a mechanism in which an individual's perception of corruption affects its probability of voting for an ethnic party. Therefore, different results on the individual and aggregated level will mean that the proposed theory is inconclusive.

Anyone of the analyses will not be as meaningful alone, as the three analyses rely on each other to test the theory. A correlation between ethnic

fractionalization and ethnic voting does not stand in conflict with reverse causality. The use of an instrumental variable indicates a causal direction but does not tell us much about the proposed theoretical mechanisms. The study on the individual level could also be consistent with various possible spurious relationships between corruption and ethnic voting. All three analyses combined will however tell a more credible story of the involved mechanisms.

## 4.2 Data

The primary sources of data in this study are the Afrobarometer and the Quality of Government dataset (Bratton et al., 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 Appendix A.1.

## 4.3 Case selection/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 politicization in electoral democracies, via the study of ethnic voting. The study of ethnic politicization in countries without electoral democracy would require other methods.

This thesis will include two levels of analysis. Two studies will aggregate individual data to conduct country-level comparisons between the countries, while one study will be conducted on an individual level.

The after sought case should be a pathway case according to the definition by Gerring (2007), which means a case in which more than one path can lead to a particular outcome. In this case, either ethnic fractionalization or corruption can be seen as paths to ethnic voting, and we will want to take ethnic fractionalization into account when we study the effect of corruption.

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 3-10 percent. However, the dataset also includes Lesotho, where the only ethnic differences are within the Sotho subgroup, that make up 99,7 percent of 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 have criticized that results from the USA are generalizable already in section 2. Many of the countries included in this study have a very high degree of ethnic fractionalization, as compared with, for example, European countries. It is therefore plausible that there exists, for example, a threshold needed for ethnic voting to occur at all. The large variance in the observed variables does, however, indicate a high degree of generalizability, at least within the spectrum we have in the data.

#### 4.4 Ethnicity, ethnic voting, fractionalization and polarization

The definition of ethnicity that is used throughout this thesis is the one by Horowitz (1985), that *"Ethnicity is based on a myth of collective ancestry, which usually carries with it traits believed to be innate. Some notion of ascription, however diluted, and affinity deriving from it are inseparable from the concept of ethnicity"*, as his definition both condenses many of the most used definition prior to his, and that his work has become one of the most cited since.

The definition of an ethnic party is not at all evident. Parties with ethnic manifestos could hypothetically either pursuit non-ethnic issues or attract voters from other ethnic groups. Parties without ethnic manifestos can pursuit policy positions that are favorable to one ethnicity.

Chandra (2007) defines an ethnic party as *"a party that represents itself to voters as the champion of the interests of one ethnic category or set of categories to the exclusion of another or others, and makes such a representation central to its strategy of mobilizing voters."*

Chandra (2011) proposes four different classifications of ethnic parties:

1. A classification based on the name.
2. A classification based on the political platform.
3. A classification based on how a majority of an ethnic group votes.
4. A classification based on which ethnic group a majority of the voters for a party belongs to.

Using explicit definitions, such as name or political platform, will avoid false positives but might produce false negatives. Using statistics of voters, ethnicities will avoid false negatives but might produce false positives.



The country-level studies in this thesis will employ a mix between Chandra's classification 3 and 4 and will be derived by processing ethnicities and party choices from the Afrobarometer for each included country. The ethnic voting score is derived by interacting the probability that two people who vote for the same party also belong to the same ethnic group with the probability that two people who belong to the same ethnic group also vote for the same ethnic party. It is, therefore, a mix between classification 3 and 4, as proposed by (Chandra, 2011). 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 the entire continent, but also to decide the "de facto"-popularity in ethnic groups. A continuous variable for the degree of ethnic voting for a party also has the advantage of bringing more possible nuance to the measure than a simple dummy variable approach. The formula is:

$$\left(\sum_{i=1}^{ethnicgroups} p_i \sum_{j=1}^{parties} q_{ij}\right) \cdot \left(\sum_{i=1}^{parties} r_i \sum_{j=1}^{ethnicgroups} s_{ij}\right)$$

where  $p_i$  stands for a 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 a fraction of party  $r$  to the entire population, while  $s_{ij}$  means the size of ethnic group  $j$  in party  $i$ . This model was initially proposed in a master's thesis by El Koubi (2016), who also includes a more detailed description of how the measure is calculated. I have however calculated the degree of ethnic voting in round 3-6 of Afrobarometer as well, instead of just round 5.

Alesina et al. (2003) suggest that ethnic fractionalization can be measured as the probability that two random people in a population will belong to different ethnic groups. I.e.,  $1 - \sum_{i=1}^{ethnicgroups} p_i^2$ , where  $p_i$  stands for a fraction of the group  $i$  to the entire population.

An alternative measure of the distribution of ethnicities in a population can be measured through the ethnic polarization index, as proposed by Montalvo and Reynal-Querol (2005). It measures how close a distribution is to an entirely bipolar set  $(\frac{1}{2}, 0, \dots, 0, \frac{1}{2})$ . The formula looks as follows:  $4 \sum_{i=1}^{ethnicgroups} p_i^2 (1 - p_i)$ , where  $p_i$  stands for a fraction of the ethnic group to the entire population. The ethnic polarization index has a stronger explanatory power than ethnic fractionalization for conflicts in a country (Montalvo & Reynal-Querol, 2005). Both measures combined provide more information about the distribution of ethnic groups in a country than each measure alone.

## 5 | Analyses

### 5.1 Country-level correlation between corruption and ethnic voting

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. 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 the just the country cases will be compared in the regression. An advantage of the use of ethnic fractionalization and ethnic polarization as constant over the time period is that changes in the ethnic demographics with an almost constant population could be correlated with ethnic politicization, such as in the cases with Montenegro or Taiwan, as described in section 2. A disadvantage is that the model will not capture the effects of migration on the ethnic composition in countries, which will have some effect if there are ethnic differences within countries in regard to emigration, or when the levels of immigration are high.

GDP per capita is used as a control variable, as the level of economic development is likely to affect voting behaviour. We use the size of the population as control variables, 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 politicization. Years of democracy is used as new democracies lack the party system stability of older democracies, which could affect the outcome in both direction. 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 used as it is a measure of

GDP per capita, population, age distribution, education level, and years of democracy have been log-transformed, because of skewness of the distribution.

### 5.1.1 Results

The only significant relationship found in table 5.1 is that of corruption, either when ethnic polarization or both ethnic fractionalization and ethnic polarization are used as independent variables. In figure 5.1, we can see the effect of when we add corruption to the regression. The almost empty second quadrant and the populated third quadrant indicate that it is uncommon with a low extent of corruption and a great extent of ethnic voting when controlled for the other variables included in the regression.

The results imply that my hypothesis that ethnic voting can be explained by the extent of corruption in a country has so far not been refuted. However, we have so far just established a correlation between corruption and ethnic voting. A correlation cannot be used as definite proof for the hypothesis that corruption leads to ethnic voting.

An alternative hypothesis, that ethnic fractionalization causes ethnic voting, which causes corruption, or that ethnic fractionalization causes both ethnic voting and corruption, can still be valid based on the information from this first analysis. To bring more clarity into which of the available alternatives seems most plausible, I will conduct another study, where the possible effects of ethnic fractionalization will be excluded. If the results are still consistent, they will form proof against the alternative hypotheses involving that ethnic fractionalization causes corruption via ethnic voting, or that ethnic fractionalization both causes corruption and ethnic voting.

Some caution can be applied in the interpretation of the results. Model 1-3 indicate an elevated level of multicollinearity, as seen in A.3.1, as a result of including both the level of democracy and the years of democracy in the model. The statistical significance however persists after the removal of years of democracy, as seen in model 4.

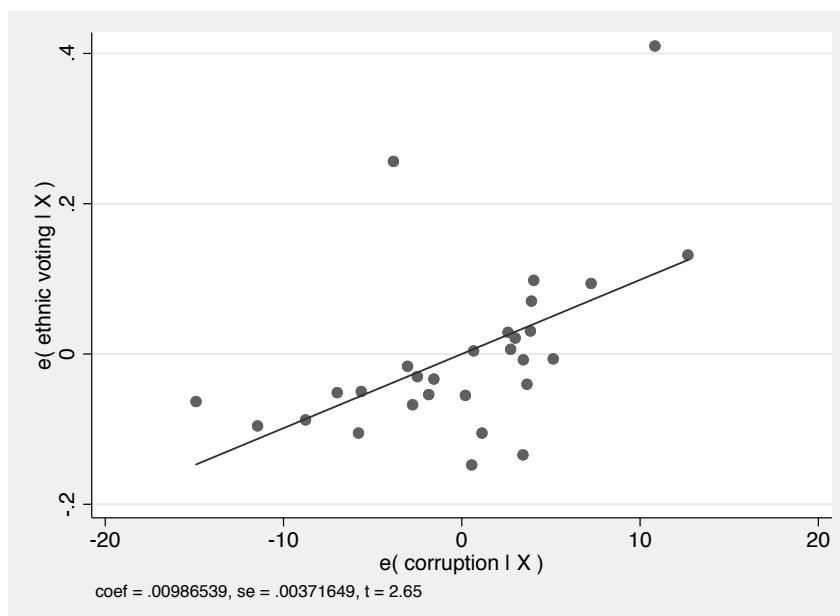
A test of heteroskedasticity in Appendix A.3.2 indicates a positive result. When we take heteroskedasticity into account in the calculation of standard errors, we will still have a significant outcome for corruption in model 3 and 4, but not in model 2. Depending on the threshold for removal of influential observations, the correlation between corruption and ethnic voting can either become insignificant at the  $p < 0.05$  level, or significant at the  $p < 0.001$  level, as we can see in Appendix A.4.1. A larger sample would probably have provided more stable results.

Table 5.1: Regression with ethnic voting as dependent variable

Ethnic voting	1	2	3	4
Corruption (0-100)	.0076363	.0094999*	.0098654*	.0097033*
Ethnic Fractionalization (0-1)	-.2090225		-.1826156	-.1755893
Ethnic Polarization (0-1)		.2528366	.2261576	.2185054
GDP/cap (log)	-.0077236	-.0331359	-.0182986	-.0237249
Population (log)	-.0096163	-.00308	.0031011	.0011554
Democracy (-10 - 10)	.0071424	.0150641	.0175063	.0117114
Population aged <14 (log)	-.2243099	-.336123	-.2671908	-.2584668
Education level (log)	-.0869075	-.0344004	-.0652779	-.0237249
Years of democracy (log)	.0074728	-.0124397	-.0207352	
R-squared	0.2647	0.3574	0.4054	0.4008
Adj. R-squared	-0.0027	0.1004	0.1238	0.1611
No. Observations	94	87	87	87
No. Groups	31	29	29	29

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Figure 5.1: Added variable plot for corruption



This added variable plot explains the effect of adding corruption to the regression model

## 5.2 Instrument variable regression on country-level

The first analysis indicated a correlation between corruption and ethnic voting. However, the correlation cannot yet tell much about the direction of causality between the two variables, with the used regression model. The first analysis could still be consistent with a causal direction going from ethnic fractionalization to corruption. We need a more sophisticated method to approach clarity about the causal direction. An experimental study would be ideal to come closer to the answer of the direction of causality. However, an experiment involving the random assignment of ethnicities and circumstances of corruption to the participants does not seem feasible.

A more accessible model is an instrumental variable approach, in which the effects of corruption on ethnic voting are isolated from the possible effects of ethnic voting or omitted variables related to ethnic voting on corruption. 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.

We have hypothesized corruption as an exogenous variable, meaning that it is not dependent of ethnic voting, but much of the previous literature has suggested that corruption is, in fact, endogenous to ethnic voting or ethnic fractionalization.

The second study recycles the variables from the first study, but this time, corruption is replaced with an instrumental variable created to minimize the possibility of measuring a spurious relationship between corruption, ethnic fractionalization and ethnic voting.

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. I use press freedom, the existence of civil society interest groups, internet access, and a measure of uneven economic development to construct the instrument for corruption. After that, the new instrument is used instead of corruption in the second stage of the regression.

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. Internet access has a similar purpose by facilitating communication and providing means to organize political demands. Uneven economic development is chosen as it is a documented source of corruption (Uslaner, 2009).

In A.6, 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 for the 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 instrument are not refuted.

### 5.2.1 Results

Table 5.2: Instrumental variable 2SLS regression with ethnic voting as dependent variable and an instrument for corruption

Model	1		2		3		4	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Ethnic voting								
Corruption		.0057139		-.009477		.010276*		.0106043*
(0-100)								
Ethnic Fractionalization	9.697718	-2.015486			11.38945	-.2078113	9.587783	-.1903837
(0-1)								
Ethnic Polarization			-7.171283		-5.983302		-4.410261	.2436762
(0-1)								
GDP/cap (log)	-1.394066	-0.183282	-2.867557	-0.052507	-2.966049	-0.045422	-1.536382	-.0522609
(log)								
Population (log)	1.427711	-0.050094	2.244932	-0.063381	1.746959	.0155757	1.962393	.0128239
(log)								
Democracy	-.877469	.0084141	-1.049396	.0180161	-1.087098	.0232032	-1.022882	.0136949
(-10 - 10)								
Population aged <14 (log)	8.988691	-1.901689	10.41443	-.3914678	11.02298	-.3210641	4.464816	-.3228846
(log)								
Education level (log)	-3.142912	-.0538854	-2.615881	.0058161	-1.294236	-.0120441	-2.601165	-.0107108
(log)								
Years of democracy (log)	2.161657	.0006513	2.870828	-.0175067	3.666908	-.0342026		
(log)								
Press freedom	.642977***		.6221962***		.7428551***		.7285321***	
(0-100)								
Civil society interest groups	2.358706**		2.638467**		2.379124**		2.043571*	
(0-10)								
Internet access	-.3406916*		-.2979701		-3.196894		-3.592081*	
(0-100)								
Uneven economic development	-2.016556		-2.432752		-.298856		-2.37894	
(0-10)								
1st stage F-test	10.931		8.066		10.781		7.050	
Kleibergen-Paap	13.150		11.859		12.650		11.775	
p=	0.0106		0.0184		0.0131		0.0191	
Hansen's J	1.250		3.463		4.832		4.575	
p=	0.7410		0.3255		0.1845		0.2057	
R-squared between		0.2494		0.3781		0.4353		0.4232
Adj. R-squared		-.05089282		-.10168514		-.13636759		-.16678165
No. Observations	87	80	87	80	80	80	80	80
No. Groups	29	27	29	27	27	27	27	27

In table 5.2, we find that the instrumental variable regression still provides a significant relationship between corruption and ethnic voting. The numerical criteria, such as the F-test, and Hanson’s J statistic, are also consistent with a strong and relevant instrument. The instrument is commented more thoroughly in appendix A.6.

The results in 5.2 are not corrected for heteroskedasticity, which should still be a problem. In A.4.2, we can find a heteroskedasticity-robust instrumental variable regression, however with a somewhat changed instrument we achieve an appropriate F-value.

The results point towards a refusal of the alternative hypothesis, that ethnic fractionalization causes corruption via ethnic voting, as the causal mechanism is most likely that corruption causes ethnic voting.

The results are also still consistent with the proposed theory and confirm the proposed causal direction as stated in the hypothesis. However, a considerable part of both my theory and the research question is devoted to the idea of an individual mechanism is involved in the causal relationship between corruption and ethnic voting. If we assume that the opposite is true, that corruption will not cause ethnic voting on an individual level, then the results from the first and second study could either mean that there is either still an omitted variable that has still not been detected by means of the instrument or that the theory suggested in section 3 is flawed. To exclude such alternative hypotheses, we will need to study the individual mechanisms involved in ethnic voting.

### **5.3 Individual perception of corruption and ethnic voting**

The third 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.

This analysis uses the Afrobarometer round 6, and contains 21711 respondents that identify with parties and have stated ethnicity (Bratton et al., 2015). 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. This will be the independent variable. We 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. The country-



Table 5.3: Multilevel, mixed-effects linear regression of with ethnic fractionalization in the party of choice as dependent variable

Fixed	
Corruption perception (0-1)	-.0400408***
More state (0-1)	-.0015645
More authoritarian (0-1)	.0146001***
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	.7403115
SE:	.0234146
Level 2 cons	.0161623
SE:	0.000
Log-likelihood	15759.834
Number of obs	21,711
Number of groups	30

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

level control variables have been retrieved from the Quality of Government dataset (Teorell et al., 2019).

A full description of the variables can be found in Appendix A.1.

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.

### 5.3.1 Results

The results in table 5.3 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 in chapter 3, the precise effect of the decrease in ethnic fractionalization will depend on how large the fractionalization is, to begin with, and the ethnic polarization.

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 a  $p < 0.001$  level.

## 5.4 Summary

In summary, the results from the first analysis suggest that there is a significant correlation between corruption and ethnic voting. The results from the second analysis suggest that the direction of causality goes from corruption to ethnic voting and that the relationship is most likely not affected by a systematic relation between the independent variable and the dependent variable, i.e., probably not affected by ethnic demographics. The results from the third study suggest the existence of an individual mechanism in which voters who perceive more corruption will also vote for more ethnically homogenous parties, i.e., vote in accordance with ethnic lines.

## 6 | Concluding discussion

The findings from the three analyses support the hypothesis that corruption leads to ethnic voting. Moreover, my theoretic suggestion that ethnic voting on a country-level is aggregated from the individual perception of corruption is also supported.

If we combine the findings from all three studies, we can concretize the findings by two hypothetical examples of a voter in a country with a majority and minority population: The first country has a problem with corruption. There are at least two parties, of which one promises to align the corruption in the voter's ethnic interest, while the other party promises to align the corruption in another ethnic groups interest. The election, becomes a clientelist competition in which the party with the largest electorate wins. If the voter belongs to a minority group that has yet been disfavored by corruption, he or she will vote for the party that promises that "it is our turn to eat." If the voter is favored by the existence of corruption, he or she will want to defend their position. The result is in any case that the voter will choose an ethnic party. The second country has no problem with corruption. Two new parties emerge in which one promises to be corrupt in the voter's ethnic interest, and one party promises to be corrupt in the interest of the other ethnic group. The voter knows that the institutions in the country will prevent the possibility of a corrupt outcome, which would make the clientelist promises along ethnic lines hard, to begin with. Besides, the absence of corruption will increase the social trust between groups, meaning that there is no reason to believe that the voters belonging to the other ethnic group will vote in only their own interest, especially if it is believed that the overall outcome of an increased ethnic competition for public goods will be harmful overall. As a result, there are few reasons for the voter to cast his or her vote on an ethnic party.

The results should be possible to generalize, with some caution. The study has not measured explicitly, for example, the hierarchy of ethnicity in countries. So if we would go from an inter-country to an intra-country comparison, it is important to take into account that minority ethnicities could have a lower place in the ethnic hierarchy, and that ethnic fractionalization does therefore not always just measure fractionalization per se. One should

want to control for the variables that contribute to the ethnic hierarchy in the first place, i.e., to closer investigate, for example, structural racism.

However, the diversity in both corruption and ethnic fractionalization should contribute to its generalizability. Except for countries such as China, Korea, and Japan, it is extremely rare with historic states that are composed of a population that is almost entirely homogenous (Hobsbawm, 2012). Large parts of Europe could also be considered an extreme case, with similar levels of corruption and a low extent of ethnic fractionalization.

Another implication of the results is that the ethnic fractionalization might not have a great effect on the existence of clientelist practices in elections, but rather on how clientelism is conducted. In multi-ethnic countries, ethnic divisions could be a more practical choice to arrange clientelism around than, for example, class divisions or urban-rural divisions, as ethnic divisions are more stable in times of economic development and urbanization.

The findings from the third individual study are apparent and significant, although the effect size is hard to conceptualize. There could also exist a possible effect that is unmeasured, as people belonging to the ethnicity that benefits from ethnic politicization and clientelism might perhaps not perceive the system as corrupt as long as he or she is on the winning side.

The suggested theory is not the most parsimonious, as it implies a somewhat more complicated causal relationship than previous thought.

The suggested theory in section 3 is references elements from many previous theories: the inequality trap, the causal direction between corruption and trust, the causal direction between inequality and clientelism and the causal direction between state capacity and the forming of ethnicities. However, my theory has the advantage that if one of the underlying theories are proven wrong, the model will still be somewhat robust. The relationship between perception of corruption and ethnic voting is supported empirically, with my somewhat independent theoretical framework of the explanation of why perceived corruption leads to ethnic voting. Even though my theory is related to the other suggested causal mechanisms, my results do in no way build on the assumption that they must all be true.

My contribution with this study is theoretic and empirical support for that corruption leads to ethnic voting. As a consequence, this means that ethnic fractionalization should not lead to corruption by means of politics. Ethnic fractionalization could still, hypothetically, lead to corruption via for example conflicts in the bureaucracy or inter-cultural communication differences, but it seems unlikely that we observe such a relationship on a systemic level without being manifested by ethnic favoritism in politics. In this sense, my theory achieves a higher degree of parsimony than the theory that ethnic fractionalization leads to corruption.

One of the major flaws of the previous research on the relationship between ethnic fractionalization and both corruption and trust is that it is in many cases, not motivated by theory. This study fills that gap, by suggesting

a theory, with empirical results that are not only consistent with the theory, but that also indicates that the mechanisms as proposed in the theory are consistent with the results.

The implications of this study cast new light on the causal direction involved in the correlation between ethnic fractionalization and all proposed measures that are also correlated with the extent of corruption, such as trust, welfare provisions or economic development. Perhaps corruption per se is a better explanation than ethnic fractionalization when it comes to for example variances in levels of trust. The findings in this thesis could make it meaningful to revisit many of the findings of proposed effects of ethnic fractionalization, especially when corruption is a contending explanatory variable.

It should be noted that corruption per se is not necessarily the best explanatory variable behind ethnic politicization on a country-level. The overall quality of government or the rule of law could be plausible contenders without having to remake the entire suggested theory. However, corruption has the advantage in this study of having a conceptual link between the individual and aggregated level in a direct way.

Corruption seems to have low explanatory power of ethnic politicization in non-democratic regimes, such as Eastern Europe in recent history. Corruption was prevalent also during the Communist era in USSR and SFR Yugoslavia, but the ethnic conflicts arose after their fall. The process of ethnic politicization could perhaps depend on the possibility of actors to articulate ethnic differences, which should not be the case when the public debate is controlled. Nationalism is a much less plausible explanation of the fall of the USSR than economic factors, meaning that the ethnic politicization in Asia and Eastern Europe during the post-soviet era should be explained by factors other than just the ethnic composition that also existed before the fall of the Berlin Wall. The extent of corruption could however perhaps be used as an explanation in a comparison of the outfall of ethnic politicization among post-soviet states. One example of this could be the development of ethnic voting in Georgia during the Rose Revolution in 2003, after which the corruption in the country was drastically reduced (Worldbank, 2012). In the early years of Georgian democracy, the extent of ethnic voting was rather great, according to the definitions by the Manifesto Project (Krause et al., 2018). In 1995, 6.84 percent voted for ethnic parties. In 1999, 25.18 percent voted for ethnic parties, and even in the election that sparked the combat against corruption, 2003, 18.8 percent of the votes went to ethnic parties. Since then, the ethnic parties have received 0 percent in every election.

The findings from this study open up many more topics to investigate. What is for example the reason of that some autocracies develop ethnic politicization, while others do not? Another topic of interest should be a more quantitative evaluation of how politics and state institutions affect ethnicities and ethnic fractionalization in the very long term. Hobsbawm created a

very plausible theory of the causal relationship between the two and provided much evidence in support for the theory, but the statistical comparisons are absent in his works.

In summary, we have reason to believe that the sense of ethnic identity has a more institutional explanation than previous thought, and that ethnic differences do not lead to corruption and conflict by necessity.

# References

- Ahlerup, P., & Olsson, O. (2012). The roots of ethnic diversity. *Journal of Economic Growth*, 17(2), 71–102.
- Alesina, A., Baqir, R., & Easterly, W. R. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4), 1243–1284.
- Alesina, A., Devleeschauwer, A., Easterly, W. R., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8(2), 155–194.
- Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, 43(3), 762–800.
- Bernhardsson, E. (2015, Apr). *Ping the world*. Retrieved from <https://erikbern.com/2015/04/26/ping-the-world.html>
- Borders, R. W. (2018). *Press freedom index. version 2018*. Retrieved from <https://rsf.org/en/ranking>
- Bratton, W., Mattes, W., & Gyimah-Boadi, E. (2015). *Afrobarometer (round 6)*. Retrieved from <http://www.afrobarometer.org/>
- Cerqueti, R., Coppier, R., & Piga, G. (2012). Corruption, growth and ethnic fractionalization: a theoretical model. *Journal of Economics*, 106(2), 153–181.
- Chandra, K. (2007). *Why ethnic parties succeed: Patronage and ethnic head counts in india*. Cambridge University Press.
- Chandra, K. (2011). What is an ethnic party? *Party Politics*, 17(2), 151–169.
- Chen, F.-Y. (2017, Feb). *The taiwanese see themselves as taiwanese, not as chinese*. Retrieved from [https://www.washingtonpost.com/news/monkey-cage/wp/2017/01/02/yes-taiwan-wants-one-china-but-which-china-does-it-want/?noredirect=on&utm\\_term=.f04ba1477dde](https://www.washingtonpost.com/news/monkey-cage/wp/2017/01/02/yes-taiwan-wants-one-china-but-which-china-does-it-want/?noredirect=on&utm_term=.f04ba1477dde)
- De Jong, E., & Bogmans, C. (2011). Does corruption discourage international trade? *European Journal of Political Economy*, 27(2), 385–398.
- Dewan, T., & Shepsle, K. (2011). Political economy models of elections. *Annual review of political science*, 14, 311–330. Retrieved from <http://search.proquest.com/docview/911909810/>
- Dincer, O. C. (2008). Ethnic and religious diversity and corruption. *Economics Letters*, 99(1), 98–102.

- Dinesen, P. T., & Sønderskov, K. M. (2015). Ethnic diversity and social trust: Evidence from the micro-context. *American Sociological Review*, *80*(3), 550–573.
- Easterly, W. R. (1997). Africa's growth tragedy policies and ethnic divisions. *The quarterly journal of economics*, *112*(4), 1203–1250.
- Easterly, W. R. (2001). Can institutions resolve ethnic conflict? \*. *Economic Development and Cultural Change*, *49*(4), 687–706.
- El Koubi, E. N. (2016). Ethnic political parties and civil conflict. *LSU Master's Theses*(1203). Retrieved from [https://digitalcommons.lsu.edu/gradschool\\_theses/1203/](https://digitalcommons.lsu.edu/gradschool_theses/1203/)
- Esarey, J., & Schwindt-Bayer, L. (2019). Estimating causal relationships between women's representation in government and corruption. *Comparative Political Studies*. Retrieved from <https://doi.org/10.1177/0010414019830744> doi: 10.1177/0010414019830744
- Foundation, M. I. (2017). *The ibrahim index of african governance*. Retrieved from <http://mo.ibrahim.foundation/iiag/>
- Gerring, J. (2007). *Case study research : principles and practices*. New York: Cambridge University Press.
- Glaeser, E. L., & Saks, R. E. (2006). Corruption in america. *Journal of Public Economics*, *90*(6), 1053–1072.
- Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance*, *3*(1), 23–45.
- Haken, N., Messner, J., Hendry, K., Taft, P., Lawrence, K., Anderson, T., & Whitehead, A. (2015). *Failed states index 2006-2018*. Retrieved from <http://fundforpeace.org/fsi/>
- Hobsbawm, E. J. (2012). *Nations and nationalism since 1780: Programme, myth, reality*. Cambridge university press.
- Holmberg, S., Rothstein, B., & Nasiritousi, N. (2009). Quality of government: What you get. *Annual Review of Political Science*, *12*, 135–161. Retrieved from <http://search.proquest.com/docview/743014576/>
- Horowitz, D. L. (1985). *Ethnic groups in conflict*. Berkeley: University of California Press.
- Hroch, M. (1993). From national movement to the fully-formed nation: the nation building process in europe. *New Left Review*(198).
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). The worldwide governance indicators: a summary of methodology, data and analytical issues. *World Bank Policy Research Working Paper*, *5430*.
- Knack, S., & Azfar, O. (2003). Trade intensity, country size and corruption. *Economics of Governance*, *4*(1), 1–18.
- Krause, W., Lehmann, P., Lewandowski, J., Matthieß, T., Merz, N., Regel, S., & Werner, A. (2018). Manifesto corpus. *WZB Berlin Social Science Center*.



- Laclau, E., & Mouffe, C. (2001). *Hegemony and socialist strategy : towards a radical democratic politics* (2. ed. ed.). London: Verso.
- Lankau, R., & Strauss, S. (2007). Mutual feedbacks maintain both genetic and species diversity in a plant community. *Science*, *317*(5844), 1561–1563. Retrieved from <http://search.proquest.com/docview/213587170/>
- La Porta, R., Lopez-de Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law, Economics, and Organization*, *15*(1), 222–279.
- Marshall, M. G., Jagers, K., & Gurr, T. R. (2018). Polity iv project, political regime characteristics and transitions, 1800-2017. *Center for Systemic Peace*.
- Marx, K. (1867). *Chapter twenty-six: The secret of primitive accumulation*. Capital: Volume One.
- Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, *69*(2), 263–279.
- Milgram, S. (1970). The experience of living in cities. *Science (New York, N.Y.)*, *167*(3924).
- Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, *95*(3), 796–816.
- Montstat. (2011). *Population census 2011*. Retrieved from <http://www.monstat.org/eng/page.php?id=57&pageid=57>
- Olson, M. (1993). Dictatorship, democracy, and development. *The American Political Science Review*, *87*(3). Retrieved from <http://search.proquest.com/docview/214434459/>
- Orjuela, C. (2014). Corruption and identity politics in divided societies. *Third World Quarterly*, *2014*, Vol. 35, Iss. 5, pp. 753-769, *35*(5), 753–769.
- Persson, A., Rothstein, B., & Teorell, J. (2013). Why anticorruption reforms fail - systemic corruption as a collective action problem. *Governance- An International Journal Of Policy Administration And Institutions*, *2013*, Vol. 26, Iss. 3, pp. 449-471, *26*(3), 449–471.
- Posner, D. N. (2004). Measuring ethnic fractionalization in africa. *American Journal of Political Science*, *48*(4), 849–863.
- Posner, D. N. (2005). *Institutions and ethnic politics in africa*. Cambridge: Cambridge University Press.
- Putnam, R. D. (2007). E pluribus unum : Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies*, *30*(2), 137–174.
- Rothstein, B. (2013). Corruption and social trust: Why the fish rots from the head down. *Social Research*, *2013*, Vol. 80, Iss. 4, pp. 1009-1032, *80*(4), 1009–1032.
- Rushton, J. P. (2005). Ethnic nationalism, evolutionary psychology and genetic similarity theory \*. *Nations and Nationalism*, *11*(4), 489–507.

- Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: anthropological and historical perspectives on the social construction of race.(author abstract). *The American Psychologist*, 60(1).
- Sovey, A. J., & Green, D. P. (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, 55(1), 188–200. Retrieved from <http://search.proquest.com/docview/822736815/>
- Standaert, S. (2015). Divining the level of corruption: A bayesian state-space approach. *Journal of Comparative Economics*, 43(3), 782–803.
- Stiftung, B. (2018). *Transformation index of the bertelsmann stiftung 2018*. Retrieved from <http://www.bti-project.org/en/index/overview>
- Stokes, S. C. (2011). *Political clientelism*. Oxford University Press. Retrieved from <http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199604456.001.0001/oxfordhb-9780199604456-e-031>
- Teorell, J., Kumlin, S., Dahlberg, S., Holmberg, S., Rothstein, B., Pachon, N. A., & Svensson, R. (2019). *The quality of government oecd dataset*. Retrieved from <http://www.qog.pol.gu.se>
- U.S. Bureau of Labor Statistics. (2018). *Current population survey (cps)*. Retrieved from <https://www.census.gov/programs-surveys/cps.html>
- Uslaner, E. (2013). Trust and corruption revisited: how and why trust and corruption shape each other. *Quality & Quantity*, 47(6), 3603–3608.
- Van den Berghe, P. L. (1987). *The ethnic phenomenon*. ABC-CLIO.
- Vanhanen, T. (2012a). Ethnic conflict and violence in heterogeneous societies. *The Journal of Social, Political, and Economic Studies*, 37(1), 38–66. Retrieved from <http://search.proquest.com/docview/1009278501/>
- Vanhanen, T. (2012b). *Ethnic conflicts*. Ulster Institute for Social Research.
- Worldbank. (2012). Fighting corruption in public services: chronicling georgia's reforms. *World Bank Publications*.

# A | Appendices

## A.1 Descriptive statistics and sources

### A.1.1 Country-level studies

Table A.1: Descriptive statistics of country-level analyses

Name	Description	Observations	Mean	Min	Max	Standard deviation	Source	Dataset name
Ethnic voting	A calculated score from the Afrobarometer. See El Koubi (2016) for methodology	95	.1523646	.0287226	.6642144	.1027122		
Corruption	Bayesian corruption indicator. A composite measure of the perception of corruption, made from over 20 surveys with 80 questions.	95	55.58202	31.82352	68.84654	9.130019	(Teorell et al., 2019). Original source: (Standaert, 2015)	bci_bci
Ethnic Polarization	Index of ethnic polarization as described in the method chapter	87	.5212184	.017	.897	.1994286	(Montalvo & Reynal-Querol, 2005)	
Ethnic Fractionalization	Index of ethnic fractionalization, involving an index of racial and lingual characteristics.	94	.6868416	.255	.930175	.1790953	(Teorell et al., 2019). Original source: (Alesina et al., 2003).	al_ethnic
Population aged <14 (log)	Total population between the ages 0 to 14 as a percentage of the total population.	95	3.712429	2.939627	3.916642	.171184	(Teorell et al., 2019). Original source: (Kaufmann, Kraay, & Mastruzzi, 2010)	wdi_pop14
Population (log)	Total population.	95	16.32323	12.20562	19.0412	1.333797	(Teorell et al., 2019). Original source: (Kaufmann et al., 2010)	log_pop
GDP/cap (log)	GDP per capita based on purchasing power parity.	95	7.899128	6.474549	9.95715	.8715598	(Teorell et al., 2019). Original source: (Kaufmann et al., 2010)	log_gdp2
Democracy	Polity measure ranging from -10 (strongly autocratic) to +10 (strongly democratic).	95	5.157895	-4	10	3.579837	(Teorell et al., 2019). Original source: (Marshall, Jaggers, & 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	3.865133	3.277145	4.446175	.2654298	(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	2.240244	0	4.26268	1.221612		
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	25.01085	5.5	54	10.39274	(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	5.613636	2	10	1.578824	(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	7.673913	3.5	9.7	.9829167	(Teorell et al., 2019). Original source: (Haken et al., 2015)	ffp_ued
Internet access	Share of population who have used the internet during the last 3 months.	95	13.3147	.3844893	58.27124	14.3402	(Teorell et al., 2019). Original source: (Kaufmann et al., 2010)	wdi_internet

### A.1.2 Individual level 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 et al., 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 in order 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 et al., 2018)

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

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

Table A.2: Descriptive statistics of individual-level analysis

asdf	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	31,280	-.5723053	-1.30416	.9495435	.5039418
GDP/cap (log)	32,472	6689.129	796.9608	60028.93	10606.51

In table A.2 we can see the descriptive statistics of the variables



## A.2 Histograms

Figure A.1: Histograms of the used variables in the country-level regression

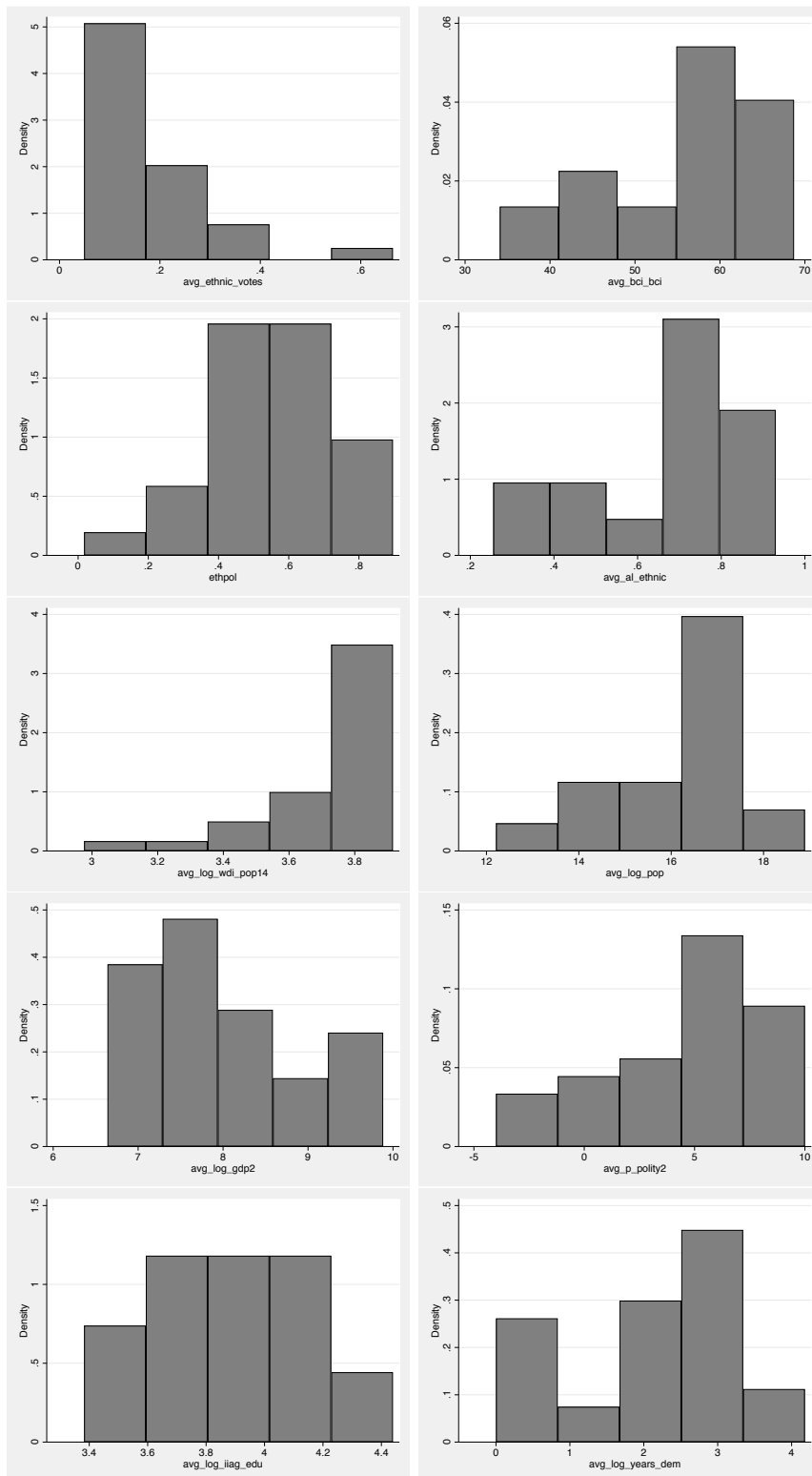
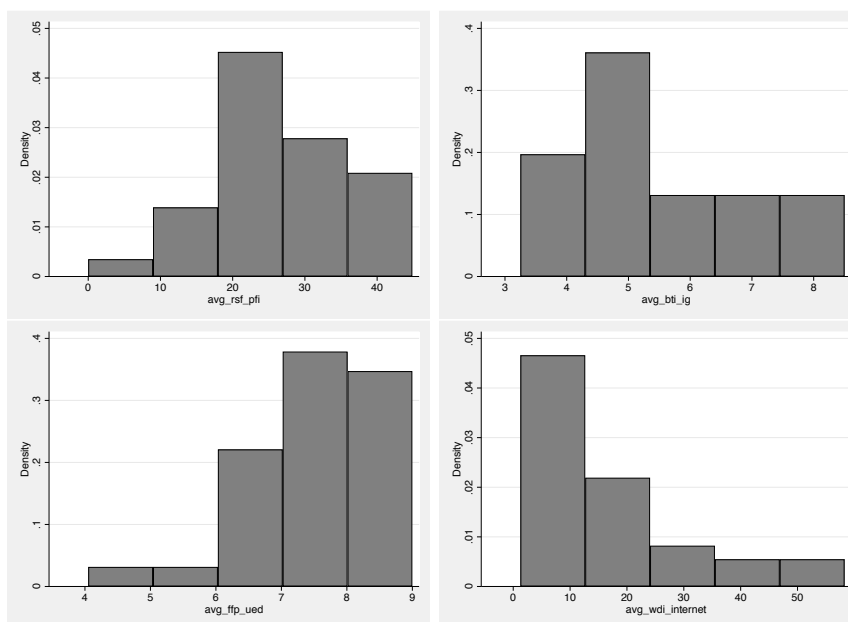


Figure A.2: Histograms of the used variables in the instrument variable



## A.3 Regression diagnostics

### A.3.1 Multicollinearity

In table A.3, 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 A.4

Figure A.3: Histograms of the used variables in the individual study

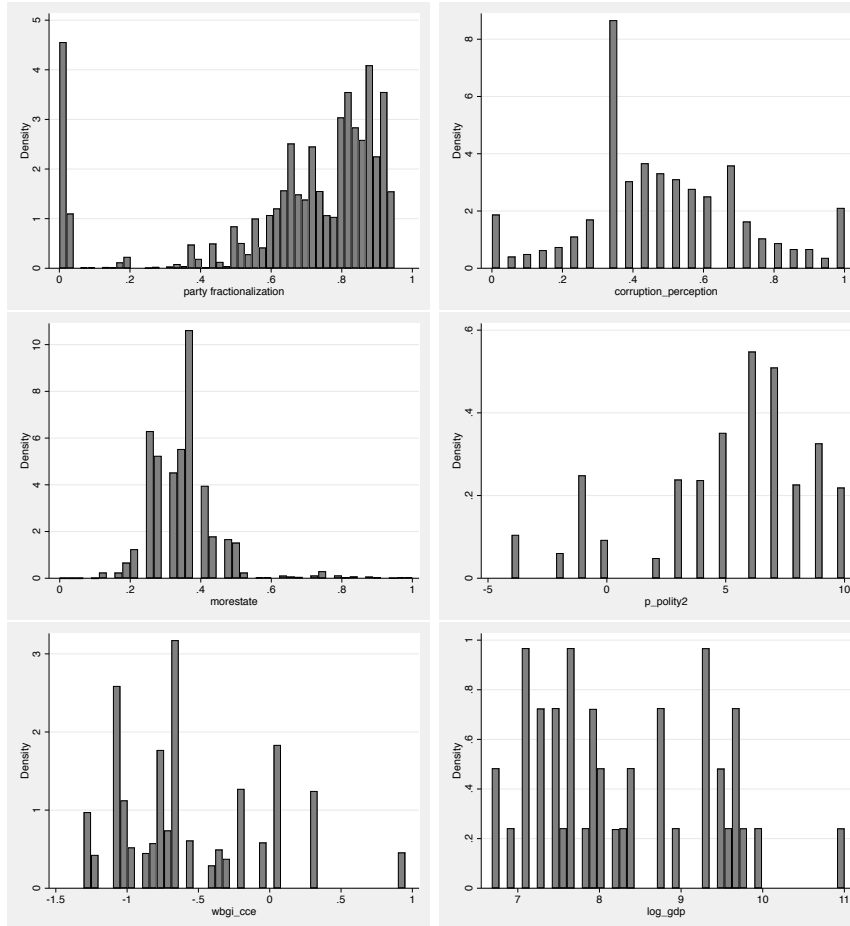


Table A.3: Multicollinearity scores

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 A.4: 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 A.5: 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

### A.3.2 Heteroskedasticity

Table A.5 and A.6 indicates that the data has problems with heteroskedasticity.

Table A.6: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

$$\begin{aligned} \text{chi2}(1) &= 5.57 \\ \text{Prob} > \text{chi2} &= 0.0183 \end{aligned}$$

Table A.7: Robustness check after removal of observations with Cook's D  $> (4/29)$

Ethnic voting	1	2	3	4
Corruption (0-100)	.0037865	.0047903	.0044162	.0041257
Ethnic Fractionalization (0-1)	-.0384626		.057935	.082352
Ethnic Polarization (0-1)		.1890546	.2038748	.1966765
GDP/cap (log)	.0170704	.0213259	.0207112	.0079682
Population (log)	-.0014763	.0036667	.0033815	-.0003543
Democracy (-10 - 10)	.0184887	.0184887	.0205195	.0072989
Population aged <14 (log)	-.3531562	-.3531562	-.2929778	-.2715339
Education level (log)	-.2108935	-.1799603	-.1775067	-.161079
Years of democracy (log)	-.043453	-.049494	-.0474234	
R-squared between	0.3484	0.4679	0.4734	0.4207
Adj. R-squared	0.0587	0.2315	0.1946	0.1632
No. Observations	82	82	82	82
No. Groups	27	27	27	27

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

## A.4 Robustness checks

### A.4.1 OLS regression

In addition to the levels in table A.7 and A.8, there are no Cook's D values below 1.

Table A.8: Robustness check after removal of observations with Cook's D  
 $>(4/87)$

Ethnic voting	1	2	3	4
Corruption (0-100)	.0084062**	.0081524***	.0085195***	.0088965***
Ethnic Fractionalization (0-1)	-.1511772		-.0694883	-.0378287
Ethnic Polarization (0-1)		.1996643***	.1846588**	.1823253**
GDP/cap (log)	.0166598	.0110051	.0096512	-.0028314
Population (log)	-.0127952	-.0066131	-.0050089	-.0099064
Democracy (-10 - 10)	.0199551*	.0196684**	.02134**	.0101585**
Population aged <14 (log)	-.3736565**	-.3373488***	-.3319749***	-.3235353**
Education level (log)	-.183181*	-.1338297*	-.1386934*	-.0953115
Years of democracy (log)	-.0382837	-.0358743	-.0398984*	
R-squared between	0.7481	0.8810	0.8909	0.8463
Adj. R-squared	0.6042	0.8130	0.8153	0.7585
No. Observations	77	77	77	77
No. Groups	23	23	23	23

\* $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table A.9: Robustness check after correction for heteroskedasticity

Ethnic voting	1	2	3	4
Corruption (0-100)	.0076363	.0094999	.0098654*	.0097033*
Ethnic Fractionalization (0-1)	-.2090225		-.1826156*	-.1755893
Ethnic Polarization (0-1)		.2528366**	.2261576	.2185054*
GDP/cap (log)	-.0077236	-.0331359	-.0182986	-.0237249
Population (log)	-.0096163	-.00308	.0031011	.0011554
Democracy (-10 - 10)	.0071424	.0150641	.0175063	.0117114
Population aged <14 (log)	-.2243099	-.336123**	-.2671908**	-.2584668
Education level (log)	-.0869075	-.0344004	-.0652779	-.0598282
Years of democracy (log)	.0074728	-.0124397	-.0207352	
R-squared	0.2647	0.3574	0.4054	0.4008
Adj. R-squared	-0.0027	0.1004	0.1238	
No. Observations	94	87	87	87
No. Groups	31	29	29	29

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$



#### A.4.2 Instrumental variable regression

Table A.10: Instrumental robustness check after correction for heteroskedasticity, with new instrument

Model	1	2	3	4
	1st	2nd	1st2	2nd
Ethnic voting (0-100)	.011972	.0242289	.0199239*	.0188464*
Corruption				
Ethnic Fractionalization (0-1)	10.74107	-2260723	9.783765	-2111163
Ethnic Polarization (0-1)				
GDP/cap (log)	-1.25692	-0.025584	-0.11868	-1.617222
Population (log)	.9066329	-0.172738	-0.0234759	.7954111
Democracy (10-10)	.1996623	.0099871	.0291125	.0496346
Population aged <14 (log)	16.96676	-2918421	-5.259629	13.53922
Education level (log)	-890743	-0.793814	.0085047	-1.042155
Years of democracy (log)	-2757081	.0046671	-.0340098	.2822893
Press freedom (0-100)	.4548647**		.4698707**	
1st stage F-test	15.503	6.520	8.675	10.003
Kleibergen-Paap	9.718 (p=0.0018)	5.531 (p=0.0187)	6.885 (p=0.0087)	6.973 (p=0.0083)
Hausman's J				
R-squared between	0.2198	neg	0.1764	0.2104
Adj. R-squared	-.0639758		-.21377638	-.0543007
No. Observations	87	80	80	87
No. Groups	29	27	27	29

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

## A.5 Ethnic voting scores

Table A.11: Calculated ethnic voting scores from Afrobarometer 3-6

Country	Round			
	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
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

## A.6 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, internet access and civil society interest groups could have a gatekeeping function against corruption, if corrupt practices have a larger chance of being detected. But the use of internet and news media could also contribute to the access to information regarding corruption, which should have an effect on corruption perception even with constant levels of corrupt practices.

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. Internet access, 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.

A well-functioning internet infrastructure should be a precondition for internet access, and the construction of such an infrastructure could be helped from the absence of corruption, meaning that a decrease in corruption could lead to an increase in internet access. However, an investigation of this relationship points towards that the absence of corruption does not seem to be a precondition for well-functioning internet infrastructure, as many corrupt countries that neighbor countries with fast internet speeds, seem to take advantage of that themselves (Bernhardsson, 2015). If this is true, internet access should first and foremost be an issue of being able to purchase a smart phone, and phone prices are steadily decreasing.

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 internet access to ethnic voting would be a higher degree of fake news and new forms of ethnic mobilization. Another could be that the internet community creates a higher degree of trust between ethnic groups, which would decrease ethnic voting. However, neither is internet directly access correlated with ethnic voting

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 & 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?** Both press freedom and internet access have theoretical maximums. It is plausible that the effect of increased press freedom and internet access will decrease with higher values of the variable, if we hypothesize information about corruption as a form of disease, where internet users and 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 at least hypothesize that internet access can be spilled over from neighboring countries with a high degree of internet access. I do not exclude that the other instruments can be subject to spillover effects, but these will probably be more subtle and indirect.