UiT THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Humanities, Social Sciences and Education

The role of inequality in state fragility

A quantitative analysis on the role of inequality in state fragility

—

Anders Torgeir Lind

Master's thesis in Political Science, May 2018

Acknowledgements

I must express my most profound gratitude to my wife Pernille for her unrelenting support, motivation and love, without her none of this would have been possible.

I would also like to thank my thesis advisor Marcus Buck for his insight, guidance and support during this process. Additionally, I would like to thank Tor Midtbø, Kristian H. Haugen and Sigbjørn Svalestuen for valuable help and input regarding the methodology of the thesis as well as the remaining members of the Stein Rokkan Research Group for Quantitative Social and Political Science and political science community at the university.

Lastly, I would like to thank my parents and remaining family for their continued support and motivation.

Tromsø, 7. Mai 2018

Abstract

This thesis examines the role of inequality in state fragility. Specifically, the direct effect of income inequality on state fragility as measured by the Fragile States Index and the interactions between income inequality and spill-over effects, national prosperity and non-tax state income. As both levels of inequality and state fragility vary over time longitudinal data from the Quality of Government dataset is used in combination with temporally fixed data from other sources.

By employing polynomial growth curve models the issue is explored, finding support for a positive effect of income inequality on state fragility as well as some support for interaction effects between income inequality and national prosperity. There is found no support for interaction effects between either income inequality and spill-over effects or income inequality and non-tax state income.

The findings contribute to elucidating the role of income inequality in state fragility.

List of abbreviations

GDP – Gross Domestic product

BIC – Bayesian Information Criterium

LR – Likelihood-ratio

ML – Maximum Likelihood

FSI – Fragile States Index

Table of contents

List of Tables

List of Figures

1 Introduction

Following the end of the cold war, the issue of the failed state, a state completely unable to fulfil its role as a state, came into the zeitgeist. As the cold war ended, so did both western and soviet policies of propping up weak states to extend their influence. As these policies ended, some states were unable to function, leading to a situation of state failure. As this impacted not only the populations of the states, but also the overall security of regions, the issue was brought into focus both among politicians, security professionals and researchers. Since the initial interest in failed states following the end of the cold war the focus of the research field has shifted somewhat, from the issue of the failed state to the more fluid terminology of the fragile state. While the term "failed state" indicates a binary situation, wherein a state is either failed or non-failed, fragility can be considered a description of a state's capabilities. As such the fragility term is more general, as it encompasses a quality rather than a specific situation. The difference of the two terms are a difference of degree rather than a difference of terms, while fragility and fragile states are more widely used terms, describing qualities of states the more restricted failed states term describes a specific situation wherein the state has ceased to provide the base functions associated with statehood. As a part of this shift, the index now known as the "Fragile States Index" changed its name from its original "Failed States Index".

The issue of fragile states is one of interest across several fields, form development and aid to international and national security. Indeed, it is quite possibly the defining international issue of our time, impacting most of the worlds states directly or indirectly in a variety of different forms from migration to security. As he quote from Gates (2010) show, the issue is one of importance even for those far removed from the fragile states themselves.

"Fractured or failing states is, in many ways, the main security challenge of our time"
– Robert M. Gates, US Secretary of Defence (Gates, 2010)

While the causes behind state fragility are varied and complex, this thesis takes a specific aim, to elucidate the role economic inequality, another central issue of our time, plays in it. Economic inequality has been the subject of much discussion, both academically as well as politically in recent years with researchers such as Thomas Piketty and Richard Wilkinson shedding lights on the wealth concentration and the societal effects of inequality. Politically, left-wing politicians such as Bernie Sanders in the USA and Jeremy Corbyn in the UK have focused on economic inequality as a driving force behind many of what they claim are undesirable developments in western society in recent years. Additionally, it is argued that economic equality has played a major part in the development of the most prosperous states, in effect paving the way for the levels of prosperity seen particularly in western countries in recent decades.

"Rich nations are rich largely because they managed to develop inclusive institutions at some point during the past three hundred years." - Acemoglu and Robinson (2012, p. 364)

Given the importance of these issues, their proposed connection as well as the interest they have both gartered in recent times it is of academic interest to further examine the relationship between the two. Boix (2015, p. 83) argues that economic inequality is a product of political institutions as well as the economy. As such inequality is a phenomenon of our own making, and perhaps more importantly, a phenomenon we can change. A reduction of state fragility is a core goal of both security and development politics. To be able to facilitate change, the causes behind fragility must be examined. It is my sincere hope that this thesis may play a part in this effort, however small the contribution may be.

1.1 Research question

Based on the foundations laid out above, the following research question has been derived.

What role does economic inequality play in state fragility?

The research question alone is broad and lacks the specificity needed for statistical testing. This is dealt with through the definition theoretically based hypotheses. These are presented in chapter 2. In addition to the hypotheses the chapter defines state fragility and economic inequality, as well as exploring theories associated with state fragility. Chapter 3 is dedicated to the methodology and research design employed to test the hypotheses, while chapter 4 describes the variables used and the process of data collection. Chapter 5 describes the variables using descriptive statistics. In chapter 6 the results of the regression are presented before these are discussed in chapter 7. The diagnostics preformed on the regression model is presented in chapter 8, before the thesis is rounded off with some concluding remarks in chapter 9. Following the concluding remarks are the list of works cited and the appendix.

2 Theoretical and conceptual frameworks

While much discussed in the decades following the end of the cold war, the idea and issue of state failure, or more generally the lack of political order, is no new subject of discussion.

Thomas Hobbes (1651) describes, in his seminal work Leviathan, the effects of a lack of political order upon society.

"In such condition, there is no place for Industry; because the fruit thereof is uncertain; and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no Instruments of moving, and removing such things as require much force; no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters; no Society; and which is worst of all, continuall feare, and danger of violent death; And the life of man, solitary, poore, nasty, brutish, and short." (Hobbes, 1651, p. 179)

As the quote above points to, and both modern-day governments and scholars argue, a lack or absence of political order blocks the way for development. Without political order, the development so sorely needed by many states is not attainable (Bates, 2008, p. 17).

In this chapter, I examine the theoretical and conceptual frameworks behind the key concepts of the thesis, economic inequality and state fragility, as well as the connection between the two. Towards the end of the chapter I examine some other explanatory variables of interest.

2.1 State fragility

State fragility is one of the core themes of international politics, development and security (Brock, Holm, & Sørensen, 2012, p. 9). In this section, I examine both the concept of state fragility as well as its causes.

2.1.1 **Ideas of statehood**

The severely fragile state can in many ways be considered the antithesis of the Weberian ideal state. Defined by Weber (1948, p. 78) as "…a human community that (successfully) claims the *monopoly of the legitimate use of physical force* within a given territory.", the basis of the ideal state rests in the wielding of legitimate physical force by a community. Building on this understanding of the state Brock et al. (2012, p. 16) proposes a twofold understanding of the state based on the two key relationships within. Between individuals as a community and between the state and the individual in terms of citizenship. Successful states manage both these relationships successfully, creating both a unity among the individual inhabitants and a relationship of privilege and responsibility between individuals and the state. Fragile states on the other hand, tend to fail at both these relationships.

In fragile states, the relationship between individuals are often dominated by ethnic, tribal, religious or societal communities rather than a state-spanning common community. While many successful states also can be said to consist of a myriad of different communities, these are often what we can describe as "second-tier communities", local or otherwise subordinate to the state community. In many fragile states these communities can be considered "first-tier communities" in competition with the state community rather than subordinate to it. The reasons for this differ from state to state but some argue that it can be traced back to the origins of the state itself. Many of today's fragile states are former colonies and have to a large extent had their borders drawn by outsiders. This has led to a situation where different communities have been lumped together in a state against their will and without any common overarching community. Some communities have also been divided by the borders, leading to cross-border communities. Simply put, producing a mismatch between the physical borders of the state and the boundaries of the communities. In addition to this, many of these states have failed, or been unable, to develop the institutions needed to produce a state-wide community of individuals (Brock et al., 2012). As mentioned above, there is a historical context to state formation. While in many fragile states a consequence of a colonial past, other factors also come into play. While the European states were formed in a context of violent competition for land, land was abundant in pre-colonial Africa, where many of today's fragile states are located. This abundance of land led to a lower level of competition compared to Europe and conflict here tended to take the form of raids for the capture of women, cattle and slaves rather than for the control of land. This weakened the impetus for state building in pre-colonial Africa. Add to this the colonial powers disinterest in building strong states in fear of losing control of their colonies and the stage is set for states with mismatched borders and weak institutions. In Latin-America the formation of the modern states took a different path, as populations were decimated and empires toppled by the invading Europeans 16th and 17th century colonial powers established a system of government where the colonial elites ruled over the poor masses, setting the stage for elite domination (Brock et al., 2012).

The relationship between the individual and state can be viewed as an exchange. Organized through institutions individuals in successful states provides the state with resources trough taxes, conscription and other forms of participation in state endeavours. In return the state delivers to the individual the substances of citizenship, legal and political rights as well as public services. This exchange creates an interdependence between the state and the individual and supports the formation of a state-wide community. Fragile states tend to fail to develop both the relationship, interdependence and institutions. In turn they face problems of cohesion and community, as well as the direct consequences from a lack of important services and institutions (Brock et al., 2012).

To summarize, the failure to both create a state-wide community of individuals and a interdependent relationship between individuals and the state leads to a lack what Brock et al. (2012, pp. 17-18) defines as "community of sentiment" and "community of citizens". This can lead to a fractured state where the state is unable to function as a unified community developing institutions and claiming the monopoly of legitimate use of physical force.

2.1.2 A definition of terms

While Brock et al. (2012) argues for using the term "fragile" rather than "failed" about the states described above others, such as Rotberg (2004) and Woodward (2017), tend to use the term "failed". As I argue in the introduction, there is merit in using both terms. However, while fragility describes a quality of the state, failed implies a specific situation in which the state is unable to fulfil its role. As such, failed can be an apt term when describing a specific situation, but the more fluid fragility a more productive term when examining the overarching phenomenon. As discussed in the introduction, there have been a shift away from the term failed in favour of fragile. However, the term failed is still useful for describing a specific situation described above. Both the OECE and the Fund for Peace, who both assess the fragility of states now use the fragility term with the Fund for Peace even changing the name of its index to reflect this change. Characteristics of fragility are discussed in following part of the chapter while a closer examination of assessments of fragility is presented in Chapter 4.

"Fragility is defined as the combination of exposure to risk and insufficient coping capacity of the state, system and/or communities to manage, absorb or mitigate those risks"
– OECD definition of state fragility (OECD, 2016, p. 22)

2.1.3 Characteristics of fragility

While they tend to share some characteristics, there is a definite heterogeneity among fragile states. While some struggle with a lack of institutional capacity, others are war-torn, fail to control domestic conflict or have suffered a complete collapse of the central state (Brock et al., 2012, p. 14). In order to better define the "fragile state", Brock et al. (2012, p. 16) constructs a summarizing concept, a Weberian ideal type of the fragile state. They define the ideal type of the fragile state as a state in which institutions and administrative structures are inefficient and corrupt, rule is based on selective coercion rather than legitimacy and rule of law, and there is a lack of effective mechanisms for holding leaders accountable to the population. This ideal is constructed as an opposite to the Weber's ideal state (Brock et al., 2012, p. 16).

In addition to the ideal type, Brock et al. (2012, p. 35) more broadly defines fragile states as "…politically, economically and socially weak entities" and provides an overview of the key characteristics of the fragile state divided into three categories, "Government", "Economy" and "Nationhood". Governmental characteristics to a large extent mirror the ideal-type explained above, including corrupt and inefficient administrative and institutional structures, rule based on selective coercion, a lack of monopoly of the legitimate use of force and a low level of state legitimacy. The economic characteristics defined include a lack of a coherent state-wide economy capable of providing a sustained basic level of welfare for the population and resources for the effective running of the state apparatus. Dependence on the world market and external economic interests as well as a national economy based on an amalgamation of fragments of modern industry, a small urban sector and traditional agriculture. The characteristics connected with nationhood to a large extent mirror the idea of

community presented in “Ideas of Statehood” above and include a lack of citizenship rights and a divided population dominated by local, ethnic or religious structures.
In real terms these states tend to be dominated by self-seeking kleptocratic leaders basing their rule on clientelism, patronage, nepotism and corruption. A lacking sense of citizenship and public accountability among the population, and external intervention (Brock et al., 2012).

While both the ideal type as well as the characteristics provided by Brock et al. (2012) are quite clear they lack operationalization. Several researchers and institutions, including the OECD and the Fund for Peace, provide such operationalization. I return to the specifics of the Fragile States Index produced by the Fund for Peace in chapter 4.

It can be argued, as Brock et al. (2012, p. 37) does, that the international standing of the right to sovereignty insulates fragile states from competition and conquest. The reasoning for this is that the international acceptance as sovereignty as an unbreakable right of the state severely limits the threat of conquest and occupation. This reduces competition between states and enables fragile states to remain sovereign without the fear of being invaded by others. In turn reducing the incentive for leaders to build stronger states able to withstand external competition (Brock et al., 2012).

The Fund for Peace describe some of the most common features of the fragile state. They summarize that the fragile state has lost both the physical control of their territory as well as the monopoly of legitimate use of force. The state also lacks a legitimate authority to make collective decisions and is unable to provide a reasonable level of public services. It is also unable to interact with other states as a full member of the international community (The Fund for Peace, 2017).

2.1.4 Criticism

While the debate about fragile states tend to focus on cause, effect and what to do with these states, there is also an ongoing debate over the core idea of state fragility. Brock et al. (2012, p. 20) responds to critique from Jones (2005) who claims that the idea of fragile states is “an ideology of the imperialism of our time” and that the “historically specific, international and local social relations” which are the root causes of fragility are instead played off as “local, and indigenous" in the prevailing explanations of state fragility and failure (Jones (2005) quoted in Brock et al., 2012, p. 20). In their response Brock et al. (2012) points to the fact that the term is "not purely descriptive" (Brock et al., 2012, p. 20) and that by identifying both core characteristics and analysis of specific cases the problems pointed out by Jones (2005) is avoidable. While I have been unable to attain the original paper by Jones, the response and quotes provided by Brock et al. (2012) leads to the conclusion that the critique chiefly concerns itself with a narrow understanding of the term, leaving out its more specific explanatory additions. If understood this way, without context, there is some validity to the claim that the term lacks historically specificity. This however, is added by accompanying analysis of cases in question. While I will argue that this critique, as I understand it, lacks validity it does bring up an important point of context and the importance of looking beyond the numbers and general definition of the term.

Ayers (2012) also critiques the term in a similar vein, arguing that it is both under-theorised as a concept and profoundly ideological. Central to this critique is an assumption that the term prescribes domestic and internal causes of state failure and argues that failed states are merely a product of global social relations in which states operate. While it is possible to ignore the external forces acting upon a state while labelling it fragile, it is not required by the term. And while the term has a natural focus on the nature of the internal situation in the state this does not preclude an external explanation for the situation. As such the critique is flawed as it fails to address that while the "facts on the ground" necessarily must be the initial subject of analysis it is entirely possible to expand the analysis by attempting to explain the external forces creating or influencing the internal situation in the state. The argument put forward by Ayers (2012) that the term is based primarily on western standards of statehood is also, in my view, flawed. This because while yes, fragile states fail to live up to western standards based upon the Weberian ideal-state these states also fail their own citizens by not providing them with security, political, social and human rights. As such these states fail in several regards.

Ayers (2012) also criticizes the term for its use, or misuse, as a pretext for intervention. While this can indeed be a problem, it is not the fault of the term but rather its use and users. The labelling of a state as fragile or failed does not necessitate intervention, it merely points out the shortcomings of the state in question. How to improve such states will depend upon the case, and it is entirely possible to argue against the need for intervention without needing to discard the term.

2.2 Inequality

Inequality is a general term describing an uneven distribution as opposed to its antonym equality which describes an even distribution. The term can be used to describe uneven distributions of a different things, apart from economics the issue of inequality is often used in describing the distributions of power, education and other subjects. As the focus of the thesis is the role of economic inequality, the distribution of monetary resources among individuals, on state fragility it is this kind of inequality which is the inequality discussed.

Inequality can be exemplified by imagining a group of 1000 people where the total income is 100.000.000. If one person earns 100.000.000 and the rest zero, there exists perfect inequality as one person receives all the income. However, if this income is evenly distributed among the people, with each earning 100.000, there is perfect income equality among the members of the group. While perfect inequality and equality are rare in the real world, the degree of economic inequality differs greatly, both between states and other entities such as cities, regions, groups etc.

"The distribution of income and wealth is shaped by both the structure of the economy and the nature of political institutions" (Boix, 2015, p. 83)

As Boix points out in the quote above, economic inequality is a result of human organization, how we choose to organize our societies and economies. I this thesis I focus on economic inequality on an individual level, and in turn my analysis is limited to the distribution of monetary assets, among individuals in different states. As I return to later in this chapter, economic inequality is often divided into income inequality and wealth inequality. While they both depict important sides of economic inequality and tend to correlate to a degree, they differ in some key respects (Piketty, 2014, 2015), it is the inequality of income that will feature most prominent in this thesis.

2.2.1 **Income inequality**

While the term “income” for many is closely associated with wages earned from labour, the term also includes earnings from sources such as self-employment, pensions, social transfers and capital income (Piketty, 2015, p. 5). In The Economics of Inequality Thomas Piketty (2015, pp. 5-6) refers to a French survey from the year 2000 in which a mere 58.8% of total household income comes from waged labour, and another 5.8% from self-employment. The remaining 35.4% coming from pensions (21.3%), social transfers (9.5%) and capital income (4.6%). Note that capital income, income from wealth, is not accurately reported as household income, and some accounts put this number at about 10%. This distribution is a general feature of western economies (Atkinson et al., 1995, p. 101 referenced in Piketty, 2015, p. 7). While, it is reasonable to expect the distribution between the different types of income to vary between states depending on welfare systems. The difference will perhaps be most noticeable between western European states such as France and fragile African states. However, the main takeaway from the numbers reported above is that income is more than just wages. Whatever the source, income inequality measures the distribution of income.

An alternative measure to income inequality could be wage inequality. Wage inequality measures inequality in wages among workers only, leaving out those not employed, and income attained from other sources such as social transfers, pensions and capital income. As non-wage incomes, especially capital income, are generally less equally distributed the measures of income inequality, in general, shows higher inequality than the measures of wage inequality (Piketty, 2015, p. 13). While wage inequality is interesting as it shows the distribution of wages, income inequality shows a more complete picture of the economy as it takes more sources and more individuals into account. Due to this, income inequality will be preferred over wage inequality as a measure of inequality in this thesis.

2.2.2 **Wealth inequality**

Wealth inequality, the distribution of wealth among the individual members of the population, is another form of economic inequality. While there is a theoretical question of how, or rather what to include, when attempting to calculate income inequality, the calculation of wealth inequality is quite straight forward provided we have reliable data. To calculate the inequality of a wealth distribution we merely add up the total wealth and calculate the percentage of it held by each individual, group or percentile. As I eluded to above the reliability of such calculations rests on the reliability and accuracy of the data used. This is in many places a challenge, as statistics tend to be based on self-reported numbers and in turn underestimates wealth, particularly of the wealthiest individuals (Piketty, 2014, p. 258). Even statistics not based on self-reporting run the risk of underreporting as wealth can be placed outside the state in question or hidden from official numbers. This is a general issue associated with all measures of wealth. The findings of Gabriel Zucman (2013) illustrates this point. Zucman (2013) argues that about 8% of global financial wealth of households is held in tax havens, of it only about a quarter is reported. This is a sizeable amount of wealth, enough to shift the status of the eurozone from the world's second largest net debtor, to net creditor.

To an even larger extent than income, wealth is unevenly distributed. While the top 10 percent of income earners tend to receive 25 to 30 percent of the total income, the top 10 percent wealth holders tend to have more than 50 percent of total wealth, with some societies reaching as high as 90 percent. These higher levels of inequality become even clearer if we instead look at the bottom 50 percent, while they typically receive about 25 to 33 percent of total income, they generally hold under 5 percent of wealth (Piketty, 2014, pp. 244-245).

2.2.3 **Measures of inequality**

To measure inequality is not as straight forward as it may seem. Assuming we have reliable data, a topic I return to in chapter 4, this can be done in a few different ways. While the GINI coefficient may be the most famous measure, the interdecile indicators such as the P90/P10 ratio are simpler and more intuitive (Piketty, 2015, p. 10).

Piketty (2015, p. 8) defines the P90/P10 ratio as "...the ratio of the lower limit of the tenth decile to the upper limit of the first decile.". While the calculation of the P90/P10 is straight forward, the key lies in what we are measuring. While wage inequality for individual workers is simple to calculate, a deeper consideration and assessment is needed for the calculation of household income adjusted for household size or inequality in disposable income (Piketty, 2015, p. 14). As I return to in chapter 4, the primary explanatory variable used in the analysis is the proportion of income garnered by the top decile. This is a measure easily accessible from the World Bank trough the Quality of Government dataset (Teorell et al., 2018), while it differs somewhat from the format of the P90/P10 ratio the top decile income measure is essentially the same as it reports the percentage of income befalling the top decile.

2.2.4 Discussion

As mentioned above, both income and wealth inequality capture important aspects of economic inequality. The use of both would have strengthened the findings of this thesis. However, in addition to the issues of hidden wealth discussed above, there is more general issue of the availability of data. While data on income inequality is available from a variety of states, comparable data on wealth inequality are far harder to come by. Additionally, it is entirely possible that wealth and income inequality have different effects on society. To avoid the issues associated with wealth data as well as isolating the effect of income inequality, this thesis focuses on income inequality. As I return to in chapter 9 the inclusion of wealth inequality in the research of inequality's role in state fragility would be welcomed.

2.3 The role of inequality in state failure

As Charles Boix (2015, pp. 46-49) summarizes, equal distribution of resources among members of foraging societies were key to achieving prosperity. In these small and early societies, distribution of resources was transparent and the society self-governing. Under such conditions an unequal distribution of resources quickly lead to plunder, violence and chaos. To avoid these outcomes and secure the stability and prosperity of the society, equal distribution was key. Simply put, it can be said to have been a requisite for a functioning society. Since the days where the foraging society was the primary societal unit of the world, human society have evolved into much more complex forms. The question then turns to whether the importance of equality have changed alongside it, or if the lessons learned by the foraging societies of the past still maintain their relevance today. Have we lost the need for equality, or have we merely constructed a society so complex we are no longer able to identify and react to inequality in the same way as our forbearers? Obviously, this question is not one to which a definite answer can be provided. We can, however, examine the effects of economic inequality on society.

In 'The Spirit Level – why equality is better for everyone' Wilkinson and Pickett (2010) find strong evidence for an adverse effect of higher levels of inequality on societal bonds as well as health issues. In their findings, there is a strong correlation between levels of societal and health issues and levels of inequality. The effects found are not limited to at-risk groups or sections of society, but rather whole populations and cover a plethora of different issues associated with health and societal issues. While their research focuses on rich developed countries the details and nuances of their findings might not be directly transferable to other states. As their research is thorough and the results robust it is nonetheless reasonable to expect a general negative impact of inequality on society and health. While it may make sense intuitively to assume that these effects are the result of a lower levels of poverty rather than inequality, the research show that this is only a small part of the issue. While it is conceivable that poverty may indeed play a larger part in states that are worse off, this does not negate the findings concerning inequality. As the findings of Wilkinson and Pickett (2010) show that a vast majority of individuals, at all levels of prosperity reap the benefits of lower inequality the effect is clearly separate from that of poverty. When discussing the research of Wilkinson and Pickett (2010) it is interesting to note that while it is cross-sectional, and as such unsuited to uncover changes over time and the direction of causality, it is backed up by a number of other studies conducted at different times. Wilkinson and Pickett (2010, p. 31) also argues that because inequality is a product of the broader societal structure it is in effect showing the effects of said societal structure upon the individuals within. As a society is defined by the individuals within, this points to a feedback loop where inequality in practicality can be argued to be the effect of society upon itself.

2.3.1 Societal capabilities

I argue that in sum, the effects of inequality on society described above is the weakening of societal capabilities. Building on the works of Boix (2015) and Wilkinson and Pickett (2010) it is reasonable to assume that inequality, trough the increase of societal issues and the weakening of societal bonds both present a threat to societal institutions as well as reduce the capacity of the population to change the institutional status quo. The quality and inclusiveness of institutions are key in the economic development of states (Acemoglu & Robinson, 2012, p. 364).

In many fragile states extractive institutions play an important part in both politics and economy. These institutions are set up to extract riches from the population. Coupled with corrupt regimes, extractive institutions create what Acemoglu and Robinson (2012, p. 343) aptly names “the vicious circle”. In it the extractive political institutions create, or upholds, extractive economic institutions funnelling revenue from the economy to the institutions under private control of the ruling elite. The extractive economic institutions enrich the ruling elite, giving them incentive to further strengthen and uphold the extractive political institutions. As this circle both creates unconstrained power and income inequality benefitting those in power, the stakes of the political game are heightened. As there are few paths to prosperity in such societies except through the state, those with ambitions have two choices, to align themselves with the existing elite and attempt to work their way towards the top, an option often closed to those who are not already a part of or associated with the ruling elite, or to challenge the rulers and take their place atop the state, often turning the struggle for power violent in the form of rebellion, infighting and civil war (Acemoglu & Robinson, 2012). Thus, the combination of self-seeking elites and extractive institutions can lead to, not only the extraction of riches but also prolonged power-struggle, violent clashes and civil-war. These challenges to the ruling elites take different forms, but often end up continuing the vicious circle, merely replacing the old elite with new elites continuing the extractive and self-seeking practices left behind. Described by Robert Michels as “the iron law of oligarchy”, where the extractive institutions ensure the continuation of extractive practices even when the old rulers are replaced. As Dawit Wolde Giorgis, a former minister to President Mengistu of Ethiopia points out recalling the changing face of Mengistu following the successful overthrow of Emperor Haile Selassie in 1974 (Acemoglu & Robinson, 2012).

“We were supposed to have a revolution of equality; now he had become the new Emperor.”
– D.W. Giorgis (quoted in Acemoglu & Robinson, 2012, p. 360)

The vicious circle also manifests itself in other ways, such as in Colombia where the lack of inclusive institutions leads to a failure of incentives for politicians to provide public services, law and order in large parts of the territory. The additional lack of constraints on politicians fail to prevent them from striking deals and arrangements with paramilitaries (Acemoglu & Robinson, 2012, p. 383).

Discussing poverty Kalyvas (2007, p. 419) states that poverty, in its relative form a consequence of high inequality, both lowers opportunity cost for joining a rebellion as well as creating grievances and demand for social change. It is reasonable to assume that these effects are transferable outside of the civil war case and that poverty also lowers opportunity cost for joining in other groups or activities undermining state institutions. As such, this can be viewed as a mechanism part of the vicious circle. As low opportunity costs of behaviour negatively impacting society weaken institutions and the state, the opportunity costs are further lowered.

Whether emerging as the result of lacking societal bonds, or through other processes, and whatever forms they take, the extractive institutions through the effects of the vicious circle pave the way for state failure (Acemoglu & Robinson, 2012, p. 372).

In more prosperous and democratic states, such as the United Kingdom, a different circle can be said to have taken effect. "The virtuous circle" where inclusive political institutions create and strengthen inclusive economic institutions. These sets of institutions include an increasing number of citizens, distributing wealth and power among the people, in turn strengthening the inclusive political institutions. Part of this circle is the distribution of resources (Acemoglu & Robinson, 2012, p. 364), or in the terms of this thesis, economic equality. Just as economic inequality plays a part in allowing for the extractive institutions of fragile states, economic equality plays a part in strengthening inclusive institutions in successful states.

Discussing the importance of inclusive economic institutions where mass participation is possible, Acemoglu and Robinson (2012, pp. 74-75) argues that to achieve this the institutions must include a set of key features. Secure private property, an unbiased legal system, individual freedom of choice regarding career, the permission of new business and the provision of public serviced providing a level playing field. A lack of these features enables self-seeking elites to cement their power and continue their exploitation of the state. In turn increasing inequality. At the same time, existing economic inequality can weaken or inhibit the formation of these institutional features by strengthening the position of the elites vis-à-vis the remainder of the population. As such it is important to note that it is the universal availability of the features that is key, not their mere existence for a portion of the population. As an example, we can consider secure private property. If this is not extended to the whole of the population, it serves only to concentrate and preserve the property of those in power.

While obviously not the singular cause, I argue that inequality, through the weakening of societal capabilities, contribute for a situation wherein society fails both to challenge the status quo of extractive institutions as well as allow for situations where extractive institutions can be established. This in turn sets the stage for increased state fragility.

"Nations fail economically because of extractive institutions"
(Acemoglu & Robinson, 2012, p. 398)

The ideas of the vicious and virtuous circles can be viewed as a form of path dependency where the initial position of a given state is a strong predictor of its future position. I argue that inequality, through its effect on societal capabilities impact a state's chances to escape the vicious circle.

Based in the theories described above H_1, the first hypothesis derived from the research question is:

H_1: Increased economic inequality increases state fragility.

Discussing the 'conflict trap' P. Collier et al. (2003, p. 53) argues that some of the root causes of violent conflict are related to economic development, particularly low and unequally distributed per capita income. P. Collier (2007, p. 19) also finds a strong correlation between countries incomes and the likelihood of civil war, these findings are also supported by a consensus amongst researchers (Cunningham & Lemke, 2014, p. 328), supporting his earlier arguments. As mentioned both earlier in the chapter, violence is an important component of civil war and, as I return to later in the chapter, violence and civil war shares many causal factors with state fragility. Given the apparent connection between overall country wealth and fragility its interaction with inequality is explored in hypothesis H_2:

H_2: Increased GDP pr. capita lowers the impact of inequality fragility.

P. Collier (2007, p. 31) argues that the effects of conflict, and by extension state fragility, spills over across borders. While I return to a deeper discussion on both the theoretical assumptions behind spill-over as well as the operationalization of the effect later in the thesis, hypothesis H_3 explores the interaction between such spill-over effects and inequality:

H$_3$: Spill-over effects increases the impact of inequality fragility.

Bates (2008, pp. 25-27) idea of the “state failure equilibrium” presented below, the levels of non-tax state revenue are also an interesting economic metric when discussing the impact of inequality on state fragility. Assuming that there is an intertwining of effects behind state fragility, the interaction of non-tax revenue and inequality is one of interest. To explore this relationship hypothesis H4 is specified as follows:

H$_4$: Non-tax revenue increases the impact of inequality on fragility.

The direct effects of country wealth, spill-over and non-tax revenue, along with other explanatory variables, on state fragility are discussed in the last part of this chapter.

2.3.2 Theoretical model

Presented graphically Figure 1 below shows the effects described above. While inequality directly impacts state fragility, GDP pr. capita, non-tax revenue and spill-over effects impact the effect of inequality. The graphical representation of the model does not encompass the expected direct effects of GDP pr. capita, non-tax revenue, spill-over effects or other explanatory variables. These effects are described in the latter part of this chapter.

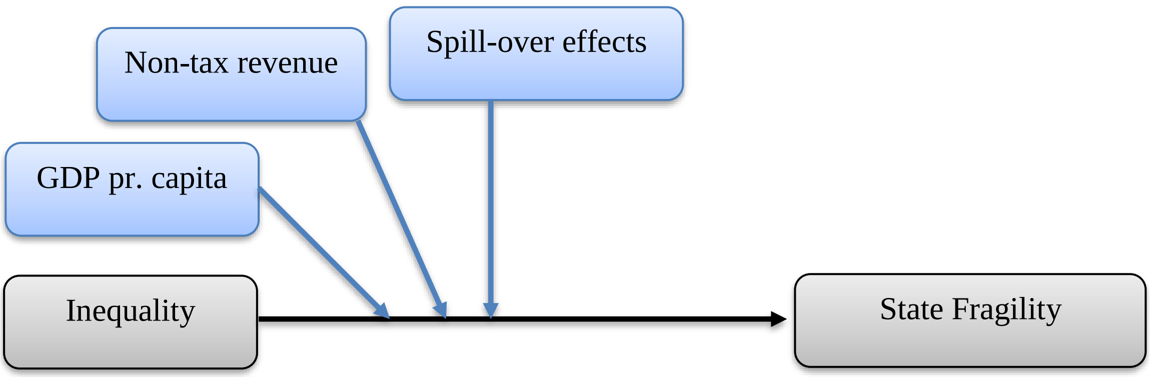


Figure 1 - Theoretical model

2.3.3 Hypotheses

Stated as hypotheses, the theoretical model based on the discussion above becomes as follows:

H_1: Increased economic inequality increases state fragility.

H_2: Increased GDP pr. capita lowers the impact of inequality fragility.

H_3: Spill-over effects increases the impact of inequality fragility.

H_4: Non-tax revenue increases the impact of inequality on fragility.

2.4 Other determinants of state failure

Following the findings of Cunningham and Lemke (2014, p. 339) and a theoretical understanding of the causes of state fragility. A number of factors initially researched as causal factors affecting the likelihood of civil war are below taken into account as potentially having a similar effect on fragility.

"…factors identified as affecting civil war have similar effects on a variety of other types of internal violence." - Cunningham and Lemke (2014, p. 339)

As Kalyvas (2007, p. 416) notes, civil war can also be a precursor to prolonged situation of state fragility as civil war degrades whatever remaining state capacity there is a its onset. This strengthens the linkages between civil war and state fragility.

2.4.1 Taxation and state revenue

Using data from the era of decolonization, P. Collier (2007, p. 19) finds a strong correlation between countries income and the likelihood of civil war. A finding that is supported by a consensus amongst scholars (Cunningham & Lemke, 2014, p. 328). This relationship works both ways, as poverty increases the chances of civil war, civil war produces poverty. P. Collier (2007, p. 20) also finds that growth rates impact this relationship, with higher growth rates lowering the risk of civil war. He also provided a possible explanation for this relationship, arguing that low incomes and growth equates poverty and hopelessness, conditions under which the opportunity costs of engaging in rebellion and other activities inductive to state fragility, are low. Poverty and hopelessness particularly impacts young men, who, when deprived of work, income and hope, can be a destabilizing factor in society. As with other indicators of civil war it is reasonable to assume a similar correlation with state fragility. The findings of P. Collier (2007) echo the financial component of Brock's (2012) definition of the fragile state presented earlier in this chapter. In this definition the inability of the economy to provide a sustained basic level of welfare is a defining feature of fragility.

Bates (2008, pp. 25-27), discussing problems of state failure in Africa, presents the idea of a "state failure equilibrium". A situation in which a sufficiently high tax income makes taxation more profitable for rulers than predation. At the same time tax rates must fall below a specific threshold for the payment of taxes to be more profitable than the individual provision of security by citizens. Provided these two levels of taxation do not overlap, meaning that there is in fact room for a situation where both ruler and citizenry gain more from taxation than they lose without it, state fragility is avoided. As there are costs associated with state failure, primarily in the form of a loss of prosperity, the gains to either party must outweigh the costs of failure. This equilibrium can be shifted by external sources of revenue benefiting the rulers of the state. Assuming such revenues are sufficiently large relative to earnings from taxation, it may become more efficient for rulers to move their efforts away from the protection/taxation arrangement with the citizenry and rather direct these towards the sources of external revenue, securing these at the cost of the state apparatus. This can create a situation under which the development and protection of external revenues is more rational for self-seeking elites than the maintenance of the state and taxation system. In such a situation the state would certainly move towards fragility, leaving the citizens to fend for themselves and orienting the remainder of the state to the extraction of revenues external to the ruler-citizenry relationship. Such external revenues tend to consist of natural resources, such as gemstones or oil, but are not necessarily limited to these. High-value, easily extracted natural resources such as the ones mentioned above, with the addition of narcotics, are also often used as funding for rebel groups providing both the incentives and means of civil conflict (Buhaug, Gates, & Lujala, 2009, p. 555; P. Collier, 2007, p. 21). Gemstones and oil also has a positive correlation with prolonged conflict duration according to research done by Buhaug et al. (2009, pp. 560-562). Provided a sufficiently low state capacity where the state is unable to maintain control over the resources, this kind of high value, easily accessible natural resources can be a catalyst for state fragility by working as an incentive for both the elite and citizenry to engage in activities undermining the state.

2.4.2 Demography

Cunningham and Lemke (2014, p. 336) find a positive relationship between population size and the likelihood of all types of violence, apart from coups. This indicates an increased chance of violence in more populous states.

Discussing ethnic fractionalization as a possible explanatory variable for civil war Kalyvas (2007, p. 419) recounts three different explanations for how ethnic disputes and nationalism cause civil war. A "security dilemma" between the different ethnic groups created by state collapse wherein the lack of a state power to govern relations. Leading both sides to build up capacity. A "commitment problem" between ethnic groups when there is no one to guarantee agreements between them. And ethnic secessionists attempting to break away from the state. Cunningham and Lemke (2014, p. 336) also finds a positive correlation between ethnic fractionalization and violence, however their findings indicate that the relationship takes the shape of an inverted u, where conflict is more likely at moderate levels of ethnic fractionalization.

Fearon and Laitin of Stanford however, finds no correlation between political repression of ethnic minorities and civil war (P. Collier, 2007, p. 23), this is counter to the findings of Kalyvas (2007) and Cunningham and Lemke (2014).

2.4.3 Geography

Examining the effects of geography on civil conflict, Buhaug et al. (2009, p. 560) finds that distance to capital prolongs conflict as military capacity decays as it is projected over distance. Taking this into the state fragility context the effect of decaying military capacity over distance indicates in a situation where the state's capacity to maintain its monopoly of force, as well as law and order, decay over distance. This indicates that geography can play a part in state fragility. Rugged terrain offers additional challenges for the employment of state power, particularly against smaller irregular units such as guerrilla fighters and criminal gangs (Buhaug et al., 2009, p. 552). This claim is backed up by the findings of Cunningham and Lemke (2014, p. 336) who find a positive correlation between mountainous terrain and violent conflict.

P. Collier (2007, pp. 53-58) argues that geography also matters in other ways, primarily focusing on the effects of neighbours and access to the sea he makes the point that landlocked states are worse off than others, depending on their neighbours and natural resources. Not all landlocked states suffer the effects of their lack of access to the sea, some are blessed with good neighbours with whom they can both trade and rely upon as a route of transport, others stuck in a resource trap regardless of their access to the sea. But for those without both natural resources and good neighbours, P. Collier (2007, p. 56) argues, access to the sea matters. Those in this situation are, overall, worse off than those who are not landlocked. The landlocked state with bad neighbours and no natural resources to speak off, is to a large extent an African phenomenon, as these areas elsewhere have not become independent states. This is a result of geography and history, and not something witch I will go into further detail on here.

As P. Collier (2007, pp. 56-57) notes in respect to landlocked states, growth spills over. As such, growth in one state results in growth in its neighbours. Along the same lines it is reasonable to assume that this effect also holds true for fragility, creating a spill-over effect where fragility in one state increases the chances of fragility in its neighbour. Territory outside of governmental control, such as territory held by rebels or areas where the state capacity is simply too low to allow for control, are much sought after “safe havens” for illicit and nefarious groups such as terrorist organizations and criminal networks, as well as rebels and insurgents. Having such potential safe havens close to one’s borders may prove destabilizing as the activities and fallout from such groups spill over between countries. In addition to this the costs of war, primarily violence, disease and crime, spill over from one state to another (P. Collier, 2007, p. 31).

2.4.4 Political

In A Theory of Political Transitions, Acemoglu and Robinson (2001) make an interesting point regarding the role inequality plays in determining the regime type of states. Given the reasonable assumption that democratic rule is inherent redistributional in nature, the poor in nondemocratic societies are assumed to favour democratization while the rich are assumed to resist it. Given sufficiently low opportunity cost, the poor masses in such a state can use revolution, or the threat thereof, to forcibly instigate change in the system of government. This, however, is no guarantee of democracy. Instead, given a sufficient level of inequality and power the rich elite resist such processes by mounting a coup. As these two forces, the poor masses and the rich elite, go head to head over the system of government the state can end up oscillation between regimes, landing them in a situation where government in any form fails to consolidate. This creates weak regimes, unable to exact the levels of control associated with neither democracy nor autocracy. Such states, often dubbed "anocracies", are in turn vulnerable, further worsening their predicament (Fearon & Laitin, 2003). Anocratic states have a higher likelihood of both civil war and state fragility as they lack both the controlling effects of a powerful autocratic state and the stabilizing effects of a democracy (Cunningham & Lemke, 2014, pp. 331-340; Kalyvas, 2007, p. 418). Cunningham and Lemke (2014, p. 336) also finds a positive correlation between both anocracy and political instability, and violent conflict. It is reasonable to expect both frequent regime changes and a situation of anocracy to weaken state capacity as resources are diverted towards the struggle for power and existing structures are weakened by conflict.

Regime changes are more likely to occur during economic recession or crisis (Acemoglu & Robinson, 2001, p. 939), in the case of fragile states the state can be said to be in a persistent state of turmoil creating a situation where the chances of regime change remain high over long periods of time leading to a situation where all parties remain vigilant and in struggle for control of the state.

P. Collier (2007, p. 17) argues that states experiencing repetitive, violent political conflict run the risk of getting trapped in a cycle of poverty and violence. While conceding that the outbreaks of civil wars are multifaceted with layers of causality, P. Collier (2007, pp. 18-27) dismisses the idea of grievances. Labelling them as nothing more than imagery, and instead focusing on the self-feeding loop of economic and social issues associated with violent political conflict wherein one ended civil war increases the chance for the breakout of civil war during the next decade, particularly in low income countries. A similar effect takes place regarding coups, where states who have suffered a coup is more likely to suffer another.

2.4.5 Discussion

As shown above, there is several factors assumed to influence state fragility alongside inequality. While some are derived from the study of civil war, others are more generally associated with fragility or conflict. However, state fragility is a complex issue that varies between cases. A unified theory explaining the phenomenon is unlikely, but by studying causal links behind fragility we can increase our understanding and lay the foundations for actions minimizing the likelihood of state failure.

To account for the factors discussed above, demographic, geographical, economic and political data were included in the dataset and statistical model. I return to these issues in chapters 3 and 4.

3 Methodology and research design

To statistically test the hypotheses developed in chapter 2 it is necessary to employ a methodology of research. As the research question and hypotheses focus on economic inequality and state fragility, two phenomenon who vary in time, the use of longitudinal data is appropriate. Longitudinal data, in turn, poeses an inherent hierarchical structure, where observations can be considered as nested within subjects, as a result a multilevel modelling approach is applicable to the data. To illuminate the relationship between inequality and state fragility I will utilize a growth curve model. The choice of model is grounded in the idea of the virtuous and vicious circles presented in chapter 2. These can be likened to the idea of path dependency or more colloquially the proverbial a slippery slope. Related to growth curve modelling, the core concept is the way in which the past condition of a given state is the key predictor of its future condition. In this view, state fragility has the potential to grow or diminish over time due to the conditions of the state, the very makeup of state fragility. More specifically the conditions at the outset, set the stage for what is to come. These models allow for the incorporation of natural growth into the estimations of effects, a necessity when analysing subjects in growth. To exemplify the way in which growth curve models work, we can imagine a study on the effect of weight training on teenagers. Regardless of their amount of training, we can expect teenagers, who are still growing after all, to become bigger and stronger as time goes by. Using a standard panel analysis runs the risk of not separating these effects from the effects of the training, causing faulty estimations. Here the assumption is that state fragility will develop over time, with the degree and direction being influenced by the variables.

As there is no indication that the relationship between economic inequality and state fragility is perfectly linear, the specific type of model used will be a polynomial growth curve model allowing for a curvilinear relationship.

In this chapter I present the methodology and research design used to test the hypotheses.

3.1 Methodology and the purpose of inquiry

The statistical method, which is held in high regard among naturalist scholars who see it as the task of science to uncover the patterns and regularities of the world. Can be split in to two distinct types, descriptive and inferential statistics. While descriptive statistics merely describe phenomenon through systematic collection of quantitative data, often used to illustrate and supplement scientific claims. And for this reason, also used by constructivist scholars. Inferential statistics infers from sample data, population characteristics, provides predictions as well as explanations and hypotheses, attempting to see beyond the data at hand (Moses & Knutsen, 2012, pp. 70-83).

To answer the research question at the outset of this thesis research design and methodology a necessity. The selection of method and research design should be a based on a critical review of the needs of the research question as well as a pragmatic approach to the feasibility of the research (D. Collier, Brady, & Seawright, 2010). As well as dictating the overall methodology, the research question in conjunction with the practical feasibility of the research in question, also dictates more particular aspects of the research design.

Observations in longitudinal data cannot be considered independent as there is a likely correlation between repeated measurements and a structure to the data. Using classical regression models to analyse longitudinal data thus results in too small standard errors for the parameters, inflating the t-values and impacting the significance level (Rabe-Hesketh & Skrondal, 2012, p. 2). Multilevel models are better suited to handling data with this structure.

Multilevel modelling is inherently hierarchical, directly including group indicators in the analysis, and is as such the appropriate approach for analysing hierarchical and nested data while mitigating the possibility of ecological fallacies (Gelman & Hill, 2007, p. 7; Luke, 2004, pp. 4-5). Multilevel modelling builds upon classical regression, specifying the differing intercepts and slopes of regression lines, as well as allowing for hierarchical clustered data structures. Thus overcoming both the limiting scope of the single-level approach as well as the individual/aggregate dilemma (Ruspini, 2002, p. 122)

"The multilevel model provides a coherent model that simultaneously incorporates both individual- and group-level models." (Gelman & Hill, 2007, p. 8)

While classical regression allows for the study of variations in terms of interaction, multilevel modelling allows for the modelling of varying effects between groups in a much clearer manner as it allows for the estimation of group level effects and averages even with small group sample size, allowing for a better understanding of the phenomenon in question (Gelman & Hill, 2007, p. 6).

"A multilevel model is a statistical model applied to data collected at more than one level in order to elucidate relationships at more than one level." (Luke, 2004, p. 7)

In longitudinal data, the nesting of individual observations within subjects creates a hierarchical structure suited for multilevel modelling. Here time becomes the lowest level of measurement, nested within the subject, with the subjects again able to be nested within groups (Ruspini, 2002, p. 120). Pertaining to this thesis *year*, denoted as *t*, is the lowest level of measurement, nested within *state*, denoted as *i*. As multilevel modelling allows for the study of effects that vary by group, the nesting of years within states in longitudinal data makes multilevel modelling an appropriate approach for studying changes in longitudinal data.

The research question at the base of this thesis focuses on the role of economic inequality in state fragility. To examine the interplay between these two phenomenon it is necessary to examine their change over time. Using multilevel methodology allows for a longitudinal analysis of these changes and allow for an examination of how they vary across states.

3.2 Longitudinal data

Longitudinal data, in some cases referred to as pooled cross-sectional time series, can be viewed as a collection of cross-sectional data consisting of repeated measures of variables across different points in time. There are several advantages to employing longitudinal data, chiefly the ability to examine and elucidate dynamic relationships, to model heterogeneity among the subjects of study (Frees, 2004, p. 6) and to separate the subject-specific random effects from the error term of the overall model (Frees, 2004, p. 22). Using repeated measurements of specific subjects, longitudinal data, in the form of repeated time series, allows for the modelling of subjects' behaviour additionally, longitudinal data design can provide increased efficiency of the estimators than alternative designs utilizing a comparable amount of data (Frees, 2004, pp. 7-10).

While the increased number of observations associated with longitudinal data is generally considered a good, increasing the efficiency of the model as compared to cross-sectional data, the repeated measurements of the same subject tend to be related, resulting in heterogeneity.

While longitudinal data, with its repeated measurements, contain more information than cross sectional data, the amount of additional information added need to be assessed. Heterogeneity is an issue that demands attention when modelling longitudinal data. Heterogeneity can be interpreted as correlation between observations from the same subject. A failure to properly account for this correlation in the modelling can cause serious bias to the model estimators, oft referred to as "heterogeneity bias". There are two main approaches to tackling heterogeneity when modelling longitudinal data, both incorporating subject uniqueness in the modelling. Fixed-effects models in which the subject-specific parameter (denoted as α_i) is treated as unknown fixed parameters to be estimated, and random-effects models where the subject-specific parameter is treated as draws from an unknown population and as such a random variable. In addition to the facts of the data, or assumptions of the real-life characteristics of the data, the inclusion of heterogeneity quantities into the modelling of longitudinal data can be motivated by the fear lurking variables causing omitted variable bias. Here the issue at hand concerns the variables left out of the model (Frees, 2004, pp. 8-10).

3.3 Multilevel modelling

Multilevel models applied to longitudinal data differs from other forms of multilevel models in two primary concerns. The goal of the analysis changes when applied to multilevel data with the focus moving to the assessment of change over time while the core difference separating the setup of longitudinal multilevel models from ordinary multilevel models is the definition of level-1 and level-2 units of analysis. While ordinary multilevel models often define subjects (such as individuals or states) as level-1 units of analysis, these are in longitudinal multilevel models defined as level-2 units of analysis while the observation is placed at level-1. As such the nesting structure of a two level model is changed from subject within group, to observation within subject (Frees, 2004, p. 174). Additionally longitudinal data is ordered in time, granting importance to the temporal ordering of occasions, thus removing the interchangeability common in other multilevel structures where the ordering among occasions lacks meaning (Rabe-Hesketh & Skrondal, 2012, p. 228). Variance is implicitly represented in multilevel models as a function of explanatory variables (Frees, 2004, p. 184).

While a multilevel model initially is specified in multilevel terms, the combined model can be written as a linear mixed-effects model. As a result the techniques of statistical inference used with linear mixed-effects models are valid and applicable also for multilevel models (Frees, 2004, pp. 166-168).

The need for a multilevel model tends to rest on three different causes, empirical, statistical and theoretical (Luke, 2004, p. 15). The theoretical foundations for using a multilevel model is the need to observe change in states over time. As such it is the research question, "What role does inequality play in state failure" that necessitates the use of two levels, state and year, in the model. Statistically the data, being hierarchical in nature due to the nesting of state-years in states, can be assumed to breach the assumption of independence required to perform a single level OLS analysis. This prompts the need for multilevel modelling which allows for correlated error structures (Luke, 2004, pp. 18-19). Empirically the high intraclass correlation score produced by the full model[1], .09972914 with a standard error of 0.0015479, further substantiates the need for a multilevel model.

3.3.1 Maximum Likelihood

Maximum likelihood is a widely used method for the estimation of parameters in statistical analysis. It can be summarized as the joint probability density of all observed responses in the model as a function of the parameters. The key concept of the maximum likelihood method is to find the parameters which maximise likelihood function (Hox, 2002, p. 37; Rabe-Hesketh & Skrondal, 2012, p. 101). Missing data, due to different causes such as attrition and omission, is an ever-constant issue when working with longitudinal data. One of the beneficial features of maximum likelihood is it's robust handling of data missing at random, allowing for missing observations while still providing consistent and efficient estimates of parameters (Allison, 2002, p. 12; Duncan & Duncan, 2004, p. 354).

3.3.2 Assumptions of Maximum Likelihood

While robust towards some breaches of the assumptions for the use of maximum likelihood the assumptions must still be tested and accounted for (Hox, 2002, pp. 37-38; Midtbø, 2012, p. 105; Rabe-Hesketh & Skrondal, 2012, p. 360). Below the assumptions of maximum

[1] STATA command: *estat icc* – following estimation of the full model.

likelihood regression are described and their handling discussed. Diagnostics of the assumptions are presented in chapter 8.

Validity

A prime assumption behind all statistical analysis is the validity of data. To achieve reliable results from an analysis, the data used must match the research question. This includes both using a dependent variable which in fact measures what we wish to examin as well as including all explanatory variables acting on the dependent variable (Gelman & Hill, 2007, pp. 45-46). When studying complex issues of human society and interaction, full validity can never be guaranteed due to the possibility of measurement errors and lurking variable. Even so, it is important to keep the issue of validity in mind when designing research. I return to this issue in chapter 4.

Linearity and additivity

Mathematically, the most important assumption of regression analysis is the linearity and additivity of the relationship between the dependant and independent variables in the model (Gelman & Hill, 2007, p. 46). To allow for breach of this assumption, curvilinearity is accounted for using a polynomial growth curve model. A model featuring log-transformed variables where applicable, is also run and presented in chapter 6.

Residuals

Homoscedastic residuals, where the variance of the residuals is independent of the values of the explanatory variables is a key assumption of regression. A breach of this assumption, where residuals are heteroscedastic is a general issue in regression (Midtbø, 2012, pp. 106-107; Rabe-Hesketh & Skrondal, 2012, p. 360).

Normally distributed residuals are a base assumption of regression analysis in general. A normal distribution of residuals implies a symmetrical distribution giving equal chances of over- and underestimation. This is, however, not required for a consistent estimation of model parameters and standard errors, nor for asymptotic normality of the estimators themselves (Luke, 2004, p. 30; Midtbø, 2012, p. 114; Rabe-Hesketh & Skrondal, 2012, p. 101). Gelman and Hill (2007, p. 46) argues that normally distributed residuals are the least important regression assumption, and even goes as far as not recommending running diagnostics on the residuals to examine the distributions. Yet in the interest of thoroughness, diagnostics on the distribution of the residuals are presented in chapter 8.

The use of regression analysis also assumes independence among the residuals. In a two-level multilevel model such as the one employed in this thesis dependence among residuals can take two forms, autocorrelation and intraclass correlation. While autocorrelation measures the dependence between the individual residuals, intraclass correlation measure the dependence between groups. STATAs measures of intraclass correlation in terms of rho. A high rho indicates that a large portion of the variation is between groups indicating a low variation within groups (Midtbø, 2012, pp. 112-113).

Graphical and statistical examinations of these assumptions are presented in chapter 8.

Absence of multicollinearity

Another assumption of regression in general is the absence of multicollinearity among the variables. Multicollinearity can be defined as a situation in which a there is a strong correlation between two or more of the variables. While only a perfect multicollinearity breaks the assumption for regression, high levels of multicollinearity can cause problems for the estimation of coefficients (Midtbø, 2012, pp. 128-130). The results of the Variance Inflation Factor testing are presented in chapter 8.

3.4 Growth curve

Initially developed for use in biology, growth curve models have become a popular multilevel approach for modelling longitudinal data. Growth curve models take growth into account by modelling both the shape of trajectories of individual subjects over time as well as the variation between trajectories, systematically due to variables on both the occasion- and subject-levels as well as randomly. Growth curve modelling can be considered as a special case of the random coefficient model. The random-coefficient model builds on the simpler random-intercept model and includes random coefficients in addition to random intercepts. This allows for varying effect of variables between subjects or clusters. In a growth curve model, it is the coefficient of time that varies randomly between subjects (Rabe-Hesketh & Skrondal, 2012, p. 343).

There are two main types of growth curve models for nonlinear growth, polynomial and piecewise models. While the trajectories in piecewise models are made up of separate linear segments, splines, the polynomial models employ pth degree polynomial functions of time to allow for curvilinear trajectories. The number of extrema allowed is a function of the degree of the polynomial, $p - 1$. As a result a quadratic model allows for one extrema, while a quartic model allows for up to three (Rabe-Hesketh & Skrondal, 2012, p. 345). In this thesis I utilize a cubic polynomial growth curve model, the modelling process of which is described below.

To introduce the polynomial into the practical model, trend variables was created using STATA[2].

Table 1 - Trend variables

Variable name:	
trend	Linear trend. Generated on a yearly basis starting with 0 at year 2005.
trend2	The quadratic trend term. Defined as $trend^2$.
trend3	The cubic trend term. Defined as $trend^3$.
trend4	The quartic trend term. Defined as $trend^4$.

3.4.1 **Model testing**

To build the model both theory and statistical testing was employed. Likelihood-Ratio (LR) test and the Bayesian Information Criterion (BIC) were used to assess different iterations of the models.

BIC is a measure of the explanatory power of the model, showing how well the tested model fit the data. The actual scores of BIC measure the lack of explanatory power yielding high scores for low explanatory power. While similar to $\bar{R}^2$ in terms of controlling for the number of variables, effectively punishing unnecessarily large models in order to limit the risk of overfitting, BIC is unstandardized and therefore used exclusively for the testing of models against one another (Hox, 2002, pp. 45-46; Midtbø, 2012, pp. 103-104).

[2] STATA command: *generate* (for full process, see accompanying .do file)

It is customary to employ the Likelihood-Ratio test to assess overall model fit in nonlinear regression models. The LR test is in principle a form of hypotheses testing, testing whether one model provides the same amount of information as another. The test provides a result in the form of a p-statistic, where a significant result indicates that the full model should be kept, and the alternative model rejected. In the LR testing I have used a significance level of 95% as this is customary in the academic field. LR tests requires the alternative model to be nested within the full model, necessitating the full model to include all parts of the alternative. In other words, the alternative model must be a reduced version of the full model (Rabe-Hesketh & Skrondal, 2012, p. 326).

While both BIC and LR test requires an identical number of observations in each model, it is possible to force the execution of an LR test in STATA[3]. While not ideal, this allows for a greater comparison of models.

The actual testing was done using the STATA commands *lrtest* and *estat ic* for LR test and BIC respectively. Additional information on the practical execution of the testing is presented in the accompanying .do file.

3.4.2 **Modelling**

Following the advice of Luke (2004, p. 20) the model was built from the bottom up, starting out with a simple model featuring only the dependant variable and the linear trend term. To this model, increasing degrees of polynomial trend terms, listed in Table 1, were added and tested using LR and BIC testing. At the end of this initial process, a model consisting of three polynomial degrees in the fixed part and two in the random part was selected. While the variable trend4, representing a quartic relationship, the 4th degree polynomial, was created it was ultimately not included in the model following testing.

The second tier of model building consisted of the selection of the appropriate covariance structure. The validity of model-based standard errors, and if there is missing data the consistency of point estimates of regression coefficients, is dependent upon a correctly specified covariance structure. As there is data missing from the dataset, the selection of a

---

[3] By adding *,force* to the end of the STATA command.

fitting covariance structure is important. As the data has regular time intervals, it is possible to utilize an unstructured covariance structure. This structure also allows for missing data (Rabe-Hesketh & Skrondal, 2012, p. 298). Different covariance structures were tested using BIC and LR testing. As independent covariance structure is the standard structure when using the *mixed* command in STATA, other covariance structures were tested against this. As a control, both exchangeable and identity structures were tested in addition to the unstructured structure. Following the testing, an unstructured covariance structure was selected.

3.4.3 Model revision and testing

Continuing the modelling process the explanatory variables were added and tested one by one. As the explanatory variables are all rooted in theory they were added to the model regardless of their testing scores. This contrast the handling of the control variables to which I return later in the chapter. Some variables, considered to essentially be different versions of the same variable, were tested using BIC only, as this test, unlike the LR test, allows for the replacement of variables and does not require one model to be nested within another (Hox, 2002, p. 45).

Among those tested by using BIC only, were five different variables representing spill-over effect in the form of failed neighbours. Two measuring the number of failed neighbours by maximum score on the Fragile States Index, setting the threshold at 90 and 100 respectively. The next two measuring the number of failed neighbours by mean score on the Fragile States Index, again setting the threshold at 90 and 100 respectively. Finally, the fifth variable is a dichotomous variable representing whether the state has failed neighbours (one or more) or not. The dichotomous variable is based on the variable measuring the number of failed neighbours by maximum score, using a threshold score of 90. The five different variables were tested using BIC and the dichotomous variable was kept. The results are presented in Table 2 below. For the full model and procedure see the attached .do file.

Table 2 - Selection of spill-over/neighbour variable

Variable:	**BIC:**
FailedNeighbourMaxFSI100	3729,762

FailedNeighbourMaxFSI90	3718,791
FailedNeighbourMeanFSI100	3728,336
FailedNeighbourMeanFSI90	3725,133
FailedNeighbour (dichotomous)	3704,447

Using the same procedure, two different variables measuring ethnic fractionalization was tested, resulting in the selection of **fe_etfra** with BIC 3454.791 over **al_ethnic** with BIC 3566.888.

The model featuring the base model and added explanatory variables is named 'the explanatory variables model'.

Extensions to the model

Allow for testing of hypothesis H_2, H_3 and H_4 as well as control for effects outside those included in the theoretically based explanatory variables the model was expanded. First by the addition of the interaction effects demanded by the hypothesis. The resulting model, adding interactions to the explanatory variables model is named the 'interactions model'. Table 3 below lists the interactions added to the model in this step.

Table 3 - Interactions

Non-tax revenue # Income share top decile
GDP pr. capita # Income share top decile
Border distance # Income share top decile
Failed neighbour # Income share top decile

To control for non-immediate effects, lagged variables were tested and added using LR and BIC testing. While it is customary to include the lower levels of lagged variables, when run with all yearly lags of the lagged variables, the model fails to estimate. As a result, only the ones found to strengthen the explanatory power of the model through LR and BIC testing was included. Lags included are displayed in Table 4 below. The resulting model, named the 'full model', thus includes both explanatory variables, interactions and lagged variables. The results from the three models are compared in chapter 6 and discussed in chapter 7.

Table 4 - Lags included in model

L3. GDP pr. capita
L1. Income share top decile
L2. Income share top decile
L3. Resource Revenue
L1. Non-tax revenue
L2. Non-tax revenue
L1. Tax revenue (% of GDP)
L2. Tax revenue (% of GDP)
L3. Tax revenue (% of GDP)

Finally, the control variables along some control interactions were tested as extensions to the full model. As none of these provided a better fit and increased explanatory power they were all rejected. The full process of both the control variable testing as well as the overall testing of variables and interactions are included in the accompanying .do file.

3.4.4 **Model specification**

Presented as a two-level model the formal model becomes:

Level 1: $y_{it} = \alpha_i + \beta_1 X'_{it} + \beta_2 t_{it} + \beta_3 t_{it}^2 + \beta_4 t_{it}^3 + \varepsilon_{it}$

Level 2: $\alpha_i = \gamma + Z'_i \beta_5 + \beta_6 t_{it} + \beta_7 t_{it}^2 + \varepsilon_i$
$\beta_1 = \beta_9 X'_{2it} + \beta_{10} t_{it} + \beta_{11} t_{it}^2 + \mu_{it}$

Where y_{it} is the Fragile States Index score for a state (i) in a given year (t). X'_{it} represents a set of variables with temporal variation (variation across years, within the given state), Z'_i a set of variables fixed in time (variation between states, fixed within years) and X'_{2it} a set of parameters impacting the effect of the within year variables X'_{it}. The curvilinear relationship is brought into the model by the cubic polynomial $\beta_4 t_{it}^3$, as well as its lower order terms $\beta_2 t_{it}$ (linear) and $\beta_3 t_{it}^2$ (quadratic), additionally $\beta_6 t_{it}$ + $\beta_7 t_{it}^2$ and $\beta_{10} t_{it} + \beta_{11} t_{it}^2$ adds quadratic nonlinear relationships to the second level of the model. γ represents the controlled grand mean intercept while ε_{it}, ε_i and μ_{it} represents the error terms.

Rewritten as a STATA command, the final model becomes:

mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

3.5 Issues and alternatives

As with all research in general choices have been made regarding the methodology. Other types of methodology could have been used to illuminate the research question. Below I go through some of the roads not taken. As well as discuss some issues related to inferential statistics.

3.5.1 General issues with statistical inference and longitudinal data

Sir Francis Galton, a pioneering figure of the field of statistics, lends his name to an issue of relevance. Galton's problem, wherein the lack of independence among the subjects cause false positives when examining correlation (Moses & Knutsen, 2012, p. 82). In this thesis neither the states as subjects nor the years as observations are truly independent of one another. While this is accounted for in the selection of a multilevel approach and the design of the model us, it is still important to keep in mind through all the phases of the analysis.

Another issue, not entirely independent from Galton's problem, is the issue of endogeneity. When analysing the causal linkages behind a phenomenon, we make assumptions as to which variables impact the phenomenon. What we cannot control for in statistics, is what some term 'lurking' variables. Variables which are not included in the analysis yet have an impact on the variables in question. As such, the effects we find of a given variable, may in fact be the product of an unknown variable acting upon the variable. While there is no sure fire way of eliminating the danger of lurking variables, an analysis grounded on sound theoretical ground will be more likely to avoid the issue (Moses & Knutsen, 2012, p. 90).

Attrition in the classical sense, the nonresponse of subjects in longitudinal studies, is not an issue when studying states as there are other mechanisms regulating participation than with individuals. However there remains a similar issue of attrition wherein some states fail to produce or report on the variables in a given year. While this in many cases can be a result of random events, a systematically missing data can be an issue (Frees, 2004, pp. 11-12). I return to this in chapter 8.

3.5.2 Alternative methodologies

While the methodology employed in this thesis is appropriate for providing an answer to the research question, such a broad research question can be answered in many ways.

Staying in the world of longitudinal statistics, there are variety of different versions of longitudinal analysis models which could be employed. Another alternative would be in depth case studies, focusing on the linkages and causal mechanisms between inequality and fragility in a single, or limited number of states. Methodologies such as Qualitative Comparative Analysis also offer up some interesting approaches to the study of the role of economic inequality in state fragility. As there exists little research on this specific relationship there is room for a wide variety of approaches.

4 Data collection and variables

In this chapter the data collection process and variables are described. Based on the theoretical foundations of chapter 3, the operationalization through variable selection is described below.

The data used in this analysis is primarily data from the time-series version of the Quality of Government Standard Dataset published by the Quality of Government Institute at the University of Gothenburg (Teorell et al., 2018). An exception from this is the geographical data which have been manually added using the CIA World Factbook (Central Intelligence Agency, 2017) and the world map or imported from the PRIOGRID (Tollefsen, Strand, & Buhaug, 2012) dataset. Below, both the source dataset and original data source of the variables used are described.

Table 5 presents the variables used in the model, for more details about the variables and a complete list of all variables included in the dataset please see appendix. Note that mean and standard deviation are not listed for the categorical or identifying variables.

Table 5 - Variable list

Variable:	**Identifier:**	**Mean:**	**Std.dev:**	**Source:**
DEPENDENT VARIABLES				
State Fragility – Fragile States Index:				
FSI Total score	ffp_fsi	70.712	23.478	QOG
GINI coefficient	wdi_gini	39.814	9.864	QOG
INDEPENDENT VARIABLES				
Inequality (Core model):				
Income share top decile	wdi_incsh10h, centered_incsh10h	31.235	7.588	QOG
Other explanatory variables:				
Population (in thousands)	wdi_pop1k	32228.93	118355.5	QOG
Failed neighbour *	FailedNeighbour	n/a	n/a	Constructed
Landlocked *	Landlocked	n/a	n/a	Constructed
GDP pr. Capita (US$)	wdi_gdpcapcur	6429.105	12254.12	QOG

GDP pr. capita growth	wdi_gdpcapgr	2.063	6.255	QOG
Ethnic fractionalization	fe_etfra	0.476	0.260	QOG
Regime type duration	bmr_demdur	40.304	47.035	QOG
Distance from border	bdist	166.843	166.843	PRIOGRID
Distance from capitol	capdist	406.472	480.167	PRIOGRID
Travel time to major city	ttime	502.373	551.823	PRIOGRID
Resource revenue	ictd_revres	6.661	11.401	QOG
Non-tax revenue	ictd_nontax	5.747	8.225	QOG
Tax revenue (% of GDP)	wdi_taxrev	16.985	8.580	QOG
Identifying Variables:				
Geographic region	Region			Geographic
Continent	Continent			Geographic
Country Code	ccode			QOG
Country Code	ccodecow			QOG
Country Name	cname			QOG
Year	year			QOG
Country Code and Year	ccodealp_year			QOG
FSI category (mean)	FSIMeanCategory			Constructed
FSI category (mean)	FSIMeanCategory2			Constructed
FSI category	FSICategory			Constructed
Regime type	ht_regtype			QOG
Regime type (simplified)	ht_regtype1			QOG

* = Dichotomous variable

4.1 Dependant variable - Fragility

As main measure of state fragility the Fragile States Index is used. The index was first launched in 2004 and is produced by the Fund for Peace, an independent non-profit research organization focused on the field of security (Fund For Peace, 2018a, 2018b). Using the proprietary CAST framework, the index quantifies state fragility in twelve different categories before aggregating the scores to create the main index. The index ranks all countries recognised by the UN providing there is reliable data available, a total of 178 states as of 2018. The CAST framework was originally developed in the 1990s, prior to the creation of the Fragile States Index and uses a mixed approach integrating a large amount of both quantitative and qualitative sources. The framework uses three different streams of data, content analysis, quantitative data and qualitative review, which are triangulated and subject to critical review to compile the index. The content analysis is done by using Boolean search phrases to filter through a large number of media sources on a sub-indicator basis. To create the indicator scores, a total of 45-50 million articles and reports are analysed annually. The quantitative data stream is comprised of data from pre-existing datasets from other statistical agencies, such as the United Nations and the World Bank, identified as representing key aspect of the different indicators. Following normalization and scaling for comparative analysis the quantitative data is compared to the findings from the content analysis. Matching scores are confirmed while mismatches are treated based on a fixed set of rules. Separately from the rest of the process a team of scientists review the individual states, serving both as a control and to pick up on year-on-year trends in the different states. As a final check the completed index is subject to a panel review (Fund For Peace, 2018b). Figure 2 depicts the process visually. Given the through process as well as the large number of sources the index is assumed to accurately measure fragility, providing validity to the analysis. Nonetheless, the validity of the assessments made by the index can be subject of discussion as the process is proprietary and as such not easily accessible for outsiders. The index is, however, widely used, a sign of trust in their assessments.

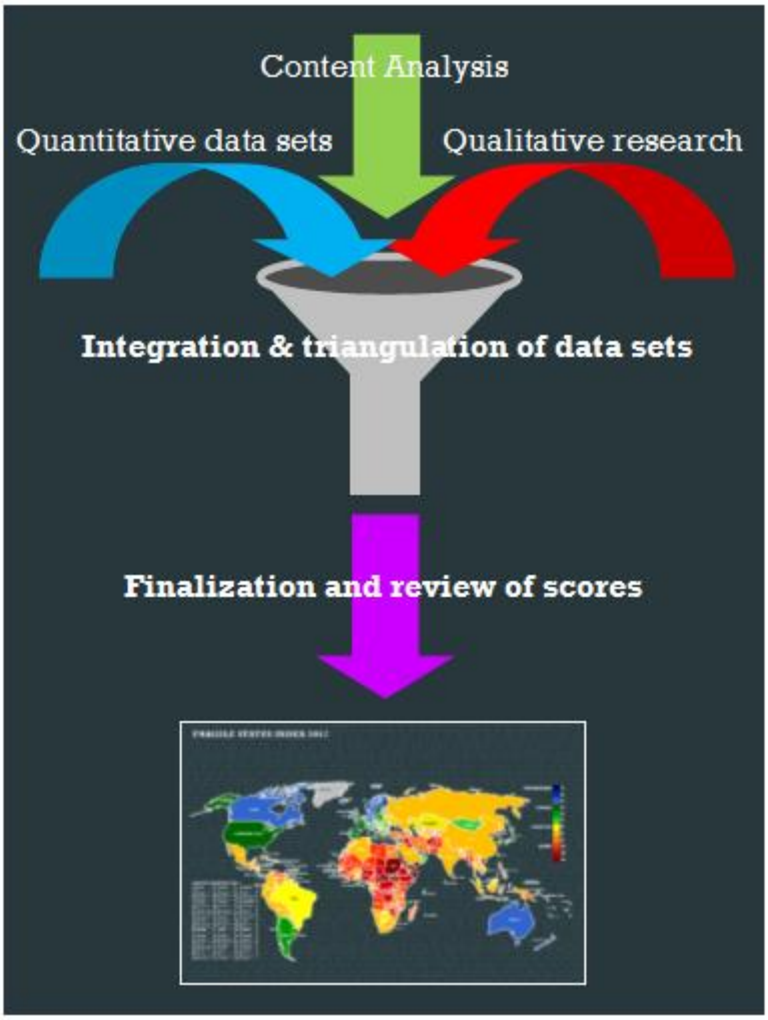


Figure 2 - FSI Process (graphics provided by FundForPeace.org)

The Fragile States Index is comprised of twelve sub-indexes covering different aspects of state fragility. I have included both the overall Fragile States Index score as well as the scores for the twelve sub categories in the dataset. Table 6 lists the categories and variable names. Additionally, a variable of each state's mean FSI score (**FSIMean**) and a variable of max FSI score (**FSIMax**) providing the max score of each state were constructed. Three different grouping variables based on the index were also constructed, **FSICategory**, **FSIMeanCategory** and **FSIMeanCategory2**. I return to these under identifying variables.

Table 6 - FSI Indicators

Aspect:	Indicator:	Variable:
Cohesion	Security Apparatus	ffp_sec
Cohesion	Factionalized Elites	ffp_fe
Cohesion	Group Grievance	ffp_gg
Economic	Poverty and Economic Decline	ffp_eco
Economic	Uneven Economic Development	ffp_ued
Economic	Human Flight and Brain Drain	ffp_hf
Political	State Legitimacy	ffp_sl
Political	Public Services	ffp_ps
Political	Human Rights and Rule of Law	ffp_hr
Social/Cross-cutting	Demographic Pressures	ffp_dp
Social/Cross-cutting	Refugees and IDPs	ffp_ref
Social/Cross-cutting	External Intervention	ffp_ext

4.2 Main explanatory variable - Inequality

As the research question states, the object of this thesis is to examine the role of economic inequality in state failure. As noted in chapter 2 there are several different ways in which economic inequality can be measured. However, income share of the top decile is used as the main explanatory variable in the model. The selection is based on Piketty's (2015) arguments for the usage of the more easily interpreted measures. In addition to the complexity, the GINI coefficient has received some criticism for not accurately describing distribution as well as expected (Chitiga & Sekyere, 2014). Both the GINI index and the income shares of the top and bottom deciles are included in the dataset.

The measurements of both the GINI coefficient and the measure of income share of the top decile comes from the World Development Indicators provided by the World Bank. And are presented as the variables **wdi_gini** for the GINI index, **wdi_incsh10h** for the income share of the top decile and **wdi_incsh10l** for the income share of the bottom decile. The data is accessed as a part of the QOG dataset.

4.3 Explanatory variables

In addition to the dependant and main explanatory variables several other explanatory and control variables. Below I go into some detail on each of the other explanatory variables, offering the reasoning for their inclusion and the sources of the data.

4.3.1 Taxation and state revenue

To account for the effects of economic growth and country income discussed in chapter 2, I have included the variables **wdi_gdpcapcur**, measuring GDP pr. capita in current US$, and **wdi_gdpcapgr**, measuring growth in GDP pr. capita. Additionally, two control variables were added to the dataset to control for total GDP, **wdi_gdppppcur**, measuring total GDP (in current US$) and real growth in GDP, **wdi_gdpgr**, measuring growth in total GDP. All four variables are collected from the QOG dataset and are original data provided by the World Development Indicators dataset produced by the World Bank (Teorell et al., 2018).

To account for levels of taxation and natural resources I have included, tax revenue in % of GDP, non-tax revenue and total resource revenue respectfully using the variables **wdi_taxrev**, **ictd_nontax** and **ictd_revres** in the dataset. Tax revenue in % of GDP is the total amount of revenue received by the state from taxation in the given year, measured as a percentage of its total GDP. Non-tax revenue comprises of all revenue received by the state in the given year categorized as non-tax or other. Total resource revenue includes all natural resource revenue accrued by the state in the given year and includes both tax and non-tax revenue from sources including a significant portion of economic rent. While all three variables are collected from the QOG dataset, **wdi_taxrev** originally stems from the World Development Indicators dataset provided by the World Bank while **ictd_nontax** and **ictd_revres** from the

Government Revenue Dataset provided by the International Centre for Tax and Development (Teorell et al., 2018).

4.3.2 Demography

To test for the effect of population size found by Cunningham and Lemke (2014, p. 336) the variable **wdi_pop**, measuring total population, is included in the analysis. Additionally, **wdi_popurb**, the percentage of the population living in urban areas, and **wdi_popden**, population density per square kilometre, are included as control variables. The data for these variables are sourced from the QOG dataset, with original data being provided by the World Bank in the dataset World Development Indicators (Teorell et al., 2018).

As a measure of ethnic fractionalization, I have included the variable **fe_etfra** in the dataset. **fe_etfra** uses Fearon's identification of 822 ethnic and "ethnoreligious" groups in 160 countries and reflects the probability that two individuals, randomly chosen from the state in question, will belong to the same ethnic or "ethnoreligious" group, with 0 signifying a perfectly homogenous population and 1 signifying a highly fragmented population in terms of ethnicity. The variable is located at the state level and is assumed to be a constant over the time period. The source of the data is the QOG dataset originally provided by Fearon (Teorell et al., 2018, p. 280).

4.3.3 Geography

To account for the effects of capacity decay, distance both to the capitol as well as to major cities are included. The included variables come from the PRIO-GRID dataset and measure both distance and travel time. PRIO-GRID uses geographical analysis, dividing the world map into pixels within cells. The variable **ttime** measure travel time in minutes from the pixel to the nearest major city, defined as a city with a population of 50.000 or more. The variable is state level mean generated from the means of each cell. In effect this is a measure of rurality, where a high score indicates an overall long distance to major cities. To also test for the effects of distances from the capitol I have included the variable **capdist**, measuring the state mean of distance, in kilometres, from the centroid of each pixel. The data for the construction of these variables is collected from the PRIO-GRID dataset which uses Uchida and Nelson (2009) as the source for travel time and Weidmann, Kuse, and Gleditsch (2010) as the source for distance (Tollefsen et al., 2012).

To control for the spill-over effects of international borders on stability, either through cross-border communities (Brock et al., 2012) or "safe havens" (P. Collier, 2007, p. 31), the variable **bdist,** measuring the state mean of the distance in kilometres from the cell centroid to the nearest land-contiguous international border, is included. The reasoning for this variable lies in the concept of spill-over, as much spill-over is assumed to be direct effects of border crossings the average distance from international borders plays a role as it becomes reasonable to assume a low average distance from international borders increase the potential impact of spill-over. This variable is constructed using data from the PRIO-GRID dataset, with original data provided by Weidmann et al. (2010) (Tollefsen et al., 2012).

To also control for mountainous terrain I have added this as a separate variable, **mountains**, which is an aggregate variable representing the state mean of the variable mountain_mean from the PRIO-GRID dataset which measures the proportion of mountainous terrain in each cell (Tollefsen et al., 2012).

To test for the spill-over effect of "bad neighbours", I have included the dichotomous variable **FailedNeighbour** which displays whether or not a state has at least one neighbouring state with high levels of fragility. This variable was generated using a maximum FIS score of 90 as a limit. In practicality based on the variable FailedNeighbourMax90, which is one of four variables counting the number of fragile neighbours each state has. The others being **FailedNeighboursMaxFSI100**, **FailedNeighboursMeanFSI100** and **FailedNeighboursMeanFSI90.** During model revision these five variables were tested before **FailedNeighbour** was selected.

To test Colliers findings that landlockedness plays a part in fragilty, I have included the dummy variables **Landlocked** and **SemiLandlocked** controlling for whether a state is landlocked or not. In **Landlocked** states who have access to the ocean is coded as 0 = not landlocked and states without access to the ocean as 1 = landlocked. In **SemiLandlocked** states with access only to large inland bodies of water, Lake Victoria and the Caspian Sea, are coded as 1 = semi-landlocked, all others are coded as 0 = not semi-landlocked. The variables have been manually constructed using the world map as provided by Google Maps.

4.3.4 Regimes

To test the effect of unstable regimes and frequent regime changes, I have included the variable **bmr_demdur** which measures the number of consecutive years of the current regime type. The data is collected from the QOG dataset, and originally provided by Boix, Miller and Rosato (2013) (Teorell et al., 2018).

As democracies are generally assumed to be less prone to many of the issues associated with fragility I constructed the dichotomous variable **Democracy** from the **ht_regtype** variable for use as a control variable in the analysis. The variable is coded as 0 = non-democratic state and 1 = democratic state.

4.4 Identifying variables

As shown in Table 5 the dataset includes seven identifying variables, geographic region (**Region**), continent (**Continent**), Country Code (**ccode** and **ccodecow**), Country Name (**cname**), year (**year**) and country code and year (**ccodealp_year**). The sources for these are the QOG dataset (Teorell et al., 2018) along with geographical data from the CIA World Factbook (Central Intelligence Agency, 2017) which has been manually added to the dataset. The inclusion of an additional country code variable is based in the need for a match variable to merge data from the QOG and PRIO-GRID datasets. As both the QOG and PRIO-GRID datasets contain the Correlates of War country codes this variable was needed to facilitate the merge. The variable **ccode** is still used as the main identifier, as it is based on the ISO-3166-1 standard and recommended for usage with the QOG dataset (Teorell et al., 2018).

To enable grouping of states based on type of political regime I have included the categorical variable **ht_regtype** which registers the type of regime in the state during the given year, the variable categorizes states in 20 different regime types. I have also added the collapsed and simplified version of the variable, named **ht_regtype1**, where the number of different regime types have been reduced to a total of 7, including other. The data has been collected from the QOG dataset, and is originally provided by Hadenius & Teorell (Teorell et al., 2018).

Additionally, the variables **FSICategory**, **FSIMeanCategory** and **FSIMeanCategory2** were constructed from the values of **ffp_fsi**. **FSICategory** provides the yearly categorization of each state according to that years **ffp_fsi** score using the categories provided by the index itself. **FSIMeanCategory** and **FSIMeanCategory2** are temporally fixed categorizations using the mean of **ffp_fsi**. While **FSIMeanCategory** follows the same categories as **FSICategory**, **FSIMeanCategory2** is more fine-grained splitting the index into the categories displayed in Table 7.

Table 7 - Grouping on mean FSI - FSIMeanCategory2

FSIMeanCategory2	**Mean FSI range:**	**Number of states:**
1	0 – 29,9	13
2	30 – 39,9	12
3	40 – 49,9	16
4	50 – 59,9	10
5	60 – 69,9	16
6	70 – 79,9	38
7	80 – 89,9	36
8	90 – 99,9	22
9	100 – 120	13
Missing	n/a	17

4.5 Issues of validity

State fragility, and failure particularly, can be an impediment to data validity. Poor states with lacking state capacity often face issues connected with accurate measurements of societal factors connected both with economy, such as inequality, employment, growth and business, and demography, such as urbanization, ethnic fractionalization and demographic indicators (Kalyvas, 2007, p. 418). While this is not something that can be addressed statistically, it remains important to keep in mind as the analysis progresses. In this kind of research there is an ever-present issue of omitted variables, while longitudinal models are able to mitigate the issue (Frees, 2004, p. 9), it remains an issue that must be considered when discussing validity of the model.

4.6 Data transformation and clean-up

To provide best fit, clear easily interpreted data and prevent an unnecessarily large dataset, transformation and clean-up of the data.

4.6.1 Dropped data

While the original QOG dataset starts in the year 1946, the explanatory variables used have no data before 1960 when the first observations of **wdi_pop** and **wdi_gdpcapcur** are added. This only applies to the variables on the occasion level, those changing from year to year, as the variables on the subject level, those only varying between states, are fixed in time and identical for all years. As shown in Table 8 below the explanatory variables are introduced in the period 1960 to 1980. The dependent variable, **ffp_fsi**, is only introduced in 2005, the first year the index was published.

Table 8 - Variables with start year

Variable:	**Start year:**	**End year:**	**Observations:**
ffp_fsi	2005	2016	2085
wdi_incsh10h	1979	2015	1258
wdi_pop	1960	2016	8520
FailedNeighbour	1946	2017	12559
Landlocked	1946	2017	12559
wdi_gdpcapcur	1960	2016	7714
wdi_gdpcapgr	1961	2016	7491
fe_etfra	1946	2017	11009
bmr_demdur	1946	2010	8389
bidst	1946	2017	10543
capdist	1946	2017	12559
ttime	1946	2017	12559
ictd_revres	1980	2015	2512
ictd_nontax	1980	2015	4579
wdi_taxrev	1972	2016	3617

Democracy	1972	2014	6838
mountains	1946	2017	12119

To minimize the amount of unusable data in the dataset, the years prior to 1960 were dropped from the dataset.

In addition to this some minor adjustments were made to the dataset. Defunct states and states not assessed by the Fragile States Index were dropped from the dataset. While the latter primarily pertains to very small states such as Andorra, St. Lucia and Tonga, one case stands out. Israel is only assessed by the Fragile States Index together with the West Bank, while other data available only covers Israel. Because of this mismatch between the data, Israel, and the West Bank, were dropped from the dataset. Other minor adjustments the renaming of a few variables to provide shorter and clearer naming and a renaming of some states to remove start/end years for specific version of the state. An example of this is the renaming of "France 1963-" to "France". The old naming indicating that the France in question is without Algeria which achieves independence in 1962 and thus is included in the data for "France -1962". As this thesis primarily concerns itself with contemporary states, all such names have been replaced. For a full overview please consult the attached .do file.

4.6.2 Centering

When interpreting multilevel models using natural metrics for the explanatory variable, β_{0i} is interpreted to be the mean response for the given subject, when the value of the explanatory variables is zero (z = 0). However, in many cases, such with the income share held by the top 10%, a zero value falls outside the range of meaningfulness as zero is not a naturally occurring value for the variable. A value of zero on the variable "income share held by the top decile" is both theoretically and practically impossible, as it would entail the top 10% of earners earning nothing at all. To solve this issue, we can centre the explanatory variable, either on the overall grand mean or the group mean replacing the explanatory variable with $z_{ij} - \bar{z}_i$ or $z_{ij} - \bar{z}_j$ respectively. This changes the interpretation of β_{0i} and the variances of intercept and slope. β_{0i} to the expected response for the given subject when the explanatory variable is mean. And of the variances to the expected variance when the explanatory variable is mean(Frees, 2004, p. 169; Hox, 2002, p. 56).

Practically the centered variable, **centered_incsh10h**, was generated by subtracting the grand mean from the value of **wdi_incsh10h**[4]. As a result, **centered_incsh10h** replaces **wdi_incsh10h** in the analysis.

4.6.3 Rescaling

To improve the readability of the coefficients and observations, the population variable, **wdi_pop**, was rescaled to population in thousands. Practically this was done by generating a new variable, **wdi_pop1k**, observations of which are the observations of **wdi_pop** divided by 1000[5].

4.6.4 Non-linear transformation

For variables not adhering to the conditions of additivity and linearity, a logarithmic transformation can remedy the issue. In general, it makes sense to log-transform variables where all responses are positive. As natural logarithms allow for an interpretation of coefficients as approximate proportional differences natural logarithms are applied (Gelman & Hill, 2007, pp. 59-60). As presented in chapter 6, log-transformed variables do not replace the original variables in the full model, rather a separate estimation is run using log-transformed variables where appropriate. The full model and the model containing log-transformed variables are presented together to allow for comparison between the two.

To identify variables suitable for transformation, the g*ladder* test in STATA was used. This graphically displays up to nine different non-linear transformations of the given variable, comparing them to a normal distribution (Midtbø, 2012, p. 71). As log-transformation is generally not used on percentage measures, variables given in percentages are not tested. Along the same lines, variables with observations at zero is also exclude as log-transformation is not applicable to this kind of variable (Midtbø, 2012, p. 134). As log-transformation is all about the distribution of the variable responses, dichotomous variables are also excluded from testing. In addition to the graphical *gladder* test, the statistical *ladder*

[4] STATA command: *gen centered_incsh10h = wdi_incsh10h - r(mean)*

[5] STATA command: *generate wdi_pop1k = wdi_pop /1000*

test was used on borderline cases. This test employs a Skewness-Kurtosis test, applying it to the same nine non-linear transformation techniques as gladder.

Table 9 below, displays the selected variables and their log-transformed versions.

Table 9 - Log-transformed variables

Original variable:	Log-transformed version:
wdi_pop	ln_pop
wdi_gdpcapcur	ln_gdpcapcur
bmr_demdur	ln_demdur
bdist	ln_bdist
capdist	ln_capdist
ttime	ln_ttime

4.7 Criticism

In a constructivist vein it is possible to critique the concept of ethnic conflict, and as such the effects of ethnic fractionalization, based in an understanding that the concept in general is flawed and that ethnicity is a construct and as such fluctuant (Kalyvas, 2007, p. 420).

4.8 Alternative variables

As with most statistical models there are variables left of the analysis. Below I go through a few central variables that, while interesting are not included.

A dichotomous measure of whether the state is at war, intrastate or civil, in a given year could allow for control of the direct effects of war upon state fragility. However, as the definitions of war, especially civil war and internal conflict, are vague and the impact on the state differs strongly from case to case, the war variable was left out of the analysis. Additionally, quite a few effects of war are directly included in the Fragile States Index. In the same vein, spill-over effects of war in neighbouring states are not included as this suffers from the same issues as a measure of war. Another interesting approach to war could have been the inclusion of peace years, the number of years since last war in the state. However, this again returns us to the issues described above.

Another interesting variable that could have been included as a control is whether polygamy, particularly polygyny, the taking of multiple wives, is widespread in the state. The argument, echoing Collier's (2007) findings on the effect of poverty and hopelessness among young men, is that polygyny reduces the availability of wives for the poorer men in a society decreasing opportunity costs for destabilizing behaviour, especially among young men. The issue is closely related to both inequality and state capacity, if real the expected effect of polygyny would likely be more prominent where inequality is higher. As such this could make for an interesting addition to the analysis. However, as the issue is complex and faces some challenges of operationalization it was not included in the analysis.

An alternative to the Fragile States Index is available in the OECD measures of sate fragility. The OECD measures fragility along five dimensions, economic, environmental, political, security and societal (OECD, 2016, p. 23). While the OECD fragility publications offer a valuable insight into state fragility, it lacks some of the statistical qualities of the Fragile States Index data. As a result, the Fragile States Index was selected as the measure of fragility in this thesis.

4.9 Data availability

As discussed earlier, the availability of reliable data is an issue, especially when dealing with fragile states. Access to more reliable data from these states would have increased the validity of the results, as well as opening up the opportunity for the inclusion of additional explanatory and control variables increasing the explanatory power of the model. As well as being generally desirable for the validity of the estimations made by the regression analysis, additional data would enable more detailed analysis of specific regions or categories of states, providing a deeper understanding of the issue and allow for testing of Collier's (2007, pp. 56-57) theories of the effects of landlocked ness in African states.

5 Descriptive statistics

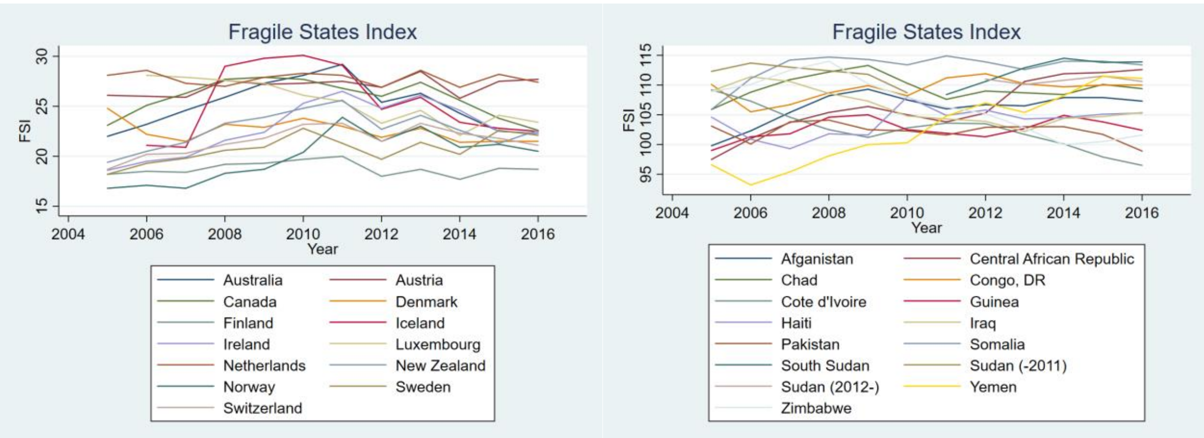
In the following chapter I explore the data and variables graphically.

The use of longitudinal data and polynomial growth curve models rests squarely on two base features of the data, change over time and variation between states. To illuminate these features of the data, graphs showing both the temporal development as well as the variation between states of key variables are provided in this chapter. For the sake of brevity and readability only the states comprising the top and bottom groups as categorized by the variable **FSIMeanCategory2** are displayed. For a more detailed overview the development of the variables please consult the appendix.

5.1 Fragile States Index

As the Fragile States Index annually measure state fragility across a variety of aspects, change in time and variation between states is expected. To examine the actual features of the data, the state’s index scores are plotted against years in the graphs below.

Figure 3 - Fragile States Index over time in top and bottom states

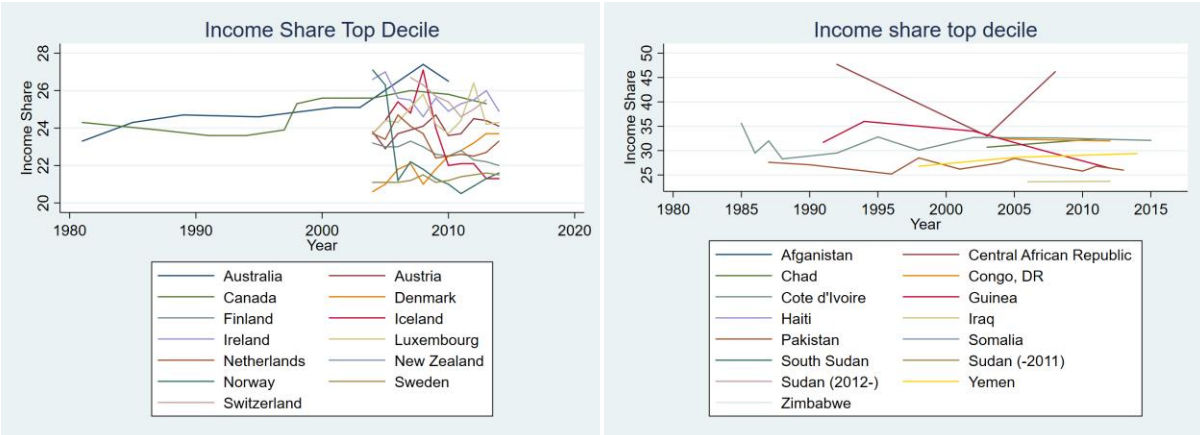


The graphs in Figure 3 show both a variation over time as well as a substantial variation between states. While some of these states exhibit relatively small changes in scores over time, other such as Yemen and Iceland show quite a substantial change in either end of the spectrum.

5.2 Income share top decile

For inequality to play a part in the variation observed in the Fragile States Index scores shown above, inequality too must vary both in time and between states. Figure 4 graphs this change.

Figure 4 - Income Share Top Decile over time in top and bottom states

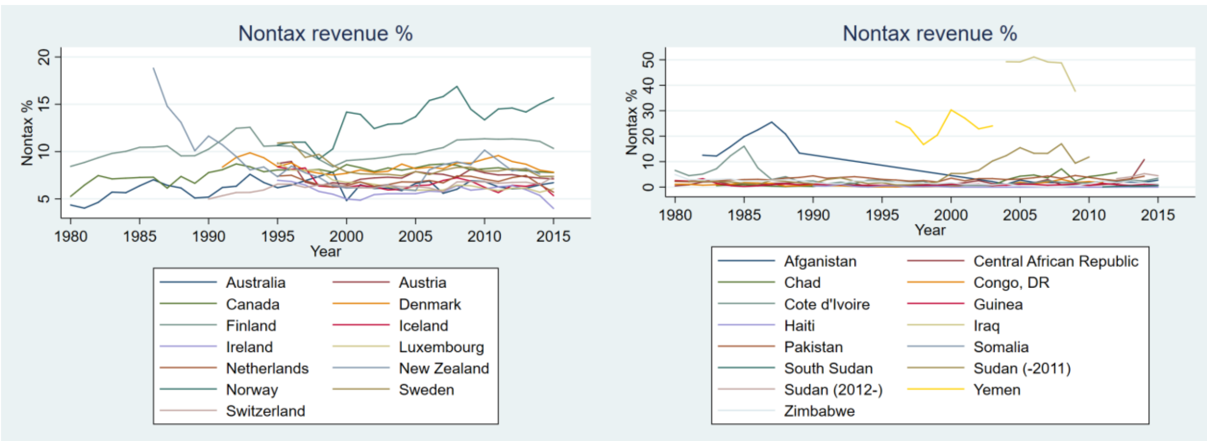


As the graph shows there is clear variation both in time and between states. Additionally, the graphs show variation in the availability of data.

5.3 Non-tax revenue

Figure 5 blow, graphs the changes in Nontax revenue over time.

Figure 5 - Non-tax revenue over time in top and bottom states

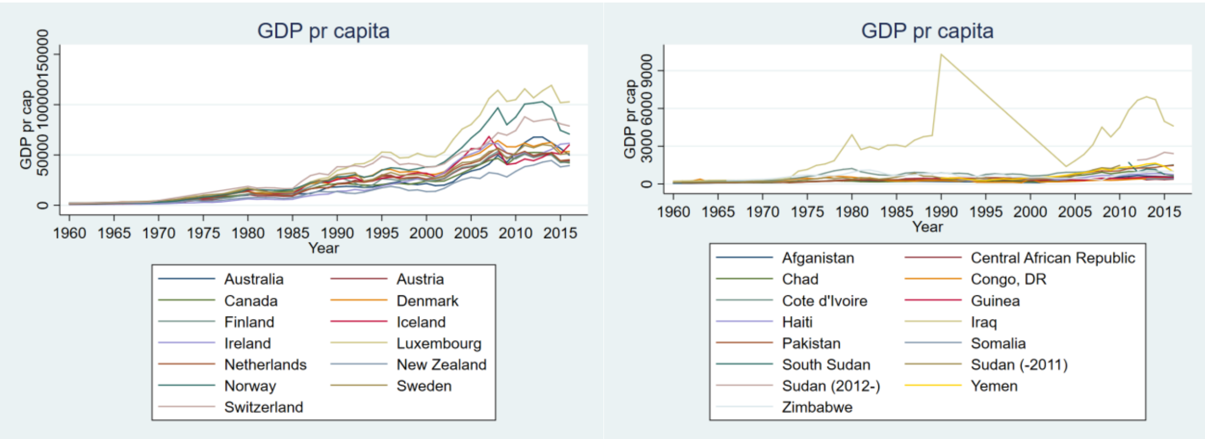


Consolidated non-tax revenue, measured as a percentage of GDP also has a few standouts alongside some issues of missing data. Among the states making up the lower end of the Fragile States Index, Norway and New Zealand stand out, with Norway's non-tax revenue increasing well beyond their counterparts while New Zealand's falls closer to the group average as time progresses. In the other group, we see a Yemen and Iraq standing out from the pack while there at the same time is a lack of data for both these states.

5.4 GDP pr. capita

GDP pr. capita is graphed against years in Figure 6. As expected there appears to be an upwards trend in the data, indicating an overall accumulative growth.

Figure 6 - GDP pr. capita over time in top and bottom states



Reviewing the graphs, most states appear to follow each other quite closely in the growth of GDP pr. capita, a likely by-product of accumulative growth, the international economy and the fact that GDP here is measured in US dollars. A stand-out from this trend is Iraq, who's graph is both much higher and irregular compared to its cohorts. Likely a result of oil revenue increasing GDP and shifting periods of war and relative peace.

6 Results

This chapter presents the results of the statistical analysis described in earlier chapters. Following a revisiting of the hypotheses, the focus of the chapter turns to the modelling process, before moving on to the main results of the analysis. A presentation and discussion of alternative models featuring log-transformed variables and an alternative main explanatory variable closes off the chapter.

When developing a growth curve model, many model versions are run and tested. Far from all of these will be presented in the chapter below. For details on the modelling process please consult the attached .do file.

When examining the results of the regression analysis it is important to keep in mind how the dependant variable is constructed. The Fragile States Index rates fragility in such a way that a higher score equals a more fragile state. As a result, positive coefficients from the regression indicates a relationship where the variable in question increases fragility while negative coefficients indicates a reduction of fragility.

6.1 Hypotheses

Before going into the analysis proper, it is beneficial to revisit the hypotheses that are being tested. As described in chapter 2, the hypotheses derived from the overall research question are:

H_1: Increased economic inequality increases state fragility.

H_2: Increased GDP pr. capita lowers the impact of inequality on fragility.

H_3: Spill-over effects increases the impact of inequality on fragility.

H_4: Non-tax revenue increases the impact of inequality on fragility.

6.2 Trend trajectories

As the first part of the model development, to determine the shape of the growth curve, different degree polynomials were added to both the fixed and random parts of the model. These initial models determining the shape of the growth curve contains no independent variables. Only the dependent variable and different degree polynomial trend terms. Model 1 consists of a linear trend variable in the fixed part of the model. Model 2 a linear and quadratic trend variable. Model 3 a linear, quadratic and cubic trend variable. And Model 4 a linear, quadratic, cubic and quartic trend variable. Following the LR and BIC testing of these four models, model 3 was selected based on test results. As a result, all following models contain the linear, quadratic and cubic trend variables in the fixed part. Building on model 3, model 5 introduces the linear trend variable in the random part. Model 6 expands on this, adding the quadratic trend variable. Following LR and BIC testing Model 6, featuring 3. degree polynomial in the fixed part and 2. degree polynomial in the random part was selected.

Table 10 - Trend variables

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
<u>Fixed effects:</u>							
Linear trend	-.1074578 *** (.0204828)	.3354707 *** (.0769536)	1.129007 *** (.1775887)	1.308122 *** (.3248228)	.9854316 *** (.1098188)	.9612257 *** (.1097693)	.9630679 *** (.105425)
Quadratic trend		-.0397265 *** (.0066577)	-.2243194 *** (.0378679)	-.3032309 ** (.1256692)	-.1977933 *** (.0200587)	-.1935684 *** (.0178807)	-.1938509 *** (.0174959)
Cubic trend			.0110836 *** (.0022388)	.0224289 (.0173729)	.0096832 *** (.0011834)	.0094545 *** (.0010103)	.0094703 *** (.0010365)
Quartic trend				-.000513 (.000779)			
Intercept	71.62973 *** (1.758024)	70.85674 *** (1.762914)	70.24333 *** (1.767346)	70.17508 *** (1.770385)	70.41357 *** (1.746514)	70.44858 *** (1.758649)	70.44545 *** (1.758028)
<u>Random effects:</u>							
Random linear trend					.5605121 (.0617449)	.9652027 (.1273509)	.8044624 (.1090051)
Random quadratic trend						.0043912 (.0006446)	.0021344 (.0007885)
Random cubic trend							9.86e-06 (2.93e-06)
_cons	546.9161 (58.06787)	547.0526 (58.0806)	547.1832 (58.09328)	547.1864 (58.0936)	540.3128 (57.39086)	548.4671 (58.27724)	548.08 (58.23311)
var(residual)	10.07527 (.3262857)	9.8904 (.3202989)	9.764698 (.316228)	9.762473 (.3161559)	2.687417 (.0913953)	1.927994 (.072664)	1.928526 (.0735546)
<u>Diagnostics:</u>							
BIC	11912.59	11884.95	11868.24	11875.45	10009.35	9798.72	9793.082
n	2085	2085	2085	2085	2085	2085	2085

* p < 0.10 (90% significance), ** p < 0.05 (95% significance), *** < 0.01 (99% significance)

Standard errors in parentheses.

BIC: Bayesian Information Criterium

As Table 10 shows the model was expanded gradually, from the linear fixed effects trend only in Model 1 to four degrees in model 4. The BIC score falls with the addition of the quadratic and cubic trends added in Model 2 and Model 3 but increases slightly with the addition of the quartic trend in Model 4. Based on this the quartic trend variable and the whole of Model 4 was rejected. Moving forward in model 5, a linear random effects trend is added to the model consisting of 3 trend degrees in the fixed part. This too shows a drop in BIC indicating a better model fit and increased explanatory power. In model 6 and 7 quadratic and cubic random trends are added to the model. As BIC drops for both Model 7 was initially selected for further development. At a later stage of the modelling process, models based on Model 7 fail to estimate when adding lagged variables and interactions. Causing a need for revision of the initial model. Based on the estimation issues, and the small difference in BIC scores between models six and seven, the cubic trend was dropped from the random part of the model. Allowing for more complex modelling in the remainder of the model. This is also in line with the recommendations of Rabe-Hesketh and Skrondal (2012, p. 348) to include fewer degrees in the random part . In addition to the BIC tests cited in the table, LR-tests were run on the different models, the results of the LR-tests are in line with the BIC scores. As for the estimations of the models, the estimated coefficients for both intercept and all three degrees of trend trajectories in the fixed part of the model are significant at a 99% significance level for models 1-3 and 5-7. This is in line with the BIC tests and further underpins the decision to drop the quartic trend trajectory added in model 4.

Following the fitting of the growth curve trajectory, covariance structures were tested on model 6 using BIC testing. Here an unstructured covariance structure provided the best fit and was selected for use in the model. As a result, the base model best describing the development of the Fragile States Index scores is a model with 3rd degree polynomial fixed effects part, a 2nd degree polynomial random effects part and an unstructured covariance structure. This is the model that will be used as a base when testing the hypotheses presented above.

As described in chapter 3, following the fitting of the growth curve, explanatory, control and lagged variables, as well as interactions, were added and tested using LR and BIC. The full process is described in chapter 3 and documented in the accompanying .do file.

6.3 Full model

Following the modelling procedure, the full model was constructed based on the theory presented in chapter 2. In Table 11 below, the results of the full model are presented alongside the results from two intermediary models, the explanatory variables model, featuring explanatory variables only, and the interactions model adding interactions to the explanatory variables model. In contrast the full model includes both explanatory, and lagged variables as well as interactions. For the full regression outputs, see STATA Outputs in the appendix. Both the explanatory variables model and interactions model have n=172 observations spread across 47 states with a minimum of 1 observation each. The average number of observations pr. state is 3.7 and the maximum 6. In the full model, following the addition of lagged variables, the number of observations falls to n=102, spread across 23 states, with a minimum number of observations at 2, the average at 4.4 and the max at 6.

Table 11 – Results

	Explanatory variables model	**Interactions model**	**Full model**
<u>Fixed effects:</u>			
Linear trend	-.0770973 (.5097671)	-.146749 (.5068802)	1.928285 ** (.9800925)
Quadratic trend	.5032696 ** (.2377048)	.5855837 ** (.2467556)	-.3431564 (.3483858)
Cubic trend	-.0823126 *** (.031673)	-.0959585 *** (.0332564)	.0318527 (.0399913)
Income share top decile	.0075708 (.1077948)	.3235578 * (.1702323)	.3741926 ** (.188678)
Population (in thousands)	.000123 ** (.00005)	.0001141 ** (.0000473)	-.000068 (.000099)
Failed neighbour	25.27155 *** (4.96883)	24.05854 *** (4.733728)	21.98966 ** (9.526424)
Landlocked	5.754738 (4.965312)	6.468208 (4.691896)	8.443465 ** (4.139489)
GDP pr. capita	-.0000711 * (.0000368)	-.000115 ** (.0000574)	-.0000481 (.0000687)
GDP pr. capita growth	.0049161 (.0295593)	.0150822 (.0305649)	-.0804998 *** (.0227979)
Ethnic fractionalization	34.82202 *** (9.398784)	35.04359 ** (8.90606)	16.50858 * (9.575367)
Regime type duration	-.0763407 ** (.0367321)	-.0640925 * (.0344984)	-.2382053 *** (.0396314)
Distance from border	-.0114024 (.0240297)	-.010151 (.0227382)	.0098883 (.0232227)
Distance from capitol	-.009091 (.0079795)	-.0096996 (.0075137)	.0010394 (.0113202)
Travel time to major city	.0099802 (.0067866)	.0103156 (.0064409)	-.0008107 (.0169031)
Resource Revenue	-.1720665 (.1053461)	-.1370605 (.1076882)	.1159606 (.1688722)

Non-tax revenue	.0051002 (.1316644)	-.1245954 (.1418104)	-.3994362 (.2468245)
Tax revenue (% of GDP)	.2683068* (.1455813)	.3206809 ** (.1442682)	.3436291 ** (.1501637)
_cons	32.83917*** (5.501386)	33.47396 *** (5.383177)	56.46234 *** (7.407887)
Interactions:			
Non-tax revenue # Income share top decile		-.0250353 (.0210897)	.0222502 (.0369326)
GDP pr. capita # Income share top decile		-2.19e-06 (5.78e-06)	-8.61e-06 (7.86e-06)
Border distance # Income share top decile		-.0000544 (.000579)	-.0015735 (.0014647)
Failed neighbour # Income share top decile		-.6265954 * (.3426035)	2.320665 (1.624023)
Lagged variables:			
L3. GDP pr. capita			-.2805135 *** (.0751018)
L1. Income share top decile			-.0840067 (.1406988)
L2. Income share top decile			-.0258027 (.1174169)
L3. Resource Revenue			.7131551 *** (.1855323)
L1. Non-tax revenue			-.655918 *** (.1734393)
L2. Non-tax revenue			-.4350106 * (.2429487)
L1. Tax revenue (% of GDP)			-.1390894 (.146688)
L2. Tax revenue (% of GDP)			.1277729 (.1800577)
L3. Tax revenue (% of GDP)			-.3260539 * (.1911574)
Random effect parameters:			
var(trend)	1.632974 (.6363507)	1.249625 (.569088)	4.138205 (1.70848)
var(trend2)	.0167638 (.0124269)	.0080046 (.0128911)	.0718679 (.034422)
var(_cons)	232.1599 (59.86031)	220.0053 (56.37419)	72.82532 (34.74317)
cov(trend,trend2)	-.1319615 (.0809709)	-.0773901 (.0775605)	-.4822094 (.2316434)
cov(trend,_cons)	-6.700924 (6.365739)	-6.731165 (5.524073)	.7303084 (7.198228)
cov(trend2,_cons)	.0329932 (.8587694)	-.0315277 (.7272192)	-.9480457 (1.010032)

var(Residual)	.6921383 (.1386173)	.753522 (.1652659)	.1977903 (.0561848)
Diagnostics:			
BIC	965.2132	978.5682	565.0599
n	172	172	102

* $p < 0.10$ (90% significance), ** $p < 0.05$ (95% significance), *** < 0.01 (99% significance)
Standard errors in parentheses.
BIC: Bayesian Information Criterium

At the bottom of Table 11 the BIC scores for each of the three models is displayed. Expectedly, as it is the result of a modelling process based on BIC and LR-tests, the full model reports a BIC significantly lower than the other two alternatives.

6.3.1 Hypothesis H_1

Turning once again to the hypotheses developed in chapter 2, the main coefficient of interest is the one associated with income share of the top decile. This coefficient is not significant at any level in the explanatory variables model but is significant at 90% significance level in the interactions model and, more importantly, significant at 95% in the full model. As shown in the STATA outputs in the appendix the p-values reported by the latter two is 0.057 and 0.047 respectively. As the full model provided a better overall fit to the model, this strengthens hypothesis H_1.

I addition to a significant relationship, the hypothesis also specifies a positive relationship. The coefficient for income share of the top decile is 0.3741926 indicating a positive effect of inequality on fragility. As described in chapter 4 the centered version of the variable describing income share held by the top decile, **centered_incsh10h**, was used in the analysis. This changes the interpretation of the coefficient as compared to that of the original variable. Thus, the coefficient provided above is interpreted as the change in Fragile States Index score for each additional percentage of income befalling the top decile, above the mean. Interestingly, the one- and two-year lags included in the model are not significant.

Given the results presented above, hypothesis H_1 is strengthened. I return to a discussion of the results in chapter 7.

6.3.2 Hypothesis H_2

In practicality what hypothesis H_2 asks for is a significant, negative coefficient of the interaction between GDP pr. capita and inequality. While the coefficient is negative in line with the hypothesis, it is not significant. The p-value for the coefficient is 0.273 far above the value associated with even a 90% significance level. Because of the lack of significance H_2 is rejected by the full model.

6.3.3 Hypothesis H_3

Hypothesis H_3 asks for a positive effect of "spill-over effects" between states on the effect of inequality. Two variables in the model are associated with spill-over effects, **bdist,** measuring the mean distance from international border across the state, and **FailedNeighbour**, a dichotomous variable depicting whether the state in question has a neighbouring state with a mean Fragile States Index score of at least 90. In effect these two variables measure two different aspects of spill-over, **bdist** the overall proximity to international borders and **FailedNeighbour** whether a neighbouring state is in in the highest category of fragility. The effect of these variables on the effect of inequality is estimated by the interaction effects between them and **centered_incsh10h**.

When examining the results presented in Table 11 as well as the full STATA outputs presented in the appendix, neither the interaction between border distance and inequality nor the interaction between failed neighbours and inequality prove significant in the full model. In the interactions model the interaction between failed neighbours and inequality is significant but only at a 90% significance level. This results in a rejection of hypothesis H_3.

6.3.4 Hypothesis H_4

Hypothesis H_4 asks for a significant positive impact of non-tax revenue, measured by the variable **ictd_nontax**, on the effect of inequality on fragility. The interaction between the two yields no significant coefficients at any of the models including interaction effects. As a result, the hypothesis is rejected.

6.3.5 Other results of interest

The trend variables, representing the shape of the regression line, differ in significance across the three models. While the cubic and quadratic trend variables are significant at 95% and 99% respectively in the interactions and explanatory variables models, these are no longer significant in the full model. Here, however the linear trend variable is significant at a 95% level.

Perhaps as no surprise, the presence of neighbouring states has a significant positive impact on fragility scores. The **FailedNeighbours** variable is significant at at least a 95% level in all three models. Landlockedness also achieves a significance level of 95%, but only in the full model. The coefficient of **Landlocked** is positive, indicating a positive relationship between landlockedness and Fragile States Index scores.

Among the economic variables, GDP pr. capita, **wdi_gdpcapcurr**, is not significant in the full mode, but at 95% in the interactions model and 90% in the explanatory variables model, while GDP pr. capita growth, **wdi_gdpcapgr** is significant at a 99% level in the full model but not significant in the other two. However, at a three-year lag GDP pr. capita is significant at a 99% level in the full model. Both variables have negative coefficients, indicating that if significant, the relationship between them and Fragile States Index scores are negative. Tax revenue, **wdi_taxrev**, is significant at 95% in both the full and interactions models, while significant at 90% in the explanatory variables model. Tax revenue has a positive coefficient across all three models, indicating a positive relationship between these variables and scores on the Fragile States Index. Resource revenue, **ictd_revres**, on the other hand is not significant at any level across the three models. Interestingly both these variables are significant at a three-year lag in the full model. Resource revenue at 99% and tax revenue at 90%. Additionally, non-tax revenue, **ictd_nontax**, while not significant at no-lag is significant at both one- and two-year lags. 99% level at a one-year lag and 90% at a two-year lag. The coefficients of the significant lagged variables are all, except for resource revenue, negative indicating that while resource revenue at a three-year lag positively impacts Fragile States Index scores, non-tax revenue, tax revenue, and GDP pr. capita has a negative impact.

Ethnic fractionalization, **fe_etfra**, proves significant at a 90% level in the full model, at a 95% level in the interactions level and at 99% in the explanatory variables model. The coefficient is positive across the board, a strong indication of a positive impact of ethnic fractionalization on scores of the Fragile States Index.

Regime type duration, **bmr_demdur**, is also significant in the full model. While significant at a 99% level in the full model, this variable is significant at a 90% level in the interactions model and at a 95% level in the explanatory variables model. Across all three models this variable has a negative impact on Fragile States Index scores.

Interestingly, population, **wdi_pop1k**, travel time to nearest major city, **ttime**, and distance to capitol, **capdist**, are not significant in any model.

6.4 Log-transformed variables

As presented in chapter 4, a separate model featuring logarithmically transformed variables was run. The results of this is presented alongside the main model below. The number of observations and their distribution is identical across the two models.

Table 12 - Results with log-transformed variables

	Log-transformed variables	**Full model**
<u>Fixed effects:</u>		
Linear trend	1.489382 (1.001176)	1.928285 ** (.9800925)
Quadratic trend	.28764 (.3861856)	-.3431564 (.3483858)
Cubic trend	-.0634164 (.0461435)	.0318527 (.0399913)
Income share top decile	.3139439 (.2031013)	.3741926 ** (.188678)
Population (in thousands)	1.367786 (1.792008)	-.000068 (.000099)
Failed neighbour	12.81147 * (7.238933)	21.98966 ** (9.526424)
Landlocked	3.40977 (3.497298)	8.443465 ** (4.139489)
GDP pr. capita	-7.383634 *** (1.354711)	-.0000481 (.0000687)
GDP pr. capita growth	.0223912 (.0306876)	-.0804998 *** (.0227979)
Ethnic fractionalization	9.385444 (7.55985)	16.50858 * (9.575367)
Regime type duration	-8.808515 *** (2.037408)	-.2382053 *** (.0396314)
Distance from border	3.018612 (2.496842)	.0098883 (.0232227)

Distance from capitol	-2.278563 (3.021866)	.0010394 (.0113202)
Travel time to major city	-.3342375 (3.288802)	-.0008107 (.0169031)
Resource Revenue	.1575932 (.178488)	.1159606 (.1688722)
Non-tax revenue	-.1925737 (.2511401)	-.3994362 (.2468245)
Tax revenue (% of GDP)	.380101 ** (.1558773)	.3436291 ** (.1501637)
_cons	129.0903 *** (21.63185)	56.46234 *** (7.407887)
Interactions:		Interactions:
Non-tax revenue # Income share top decile	.0256343 (.0386179)	.0222502 (.0369326)
GDP pr. capita # Income share top decile	-.00001 ** (4.53e-06)	-8.61e-06 (7.86e-06)
Border distance # Income share top decile	-.0002537 (.0015892)	-.0015735 (.0014647)
Failed neighbour # Income share top decile	.6502058 (1.726055)	2.320665 (1.624023)
Lagged variables:		
L3. GDP pr. capita	-.2100556 ** (.0814913)	-.2805135 *** (.0751018)
L1. Income share top decile	-.1303074 (.1265294)	-.0840067 (.1406988)
L2. Income share top decile	-.0525525 (.115101)	-.0258027 (.1174169)
L3. Resource Revenue	.6006354 *** (.1972839)	.7131551 *** (.1855323)
L1. Non-tax revenue	-.7664016 *** (.1833184)	-.655918 *** (.1734393)
L2. Non-tax revenue	-.1941205 (.2355465)	-.4350106 * (.2429487)
L1. Tax revenue (% of GDP)	.0306557 (.1602102)	-.1390894 (.146688)
L2. Tax revenue (% of GDP)	.1930527 (.194923)	.1277729 (.1800577)
L3. Tax revenue (% of GDP)	-.2983446 (.1999175)	-.3260539 * (.1911574)
Random effect parameters:		
var(trend)	3.54087 (2.037343)	4.138205 (1.70848)
var(trend2)	.0840775 (.0457524)	.0718679 (.034422)
var(_cons)	30.78507 (18.39587)	72.82532 (34.74317)
cov(trend,trend2)	-.5096061 (.3036842)	-.4822094 (.2316434)
cov(trend,_cons)	3.098535 (4.841813)	.7303084 (7.198228)
cov(trend2,_cons)	-.9273494 (.6945986)	-.9480457 (1.010032)

var(Residual)	.2837421 (.0938826)	.1977903 (.0561848)
<u>Diagnostics:</u>		
BIC	548.1559	565.0599
n	102	102

* p < 0.10 (90% significance), ** p < 0.05 (95% significance), *** < 0.01 (99% significance)
BIC: Bayesian Information Criterium

As Table 12 shows, the model including log-transformed variables differ somewhat from the full model. While the full model achieves a BIC score of 565, the log-transformed variables model scores 548 on the BIC test, indicating a somewhat better overall fit. As for the coefficients, the linear trend variable and key variable **wdi_incsh10h**, are no longer significant in the log-transformed variables model, the same goes for the variables representing GDP pr. capita growth, landlockedness and ethnic fractionalization as well as the two-year lag of non-tax revenue and the three-year lag tax revenue. In contrast, the GPD pr. capita variable as well as the interaction between GDP pr. capita and income share top decile are significant, at a 99% and 95% level respectively. Due to the log-transformation of the variable, the coefficient of GDP pr. capita changes significantly as compared to the natural number variable in the full model. For the remaining variables there are only minor changes in significance.

6.5 GINI model

While criticised and ultimately rejected as the main explanatory variable of this thesis, the GINI coefficient is a widely used measure of inequality. To allow for comparison a GINI version of the full model was run, wherein income share top decile was replaced as the dependant variable by the GINI coefficient. The two models have an identical amount of observations, as is the distributions of these.

Table 13 - GINI model results

	GINI model	Full model
Fixed effects:		
Linear trend	1.864931 * (.9815225)	1.928285 ** (.9800925)
Quadratic trend	-.3021587 (.3507807)	-.3431564 (.3483858)
Cubic trend	.0256982 (.0404334)	.0318527 (.0399913)
Inequality	.3255096 ** (.1654975)	.3741926 ** (.188678)
Population (in thousands)	-.0000428 (.0000962)	-.000068 (.000099)
Failed neighbour	20.66468 ** (9.235501)	21.98966 ** (9.526424)
Landlocked	8.295586 ** (4.183935)	8.443465 ** (4.139489)
GDP pr. capita	-.0000546 (.0000691)	-.0000481 (.0000687)
GDP pr. capita growth	-.0711954 *** (.0234127)	-.0804998 *** (.0227979)
Ethnic fractionalization	19.39251 ** (9.595578)	16.50858 * (9.575367)
Regime type duration	-.233715 *** (.0394216)	-.2382053 *** (.0396314)
Distance from border	.0104256 (.0232894)	.0098883 (.0232227)
Distance from capitol	-.001932 (.0112749)	.0010394 (.0113202)
Travel time to major city	.00323 (.0162819)	-.0008107 (.0169031)
Resource Revenue	.1103194 (.1721744)	.1159606 (.1688722)
Non-tax revenue	-.448773 * (.2527154)	-.3994362 (.2468245)
Tax revenue (% of GDP)	.4039765 *** (.1530543)	.3436291 ** (.1501637)
_cons	39.81925 *** (9.311909)	56.46234 *** (7.407887)
Interactions:		Interactions:
Non-tax revenue # Income share top decile	.0130837 (.0381624)	.0222502 (.0369326)
GDP pr. capita # Income share top decile	--8.54e-06 (7.93e-06)	-8.61e-06 (7.86e-06)
Border distance # Income share top decile	-.0012071 (.0014604)	-.0015735 (.0014647)
Failed neighbour # Income share top decile	1.872077 (1.612887)	2.320665 (1.624023)
Lagged variables:		
L3. GDP pr. capita	-.2703561 *** (.0763274)	-.2805135 *** (.0751018)

L1. Income share top decile	-.1267339 (.1422171)	-.0840067 (.1406988)
L2. Income share top decile	-.0110225 (.1187153)	-.0258027 (.1174169)
L3. Resource Revenue	.7310918 *** (.185538)	.7131551 *** (.1855323)
L1. Non-tax revenue	-.646664 *** (.1735226)	-.655918 *** (.1734393)
L2. Non-tax revenue	-.4156369 * (.2428502)	-.4350106 * (.2429487)
L1. Tax revenue (% of GDP)	-.1078877 (.1474745)	-.1390894 (.146688)
L2. Tax revenue (% of GDP)	.1504099 (.1800651)	.1277729 (.1800577)
L3. Tax revenue (% of GDP)	-.3090415 (.1916383)	-.3260539 * (.1911574)
Random effect parameters:		
var(trend)	4.09996 (1.701325)	4.138205 (1.70848)
var(trend2)	.0661717 (.0328821)	.0718679 (.034422)
var(_cons)	78.19514 (38.53897)	72.82532 (34.74317)
cov(trend,trend2)	-.4598959 (.2246973)	-.4822094 (.2316434)
cov(trend,_cons)	-1.767419 (7.506989)	.7303084 (7.198228)
cov(trend2,_cons)	-.6633122 (.9789529)	-.9480457 (1.010032)
var(Residual)	.2062745 (.0603633)	.1977903 (.0561848)
Diagnostics:		
BIC	565.7641	565.0599
n	102	102

* $p < 0.10$ (90% significance), ** $p < 0.05$ (95% significance), *** < 0.01 (99% significance)
BIC: Bayesian Information Criterium

As Table 13 shows, the two models are quite similar. The BIC scores are virtually identical, with a difference of only about 0.7. There are some minor differences in terms of significances, linear trend is only significant using a 90% level in the GINI model while being significant at 95% in the full model. Ethnic fractionalization also changes, from significant at a 90% level in the full model to a 95% in the GINI model, while tax revenue, moves from a 95% level in the full model to a 99% level in the GINI model. Of more interest perhaps, is the fact that non-tax revenue is significant at a 90% level in the GINI model, while not significant in the full model. Also, the three-year lag of tax revenue is no longer significant in the GINI model. All results pertaining to the hypotheses are unchanged in the GINI model, strengthening the arguments for the acceptance of H_1 and rejection of H_2, H_3 and H_4.

7 Analysis

In this chapter I discuss the findings presented in chapter 6. The first part of the chapter is devoted to hypothesis H_1, before hypotheses H_2, H_3 and H_4 are discussed. To round of the chapter the results outside the hypotheses are discussed.

As I return to in chapter 8, there is an issue with high autocorrelation among the residuals. As this is not accounted for in the main models this can have an influence on the significance testing, usually in the form of less significant results. While this is an overall issue, it calls some of the results falling just short of significance into question.

The significance of the trend variables changes between the models. While the cubic and quadratic trend variables are significant in the explanatory variables and interactions models, linear trend is instead significant in the full and following models. This indicates that as the model is expanded the trajectory best fitting the regression becomes closer to linear. A likely reason for this is that the curvilinear relationship included factors now accounted for by the explanatory and lagged variables, as well as the interactions.

7.1 The role of inequality in state failure

As shown in chapter 6 economic inequality, measured as income share held by the top decile, is both positive and significant at a 95% level in the full model as well as the log-transformed variables model. Additionally, the substituted inequality variable, GINI coefficient, is also significant at a 95% level in the GINI model. These results are strong indicators of a positive correlation between economic inequality and state fragility as measured by the Fragile States Index, pointing to an acceptance of the H_1 hypothesis. These findings are further substantiated by alternative model 1 presented in the appendix.

There is however, an issue with the addition of lagged variables in the full model, as the lags are introduced the number of observations falls from n = 172 in the explanatory variables model to n = 102 in the full model. Ideally a larger number of observations would have been preferable. The BIC scores for these three models show a stark decline with the addition of lagged variables, the full model scoring 565 as opposed to the 965 and 978 scored by the explanatory variables and interactions models respectively. This indicates a better fit and stronger explanatory power overall for the full model. While this can be considered to further substantiating the acceptance of the H_1 hypothesis it is possible that the lower BIC score is a result of the reduced number of states with observations. It can also be argued that the results are less generalizable because of the substantially lower number of states with observations.

Additionally, the results of Alternative model 2, where the Fragile States Index sub score associated with uneven economic development is subtracted from the overall score, indicate significance at only a 90% significance level.

Interestingly, the one- and two-year lags of income share top decile are not significant, indicating either an immediate relationship or more likely a temporal effect not sufficiently described by the included lags. The inclusion of multiple lagged versions of income share top of the decile could have shed some light on this relationship, however the addition of additional lags rendered the model unable to be estimated by STATA. A possible explanation for this is the model becoming too complex for the further reduced amount of available data.

7.2 GDP pr. capita

Contrary to both the consensus in the field (Cunningham & Lemke, 2014, p. 328) and the findings of P. Collier (2007, p. 19) the full model finds no significant relationship between state income and fragility as discussed in chapter 2. In its log transformed version in both the Log-transformed variables model and Alternative model 1 however, both GDP pr. capita and its interaction with inequality has a significant and negative coefficient. This indicates a correlation between country income, operationalized as GDP pr. capita, and state fragility. While the coefficient is small, the range of the variable must be considered. Ranging from 37.51812 to 119172.7 in natural numbers, the variable has a large range, this shows that while the coefficient might appear small in numbers it has a large potential impact.

As GDP pr. capita generally has non-linear growth, it appears sensible to transform the variable to achieve linearity. This is the reasoning behind the log-transformed variables model presented in Table 12 and Alternative model 1 presented in the appendix. The results from the models with the log-transformed version of GDP pr. capita lends credence to the conventional wisdom in the field and the findings of P. Collier (2007, p. 19) as well as providing support for the acceptance hypothesis H_2.

7.3 Spill-over

Hypothesis H_3 asks for a significant positive impact of spill-over effects on the effect of inequality on state fragility. The spill-over effects referred to in hypothesis H_3 is operationalized as two distinct variables in the analysis. The dichotomous **FailedNeighbours**, indicating whether a state has one or more failed neighbours, and **bidst**, measuring the mean distance from international borders across the state territory. As the two variables account for two different sides of the spill-over concept, they are run as separate interactions.

Neither of the two interaction effects has significant coefficients in any of the models run, indicating that spill-over effects, operationalized this way at the very least, does not impact the effect of inequality on state fragility. Based on these results hypothesis H_3 can be comfortably rejected.

It is however interesting to note that whether a state has failed neighbours have a significant positive coefficient in all models run, indicating that while spill-over effects appear not to impact the effect of inequality on state fragility, failed neighbours have an independent positive effect. This supports the general idea of spill-over effects presented in chapter 2, indicating a spill-over effect of fragility itself.

7.4 Nontax revenue

Building on Bates (2008, pp. 25-27) theories on the effects of taxes on fragility presented in chapter 2, hypothesis H_4 is concerned with the effect non-tax revenue has on the impact of inequality on fragility. The assumption of the hypothesis being that separate from the direct relationship between non-tax revenue and fragility, such revenue also strengthens the effects of inequality on fragility. However, none of the models find a significant interaction effect between the two, and as a result hypothesis H_4 is rejected.

Interestingly, only the GINI model shows significant direct effects of non-tax revenue on fragility, and here only at a 90% significance level. This points to a rejection of Bates theory. Tax revenue, measured as a percentage of GDP, however is significant across all models. Counter to expectations however, the coefficient is positive indicating a positive relationship between tax revenue and fragility. The coefficient is quite small, a mere 0.344 in the full model, indicating a mild effect. Add to this the lacking significance of resource revenue across all models, and Bates (2008) theory of a “state failure equilibrium” appears to be weakened by the results. However, when examining the results of the lagged variables both tax revenue, resource revenue and non-tax revenue are significant at different levels and year lags. To complicate matters even more the coefficients of the lags, are negative when significant, apart from resource revenue at a three-year lag. These findings appear to contradict one another, indicating that while there is quite possibly some validity to Bates (2008) theory, neither it nor this model is able to adequately explain the relationship.

7.5 Other variables

Among the other variables of the model there are also some interesting results from the regression models.

Among the demographic variables population and ethnic fractionalization both produce interesting results. While population is not significant in any model, including when log-transformed in the log-transformed variables model, ethnic fractionalization is significant across all models except for the log-transformed variables model. As the population variable is not significant even when log-transformed, a transformation well suited to population numbers, indicate that the findings of Cunningham and Lemke (2014) does not hold for state fragility. Whether this is a product of population not having the same effect on fragility as violence, or a counter-indication of the results found by Cunningham and Lemke (2014) is up for discussion. I am however, reluctant to call for the rejection of their findings as the basic premise for the research differs regarding dependant variable. The significant results of ethnic fractionalization however, lends credence to the findings of both Cunningham and Lemke (2014) and Kalyvas (2007). Not only is the correlation significant across all models except for the log-transformed variables model, the effect is positive and strong. The fact that the variable lacks significance in both the lagged variables model and the Alternative model 1 presented in the appendix raises some concern however. While it is possible that the effect observed is a result of the non-linear variables included in the full model, I am more inclined to accept this as a possible bi-product of the inverted u-shaped relationship found by Cunningham and Lemke (2014).

The results of the remaining geographical variables present a bit of a mixed bag. While landlockedness produces positive significant coefficients across all models except for the log-transformed variables model. Distance from capitol and travel time to nearest major city are not significant in any model. The significant positive coefficients of landlockedness provide support for the findings of P. Collier (2007), however the lack of significance in both the log-transformed variables model and Alternative model 1 presented in the appendix, sow some doubt regarding this variable. Additionally the theory put forward by P. Collier (2007) regarding landlocked states feature some rather complex interactions and qualifiers limiting the number of real world cases. As a result, drawing conclusions about the theory from the statistical model employed in this thesis is difficult and not something to be done with certainty. As mentioned, both the distance form capitol and travel time from nearest major city, geographical variables related to the theories on capacity decay presented by Buhaug et al. (2009) are not significant across all models. This suggest a rejection of the transferability of their findings to the case of state fragility. While it makes intuitive sense, and perhaps is valid for special cases where state capacity is sufficiently low, this thesis provides no support for the transfer of their findings from the study of violence to that of state fragility. As such, this, along with the findings discussed above, can point to a rejection of the idea of transferability between studies of civil war, and political violence, and of state fragility.

The results presented in chapter 6 finds that regime duration has a significant effect on state fragility in all models. The regime duration variable is related to the theories of oscillating regimes and anocracies presented by Acemoglu and Robinson (2001); Cunningham and Lemke (2014); Fearon and Laitin (2003); Kalyvas (2007) as well as closely associated with Collier's (2007) idea of the conflict trap. Perhaps not surprising based on the amount of theories and findings it is based on, the regime type duration variable reports a negative coefficient at significant confidence levels across the board. Further strengthening existing theory on the matter.

From the sphere of economy, the variable for economic growth, operationalized as GPD pr. capita growth, show a significant negative association with fragility in all models except for the log-transformed variables model and Alternative model 1, the opposite pattern of significance from the GDP pr. capita variable. This suggests that the reason for the significance of GPD pr. capita growth may indeed be caused by the non-linearity GDP pr. capita in the natural number models. It seems likely however, that there is a negative relationship between prosperity and state fragility, wherein higher levels of growth and/or prosperity are associated with lower state fragility. This is also in line with the prevailing theories.

8 Diagnostics

As discussed in chapter 3, breaches the assumptions for regression can weaken the validity of the regression results. To ensure validity, conformity to the regression assumptions were tested. Multiple breaches of assumptions can interfere with testing of isolated assumptions, making it difficult to differentiate between the breaches (Midtbø, 2012, pp. 105-106). Despite this fact, due to the practicality of testing each assumption was tested separately. This chapter presents the results and testing procedure for the diagnostics.

8.1 Multicollinearity

To examine the correlation between the variables included in the model, Table 14 displays the correlation matrix for the variables. Note that the trend variables are omitted as these are not variables describing the observations.

Table 14 - Correlation matrix

	cent~10h	wdi_p~1k	Failed~r	Landlo~d	wd~apcur	wdi~apgr	fe_etfra	bmr_de~r	bdist	capdist	ttime	ictd_r~s	ictd_n~x	wdi_ta~v
centered~10h	1.0000													
wdi_pop1k	0.0899	1.0000												
FailedNeig~r	0.3102	0.2943	1.0000											
Landlocked	0.2213	-0.2544	0.0440	1.0000										
wdi_gdpca~ur	-0.5181	-0.0998	-0.4840	-0.1772	1.0000									
wdi_gdpcapgr	0.0624	0.0654	0.1884	0.0979	-0.2989	1.0000								
fe_etfra	0.4264	0.1456	0.3265	0.2105	-0.4581	0.1619	1.0000							
bmr_demdur	-0.1586	0.2751	-0.2448	-0.1748	0.5501	-0.2148	0.0319	1.0000						
bdist	0.0613	0.5045	0.1818	-0.2170	-0.0322	0.0673	0.0917	0.2155	1.0000					
capdist	0.1177	0.6760	0.1715	-0.1963	-0.0436	0.0920	0.1588	0.2636	0.9132	1.0000				
ttime	0.1912	0.2142	0.1224	-0.0710	-0.0977	0.0473	0.2675	0.1563	0.8009	0.7373	1.0000			
ictd_revres	0.1853	0.0371	0.2985	0.0458	-0.1016	0.1344	0.1378	0.0382	0.1461	0.1735	0.1883	1.0000		
ictd_nontax	-0.1724	-0.0304	-0.0591	-0.1172	0.3118	-0.1259	-0.2286	0.3085	0.1310	0.1019	0.1381	0.5762	1.0000	
wdi_taxrev	-0.2323	-0.3162	-0.2596	-0.0428	0.4699	-0.1627	-0.4516	0.0817	-0.1318	-0.1714	-0.1233	0.0390	0.1383	1.0000

The correlation matrix shows a mixed bag of correlations across the variables. As is to be expected, the levels of correlation vary from the very weak to the very strong. As variables with high levels of correlation add little new information to the model, the standard errors get inflated. High levels of correlation indicate a situation where both variables occur together. This issue of multicollinearity only makes a marked impact at higher levels (Skog, 2004, pp. 286-288).

To test for multicollinearity the Variance Inflation Factor (VIF) test in STATA was used. While there is no set level for how much multicollinearity is too much, Midtbø (2012, p. 129) suggests setting the upper limit for multicollinearity at a VIF score of 10. As VIF is only applicable following a standard regression, one was conducted using a simplified version of the model, excluding interactions, trend and lagged variables as well as the random part of the model[6]. The results of the test are displayed in Table 15 below.

Table 15 - Variance Inflation Factor

Variable	VIF	1/VIF
capdist	12.60	0.079340
bdist	9.77	0.102306
ttime	4.09	0.244463
wdi_gdpcapcur	3.64	0.275033
wdi_pop1k	3.08	0.324411
bmr_demdur	2.57	0.388641
ictd_revres	2.14	0.467359
centered_incsh10h	2.11	0.473301
ictd_nontax	2.08	0.481091
fe_etfra	2.02	0.495302
FailedNeighbour	1.97	0.508691
wdi_taxrev	1.55	0.645950
Landlocked	1.32	0.756471
wdi_gdpcapgr	1.16	0.860360

As the table shows, the only variable exceeding a Variance Inflation Factor of 10 is the **capdist** variable. As this variable is not significant in any model, with significances quite far from the level required for a 95% significance level, this is no major issue in need of further consideration.

[6] STATA command: *regress ffp_fsi centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev*

8.2 Residuals

To assess the distribution of the residuals both statistical and graphical techniques were employed. Diagnostics of both level-1 residuals and level-2 random effects are presented below, both employing standardized measures. While the level-1 residuals are predicted directly by STATA[7] the standardized random effects are based on the random effects estimated by STATA and calculated using the procedure suggested by Rabe-Hesketh and Skrondal (2012, p. 161)[8]. The advantage of using standardized residuals lies in their approximate normal distribution if model assumptions are true (Rabe-Hesketh & Skrondal, 2012, p. 55).

8.2.1 Normal distribution

Below, the distribution of the standardized residuals of the full model are presented graphically using histogram and q-q-plot. A line depicting a normal distribution is overlaid on the graphs. As Figure 7 shows, the standardized residuals deviate somewhat from a normal distribution, with a tighter fit around 0 indicating a higher number of residuals close to 0 than expected by the normal distribution. This leads to a lower density in the remaining bins except for a few outliers outside the normal distribution.

---

[7] Standardized residuals predicted by STATA command: *predict res, rstandard*

[8] Standardized random effects calculated using STATA commands:

predict t1 t2 re, reffects

predict t1_se t2_se re_se, reses

generate diag_se = sqrt(exp(2[lns1_1_1]_cons) - re_se^2)*

replace re = re/diag_se

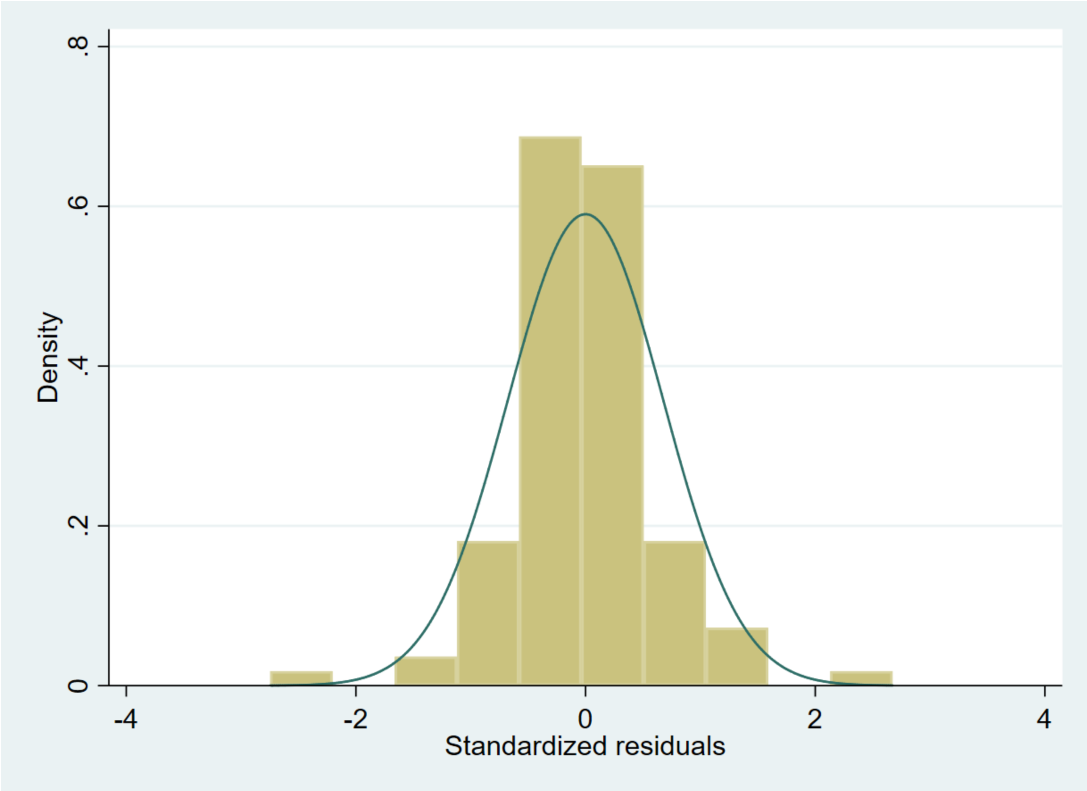


Figure 7 - Histogram of standardized residuals at level-1

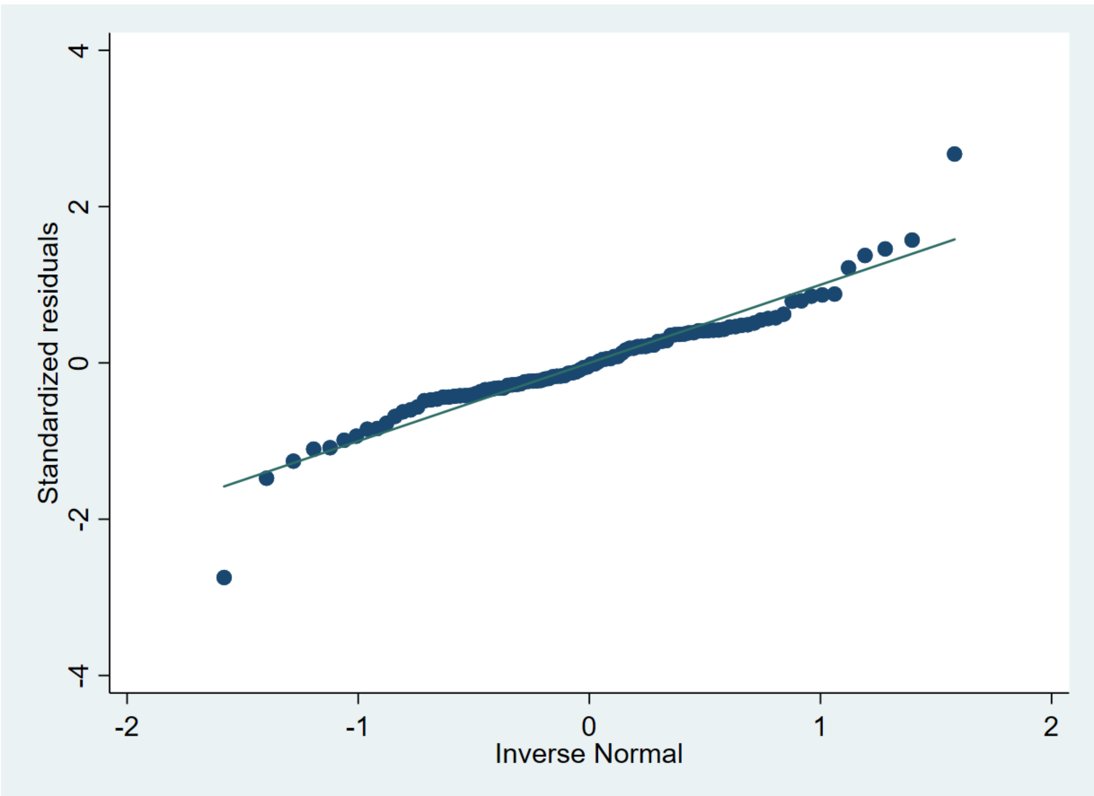


Figure 8 - Q-Q-Plot of standardized residuals at level-1

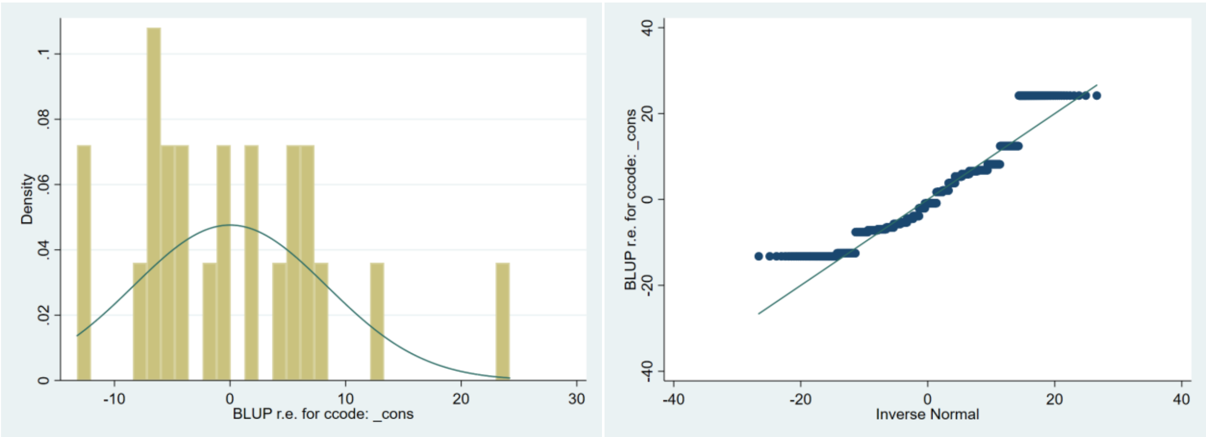
Figure 8 shows a distribution of standardized residuals in line with the histogram above. While most residuals follow the line closely, two outliers are present at the top right and bottom left corners of the plot.

Statistically the distribution of the residuals can be tested using the Skewness-Kurtosis test[9]. The test reports an overall p-value of 0.0018, rejecting normality within a 95% significance level. With p-values for skewness and kurtosis at 0.8347 and 0.0001 respectively. Practically the test tests for the absence of normality, so a significant result indicates the absence of normality.

As both the graphs and Skewness-Kurtosis test shows, the distribution of the residuals deviates somewhat from a normal distribution. This is primarily caused by kurtosis in the residuals. The normal distribution of residuals is considered a negligible issue, especially when dealing with larger datasets (Gelman & Hill, 2007, p. 46; Midtbø, 2012, p. 61).

Figure 9 depicts the distribution of the random intercepts as predicted by Best Linear Unbiased Predictors (BLUP). While some significant deviations from normality, the overall trend of the graphs indicates a situation approaching normal distribution. However, normality of the random intercepts is not a prerequisite for consistent estimation of standard errors and model parameters, nor for the asymptotic normality of the estimators.

Figure 9 - Q-Q Plot and Histogram of varying intercepts RE



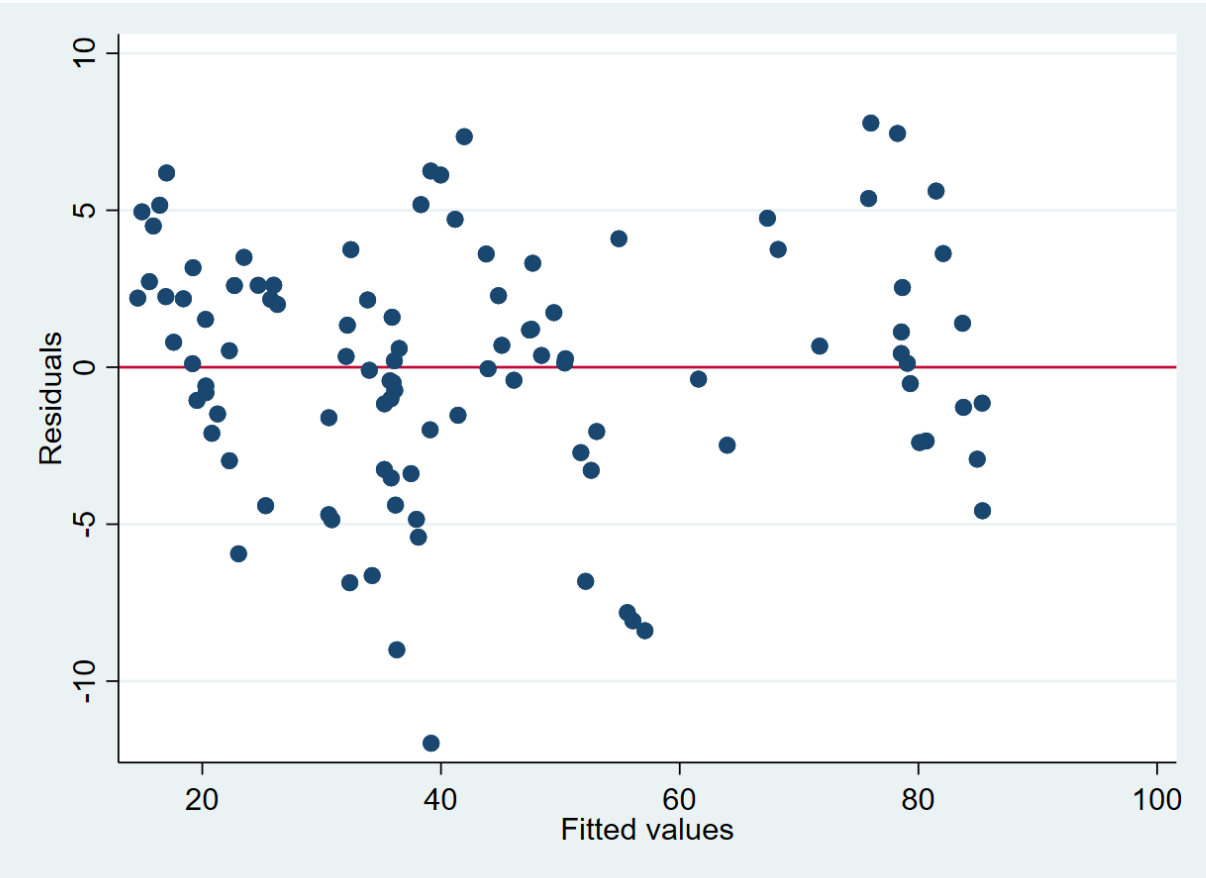
9 STATA command: *sktest res*

8.2.2 **Homoscedasticity**

The homoscedasticity of the residuals can be tested both graphically and statistically using a Breuch-Pagan test. While the graphical test consists of examining the residuals for shape or structure, the statistical test explicitly tests for heteroskedasticity and a significant result indicates heteroskedasticity (Midtbø, 2012, pp. 107-109).

To test for breach of the homoscedasticity assumption, the residuals were plotted against the predicted values using STATA[10]. This test was done against a standard regression version of the model, leaving out trend variables and the random-effects portion of the model.

Figure 10 - Residuals plotted against predicted values.



[10] STATA command: *rvfplot,* with the option, *yline(0)*

As Figure 10 shows, there is no clear shape of notable abnormality to the distribution of the residuals. This is backed up by a Breuch-Pagan test yielding a p-value of 0.3642, a comfortable rejection of heteroscedasticity in the residuals.

8.2.3 Independence

To test for independence among the residuals, an alternative model was run featuring a AR(1) residual structure[11]. In this alternative model, Alternative model 3, the covariance structure is set to the STATA default identity due to estimation failing when including both residual and covariance structure. While the full results from Alternative model 3 are presented in the appendix, the key takeaways from it are presented below.

Alternative model 3 reports a rho of 0.9919906, indicating a very high level of intraclass correlation. Additionally, the autocorrelation is reported at 0.0036892 indicating a low level. The fact that the identity covariance structure was used however, indicates a possible issue. To allow for both AR(1) residual and an unstructured covariance structure, Alternative model 4 was run. This model, while incorporating both the desired residual and covariance structures, does not include the **trend** and **trend2** variables in the random effects portion. The full results of this model are also presented in the appendix. In terms of auto- and intraclass correlation however, the model substantiates the findings of Alternative model 3, with a rho of 0.9806272 and autocorrelation at 2.93e-13. In the full model, which does not include the AR(1) residual structure, the high levels of intraclass correlation will impact significance testing, typically in the form of less significant results.

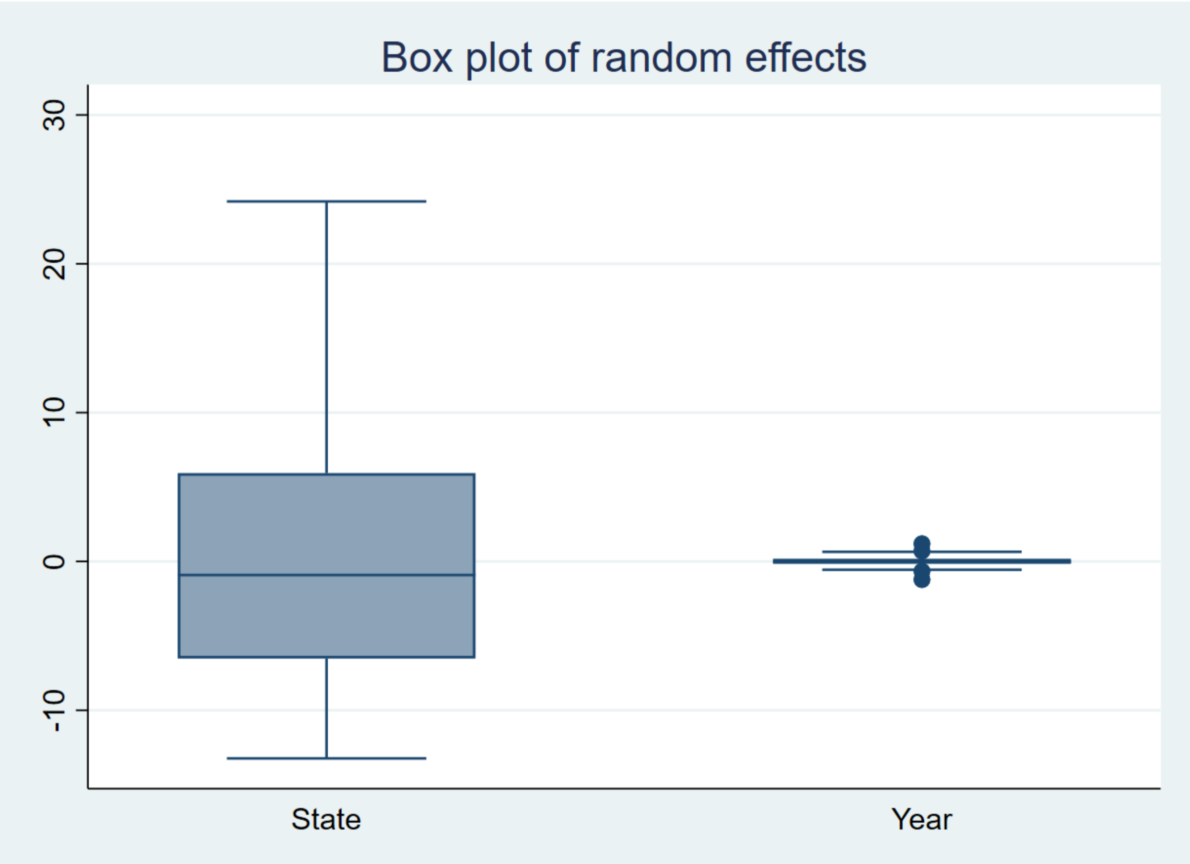
---

[11] STATA command: *mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev*|| *ccode:trend trend2 , residuals(ar 1,t(year)) mle*

8.2.4 Variability between levels

To compare the amount of variability within states and years, Rabe-Hesketh and Skrondal (2012, p. 413) suggests plotting the random intercepts using a box plot. To ensure a valid representation of the within state variability, only one observation per state is included. The observations used were selected at random by STATA[12]. The resulting graph, Figure 11, indicates a much higher variability between states than within.

Figure 11 - Box plot of random effects



[12] STATA command: *egen pickone = tag(ccode)* and *replace re=. if pickone!=1*

8.3 Missing data

As the maximum likelihood estimation readily handles data missing at random but is more vulnerable to data missing not at random, an examination of the missing data is appropriate. As Table 16 below shows, there is a significant amount of data missing from the dataset.

Table 16 - Missing data overview

Variable	Missing observations	Non-missing observations
ffp_fsi	8038	2085
centered_incsh10h	8865	1258
wdi_pop1k	1603	8520
wdi_gdpcapcur	2409	7714
wdi_gdpcapgr	2632	7491
fe_etfra	1256	8867
bmr_demdur	2715	7408
Bdist	1624	8499
ictd_revres	7611	2512
ictd_nontax	5544	4579
wdi_taxrev	6506	3617

To examine patterns in the missing data, Table 17Table 16 was produced using the *xtdescribe* command in STATA on the main explanatory variable, **wdi_incsh10h**[13]. While the table indicates that data is missing for a large number of observations, there is no clear pattern of missingness apparent in the table. As the missing values are, in fact, missing, it is difficult to establish the true nature of the missing data. It does however seem to fit the description of data missing at random and will be treated as such.

[13] STATA command: *xtdescribe wdi_incsh10h*

Table 17 - Missing data patterns

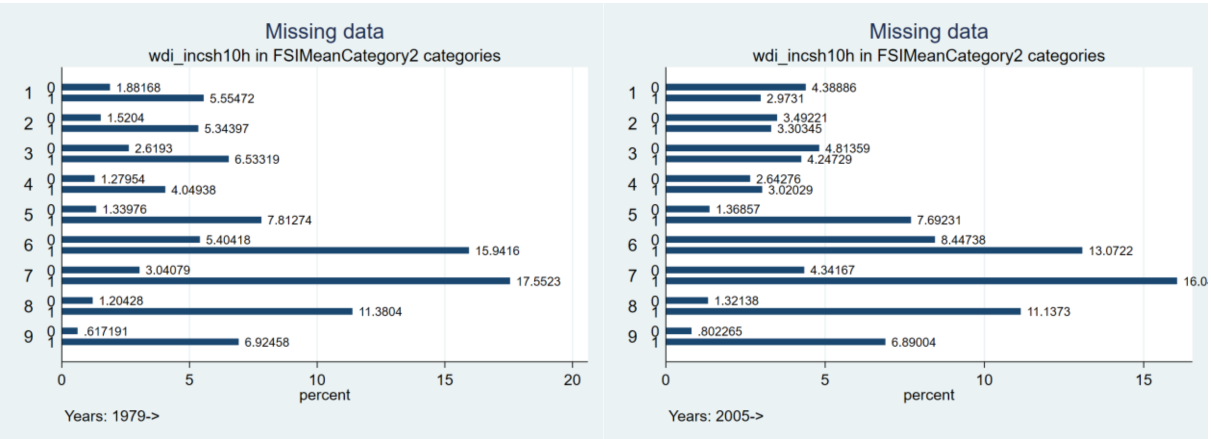
```
   ccode:  8, 12, ..., 894                                 n =        154
    year:  1979, 1980, ..., 2015                           T =         37
           Delta(year) = 1 year
           Span(year)  = 37 periods
           (ccode*year uniquely identifies each observation)

Distribution of T_i:   min      5%      25%       50%       75%     95%     max
                         1       1        3         7        11      21      31

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+--------------------------------------
       16     10.39   10.39 |  ........................11111111111.
        3      1.95   12.34 |  ...............................1....
        3      1.95   14.29 |  ............................1.......
        2      1.30   15.58 |  .................................1..
        2      1.30   16.88 |  ..........................1.....1...
        2      1.30   18.18 |  .........................1.......1..
        2      1.30   19.48 |  .......................1.......1....
        2      1.30   20.78 |  ..................1.................
        2      1.30   22.08 |  .................1111111111111111111
      120     77.92  100.00 | (other patterns)
 ---------------------------+--------------------------------------
      154    100.00         |  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

Figure 12 below displays the amount of missing (1) vs. non-missing data (0) in the categories defined by FSIMeanCategory2.

Figure 12 - Missing data wdi_incsh10h for FSIMeanCategory2 categories



While the graph on the left side of Figure 12 is displays the distribution of missing vs. non-missing data across the categories from 1979, the first year of data for the **wdi_incsh10h** variable, forward. The graph on the right side displays the same distribution from year 2005, the first year of **ffp_fsi** onwards.

As Figure 12 shows, there is a connection between the Fragile States Index scores and missing data in the more recent years. Graphing the percentages of missing vs. non-missing data for each of the categories from the initial year of the Fragile States Index in 2005 onwards, there appears to be a correlation between the amount of missing data and the category of fragility in which the state belongs. This eludes to a possible causality between the degree of fragility and missingness of data. This indicates that data is conditionally missing, however such missingness is still allowed in the missing at random definition as the missingness is dependent on the value of another variable, not the missing variable itself.

When considering this from a practical standpoint, this connection seems plausible, as the state institutions in more fragile states generally are weaker than their counterparts in less fragile states. As a result, the data produced primarily by these kinds of institutions is more likely to be missing. This lack of data from the most fragile states, is an issue with the data weakening the explanatory power of the model. It is however an inescapable problem as there is simply no way of remedying the issue. This does not, however, make the results invalid. It merely highlights the issues faced by many researches working with societal issues in the real world, where perfect data is hard to come by. For the research on fragile states, this is a particular issue.

9 Concluding remarks

The research question and overarching goal of this thesis has been to examine the role of economic inequality in state fragility. While this is a broadly formulated goal, the hypotheses derived from existing theory on the causes of state fragility and political violence narrow the scope of the thesis considerably. Examining both the direct effects of inequality on state fragility as well as the interaction effects of inequality and state income, non-tax revenue and spill-over effects the hypotheses combined attempts to elucidate the role inequality plays in state fragility. This is done from a qualitative standpoint employing a polynomial growth curve model. The key theoretical assumption underpinning the thesis is the destabilizing effects of economic inequality on institutions and societal bonds reducing societal capabilites. While the thesis is grounded in existing theories, the application and approach are novel.

Operationalized as the proportion of income befalling the top decile, inequality is found to have a significant effect on state fragility as measured by the Fragile States Index. The model finds a positive relationship, indicating that increased inequality increases fragility. While the interaction effects initially were found to not be significant, hypothesis H_2 appears to be strengthened by the results of the log-transformed variables model and Alternative model 1 employing log-transformed variables. Hypothesis H_3 and H_4 fail to achieve significance across all models and are rejected. While the growth curve model is unable to explain the mechanics behind this relationship, the theoretical foundations supports the idea that inequality has a negative impact on institutions and social cohesion, increasing the likelihood of violence, self-seeking elites and a diminished state capacity. While far from the sole explanation for state fragility, economic inequality appears to play a part in it.

The findings of this thesis open new avenues of possibility for the handling of state fragility; the reduction of economic inequality.

9.1 Further research

As data on these topics are scarce this thesis is based on the data available at this point in time. A revisiting of the questions at hand with more data at a future date would be of interest. While there is an increasing amount of data being produced in the western world, the perhaps most interesting data is that from the most fragile states themselves. In these states data is hard to come by for practical reasons, as a weak state is far less likely to keep reliable statistics. Yet, as technology and science progress we can hope for more data from these states as well.

The future may also offer possibilities for more data on the distribution of wealth. This would open the possibility of both an analysis of the role of wealth inequality in state fragility. As well as an examination of the overall effects of the combination of wealth and income inequality.

While this thesis offers some small insight into the correlations between inequality and fragility, it does not produce evidence of the mechanics behind them. This would make an interesting subject for further research. Perhaps by employing qualitative research methodologies the root mechanics at work can be explored.

The results of the analysis also provide some interesting avenues of further research, beyond an exploration of the underlying mechanics behind the correlations the role of non-tax and resource state revenue demands further exploration. This thesis finds no significant relationship between neither resource revenue or non-tax revenue and state fragility when examined as simultaneous phenomenon. However, the lagged versions of the variables are at three- and one-year lags respectively. These findings appear curious and warrant further exploration. Such research would also shed additional light on the “State failure equilibrium” theory proposed by Bates (2008).

Works cited (references)

Acemoglu, D., & Robinson, J. (2001). A theory of political transitions. *The American Economic Review, 91*(4), 938-963. doi:10.1257/aer.91.4.938

Acemoglu, D., & Robinson, J. A. (2012). *Why Nations Fail*. London: Profile Books.

Allison, P. D. (2002). *Missing data* (Vol. 136). London: Sage publications.

Ayers, A. (2012). *An illusion of the epoch: Critiquing the ideology of 'failed states'* (Vol. 49).

Bates, R. H. (2008). Probing the sources of political order. In S. N. Kalyvas, I. Shapiro, & T. E. Masoud (Eds.), *Order, Conflict and Violence* (pp. 17-41). Cambridge: Cambridge University Press.

Boix, C. (2015). *Political order and inequality : their foundations and their consequences for human welfare*. New York: Cambridge University Press.

Brock, L., Holm, H.-H., & Sørensen, G. S., Michael. (2012). *Fragile states : violence and the failure of intervention*. Cambridge: Polity Press.

Buhaug, H., Gates, S., & Lujala, P. (2009). Geography, Rebel Capability, and the Duration of Civil Conflict. *Journal of Conflict Resolution, 53*(4), 544-569. doi:10.1177/0022002709336457

Central Intelligence Agency. (2017, 23.01.18). The World Facbook. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/

Chitiga, M., & Sekyere, E. (2014). Income inequality and limitations of the gini index: the case of South Africa. *HSRC Review*.

Collier, D., Brady, H. E., & Seawright, J. (2010). Critiques, Responses, and Trade-Offs: Drawing togehter the debate. In H. E. Brady & D. Collier (Eds.), *Rethinking Social Inquiry* (2nd ed., pp. 125-159). Plymouth, UK: Rowman & Littlefield Publishers.

Collier, P. (2007). *The bottom billion : why the poorest countries are failing and what can be done about it*. Oxford: Oxford University Press.

Collier, P., Elliott, V. L., Hegre, H., Hoeffler, A., Reynal-Querol, M., & Sambanis, N. (2003). *Breaking the Conflict Trap : Civil War and Development Policy*: Washington, DC: World Bank and Oxford University Press.

Cunningham, D. E., & Lemke, D. (2014). Beyond Civil War : A Quantitative Examination of Causes of Violence within Countries. *Civil Wars, 16*(3), 328-345. doi:10.1080/13698249.2014.976356

Duncan, T. E., & Duncan, S. C. (2004). An Introduction to Latent Growth Curve Modeling. *Behavior Therapy, 35*(2), 333-363. doi:10.1016/S0005-7894(04)80042-X

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, Insurgency, and Civil War. *APSR, 97*(1), 75-90. doi:10.1017/S0003055403000534

Frees, E. W. (2004). *Longitudinal and Panel Data : Analysis and Applications in the Social Sciences*. Cambridge: Cambridge University Press.

Fund For Peace. (2018a). About the fund for peace. Retrieved from http://global.fundforpeace.org/aboutus

Fund For Peace. (2018b). Methodology. Retrieved from http://fundforpeace.org/fsi/methodology/

Gates, R. (2010). Helping Others Defend Themselves: The Future of U.S. Security Assistance. *Foreign Affairs, 89*(3), 2-6.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambidge University Press.

Hobbes, T. (1651). *Leviathan : The Matter, Forme, & Power of a Common-Wealth Ecclesiastical and Civill*. Auckland: The Floating Press.

Hox, J. J. (2002). *Multilevel analysis : techniques and applications*. Mahwah, N.J: Lawrence Erlbaum Associates.

Kalyvas, S. N. (2007). Civil Wars. In C. Boix & S. C. Stokes (Eds.), *The Oxford Handbook of Comparative Politics* (pp. 416-434). Oxford: Oxford University Press.

Luke, D. A. (2004). *Multilevel modeling* (Vol. 07-143). Thousand Oaks, Calif: Sage Publications.

Midtbø, T. (2012). *Stata : en entusiastisk innføring*. Oslo: Universitetsforl.

Moses, J. W., & Knutsen, T. L. (2012). *Ways of knowing : competing methodologies in social and political research* (2nd ed. ed.). Basingstoke: Palgrave Macmillan.

OECD. (2016). *States of Fragility 2016*: OECD Publishing.

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, Mass.: Belknap Press.

Piketty, T. (2015). *The economics of inequality* (A. Goldhammer, Trans.): The Belknap Press of Harvard University Press.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata : Vol. 1 : Continuous responses* (3rd ed. ed. Vol. Vol. 1). College Station, Tex: Stata Press.

Rotberg, R. I. (2004). *When states fail : causes and consequences*. Princeton, N.J: Princeton University Press.

Ruspini, E. (2002). *Introduction to longitudinal research*. London: Routledge.

Skog, O.-J. (2004). *Å forklare sosiale fenomener : en regresjonsbasert tilnærming* (2. [rev. og utvidet] utg. ed.). Oslo: Gyldendal akademisk.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Pachon, N. A., & Svensson, R. (2018). *The Quality of Government Standard Dataset*. Retrieved from: http://www.qog.pol.gu.se

The Fund for Peace. (2017). What does state fragility mean? Retrieved from http://fundforpeace.org/fsi/frequently-asked-questions/what-does-state-fragility-mean/

Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID v.2.0 Codebook. Retrieved from http://grid.prio.org/#/codebook

Uchida, H., & Nelson, A. (2009). *Agglomeration Index : Towards a New Measure of Urban Concentration*: Washington, DC: World Bank.

Weber, M. (1948). *From Max Weber : essays in sociology* (H. H. Gerth & C. W. Mills, Trans.). London: Routledge & Kegan Paul.

Weidmann, N. B., Kuse, D., & Gleditsch, K. S. (2010). The Geography of the International System: The CShapes Dataset. *International Interactions, 36*(1), 86-106. doi:10.1080/03050620903554614

Wilkinson, R., & Pickett, K. (2010). *The Spirit Level - Why Equality is Better for Everyone*. London, England: Penguin Books.

Woodward, S. L. (2017). *The ideology of failed states - why intervention fails*. Cambridge, United Kingdom: Cambridge University Press.

Zucman, G. (2013). The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors? *Quarterly journal of economics, 128*(3), 1321-1321. doi:10.1093/qje/qjt012

Appendix

Complete variable list

While Table 5 presents the variables used in the model, Table 18 displays all variables include in the dataset. A STATA readout of the summary statistics of the variables is also included immediately following the table.

Table 18 – Complete variables list

Variable:	Identifier:	Source:
DEPENDENT VARIABLES		
State Fragility – Fragile States Index:		
FSI Total score	ffp_fsi	QOG
FSI Public Services	ffp_ps	QOG
FSI External Intervention	ffp_ext	QOG
FSI Human flight & brain drain	ffp_hf	QOG
FSI State Legitimacy	ffp_sl	QOG
FSI Group Grievance	ffp_gg	QOG
FSI Refugees and IDPs	ffp_ref	QOG
FSI Demographic Pressure	ffp_dp	QOG
FSI Security Apparatus	ffp_sec	QOG
FSI Fractionalized Elites	ffp_fe	QOG
FSI Poverty and Economic Decline	ffp_eco	QOG
FSI Human rights and rule of law	ffp_hr	QOG
FSI Uneven Economic Development	ffp_ued	QOG
Mean FSI (mean of ffp_fsi)	FSIMean	Constructed
Max FSI (max value of ffp_fsi)	FSIMax	Constructed
INDEPENDENT VARIABLES		
Inequality (Core model):		

GINI Index score	wdi_gini	QOG
Income share highest decile	wdi_incsh10h	QOG
Income share highest decile (centered)	centered_incsh10h	QOG
Income share lowest decile	wdi_incsh10l	QOG
Explanatory Variables:		
Total population	wdi_pop	QOG
Total population in thousands	wdi_pop1k	QOG
Urban population	wdi_popurb	QOG
Population density	wdi_popden	QOG
Failed neighbours (max FSI >100)	FailedNeighboursMaxFSI100	Constructed
Failed neighbours (max FSI >90)	FailedNeighboursMaxFSI90	Constructed
Failed neighbours (mean FSI >100)	FailedNeighboursMeanFSI100	Constructed
Failed neighbours (mean FSI >90)	FailedNeighboursMeanFSI90	Constructed
Failed neighbour *	FailedNeighbours	QOG
Landlocked state *	Landlocked	Constructed
Semi-landlocked state *	SemiLandlocked	Constructed
Total GDP (current US$)	wdi_gdppppcur	QOG
GDP pr. Capita (current US$)	wdi_gdpcapcur	QOG
GDP growth	wdi_gdpgr	QOG
GDP pr capita growth	wdi_gdpcapgr	QOG
Ethnic fractionalization	al_ethnic	QOG
Ethnic fractionalization	fe_etfra	QOG
Years of current regime type	bmr_demdur	QOG
Mountainous	mountains	PRIOGRID
Distance from international border	bdist	PRIOGRID
Distance from capitol	capdist	PRIOGRID

Travel time to nearest major city (max)	ttime_max_state_mean	PRIOGRID
Travel time to nearest major city (min)	ttime_min_state_mean	PRIOGRID
Travel time to nearest major city (standard deviation)	ttime_sd_state_mean	PRIOGRID
Travel time to nearest major city (mean)	ttime	PRIOGRID
Total resource revenue	ictd_revres	QOG
Consolidated non-tax revenue	ictd_nontax	QOG
Tax revenue (% of GDP)	wdi_taxrev	QOG
Democracy	Democracy	Constructed
Log-transformed variables:		
Population in thousands	ln_pop	
GDP pr. capita	ln_gdpcapcur	
Years of current regime type	ln_demdur	
Border distance	ln_bdist	
Capitol distance	ln_capdist	
Travel time	ln_ttime	
Trend variables		
Linear trend	trend	Constructed
Quadratic trend	trend2	Constructed
Cubic trend	trend3	Constructed
Quartic trend	trend4	Constructed
Identifying Variables:		
Geographic region	Region	Geographic
Continent	Continent	Geographic
Country Code (Correlates of War)	ccode	QOG
Country Code	ccodecow	QOG
Country Name	cname	QOG

Year	year	QOG
Country Code and Year	ccodealp_year	QOG
Regime type	ht_regtype	QOG
Regime type	ht_regtype1	QOG
FSI category based on FSIMean	FSIMeanCategory	Constructed
FSI category based on FSIMean	FSIMeanCategory2	Constructed
FSI category	FSICategory	Constructed
Residuals:		
Residuals	res	Constructed
Random effects	re	Constructed

* = Number reported in billions for purposes of this table

** = Number reported in millions for purposes of this table

Variable	Obs	Mean	Std. Dev.	Min	Max
ccode	10,123	422.4816	256.7215	4	894
cname	0				
year	10,123	1988.572	16.7675	1960	2017
ccodealp_y~r	0				
ccodecow	10,123	462.4613	247.7132	2	990
al_ethnic	9,684	.4519758	.2563202	0	.930175
bmr_demdur	7,408	40.304	47.03491	1	211
fe_etfra	8,867	.4763567	.2605491	.001999	1
ffp_dp	2,085	6.186763	2.176195	.8	10
ffp_eco	2,085	5.763405	1.918973	1	10
ffp_ext	2,085	5.815252	2.302677	.8	10
ffp_fe	2,085	6.20307	2.455258	.7	10
ffp_fsi	2,085	70.71194	23.47845	16.8	114.9
ffp_gg	2,085	6.008393	2.000511	1	10
ffp_hf	2,085	5.594005	2.049066	.9	10
ffp_hr	2,085	5.877314	2.352146	.7	10
ffp_ps	2,085	5.717506	2.421643	1	10
ffp_ref	2,085	5.085947	2.343832	.9	10
ffp_sec	2,085	5.694005	2.392504	.9	10
ffp_sl	2,085	6.304125	2.453601	.4	10
ffp_ued	2,085	6.462638	1.920715	1	10
ht_regtype	6,838	46.18061	46.22747	1	100
ht_regtype1	6,838	47.59988	48.34707	1	100
ictd_nontax	4,579	5.747138	8.224871	0	89.08322
ictd_revres	2,512	6.660751	11.40106	-.7245431	76.53272
wdi_gdpca~ur	7,714	6429.105	12254.12	37.51812	119172.7
wdi_gdpcapgr	7,491	2.063189	6.25512	-64.99631	140.5011
wdi_gdpgr	7,494	3.942973	6.454945	-64.04711	149.973
wdi_gdpppp~r	4,438	3.83e+11	1.36e+12	1.72e+08	2.14e+13
wdi_gini	1,258	39.81375	9.8637	16.2	65.8
wdi_incsh10h	1,258	31.23521	7.588563	17.1	61.5
wdi_incsh10l	1,258	2.408744	1.059204	0	6.1
wdi_pop	8,520	3.22e+07	1.18e+08	60504	1.38e+09
wdi_popden	8,350	135.1058	435.3512	.6322112	7908.721
wdi_popurb	8,523	48.63819	24.3125	2.193	100
wdi_taxrev	3,617	16.98542	8.580269	.057764	132.5171
FSIMean	10,123	70.63851	23.2268	18.76667	113.0333
Region	10,123	8.873555	5.63791	0	21
Continent	10,123	1.508249	1.236189	0	4
Landlocked	10,123	.2195989	.4139954	0	1
FSIMeanCat~y	10,123	2.839079	.8285593	1	4
Fail~xFSI100	10,123	.5040008	.9044768	0	5
Faile~xFSI90	10,123	1.056801	1.611138	0	7
Fail~nFSI100	10,123	.3772597	.7462112	0	4
Faile~nFSI90	10,123	.7187593	1.28633	0	7
FSIMax	10,123	74.7751	23.77621	20	114.9
FSICategory	10,123	79.19263	38.89466	1	99
FSIMeanCat~2	10,123	5.56426	2.299475	1	9
Democracy	6,838	.421322	.4938071	0	1
mountains	9,833	.3058064	.2729276	0	.9608333
bdist	8,499	137.4737	166.8427	7.178968	1177.826
capdist	10,123	406.4723	480.1673	12.7008	3491.659
ttime_max_~n	2,293	1017.437	859.6526	73	7065.34
ttime	10,123	502.3731	551.8233	17.42888	4152.241
ttime_min_~n	2,293	262.4941	475.8093	0	4058.5
ttime_sd_s~n	2,293	155.627	117.579	14.78891	750.3627
wdi_pop1k	8,520	32228.93	118355.5	60.504	1378665
centered~10h	1,258	1.04e-08	7.588563	-14.13521	30.26479
FailedNeig~r	10,123	.4139089	.4925569	0	1
SemiLandlo~d	10,123	.0286476	.1668223	0	1
trend	2,293	6.014392	3.739472	0	12
trend2	2,293	50.15046	46.60198	0	144
trend3	2,293	469.5347	547.3725	0	1728
trend4	2,293	4685.866	6383.2	0	20736
ln_pop	8,520	8.768017	1.792483	4.102709	14.13663
ln_gdpcapcur	7,714	7.44454	1.669344	3.624824	11.68833
ln_demdur	7,408	3.041337	1.233557	0	5.351858
ln_bdist	8,499	4.480289	.9208426	1.971156	7.071425
ln_capdist	10,123	5.538674	.9947501	2.541665	8.158133
ln_ttime	10,123	5.797027	.929564	2.858129	8.331404
res	102	1.21e-09	.6760197	-2.74615	2.673244
t1	1,334	4.54e-09	1.852879	-2.789676	3.968825
t2	1,334	-1.30e-09	.2420292	-.5114832	.3223876
re	1,044	-.3941183	4.556329	-7.123818	12.11775
t1_se	1,334	.7576444	.3656397	.3469542	1.585513
t2_se	1,334	.1103935	.0338786	.0652586	.172913
re_se	1,334	1.271781	.9842502	.3918972	3.958999
diag_se	1,044	1.872672	.0624827	1.729554	1.996152

Variable development:

To expand on the development of the main variables over time presented in chapter 5 the following section presents the development for all states in trellis graphs.

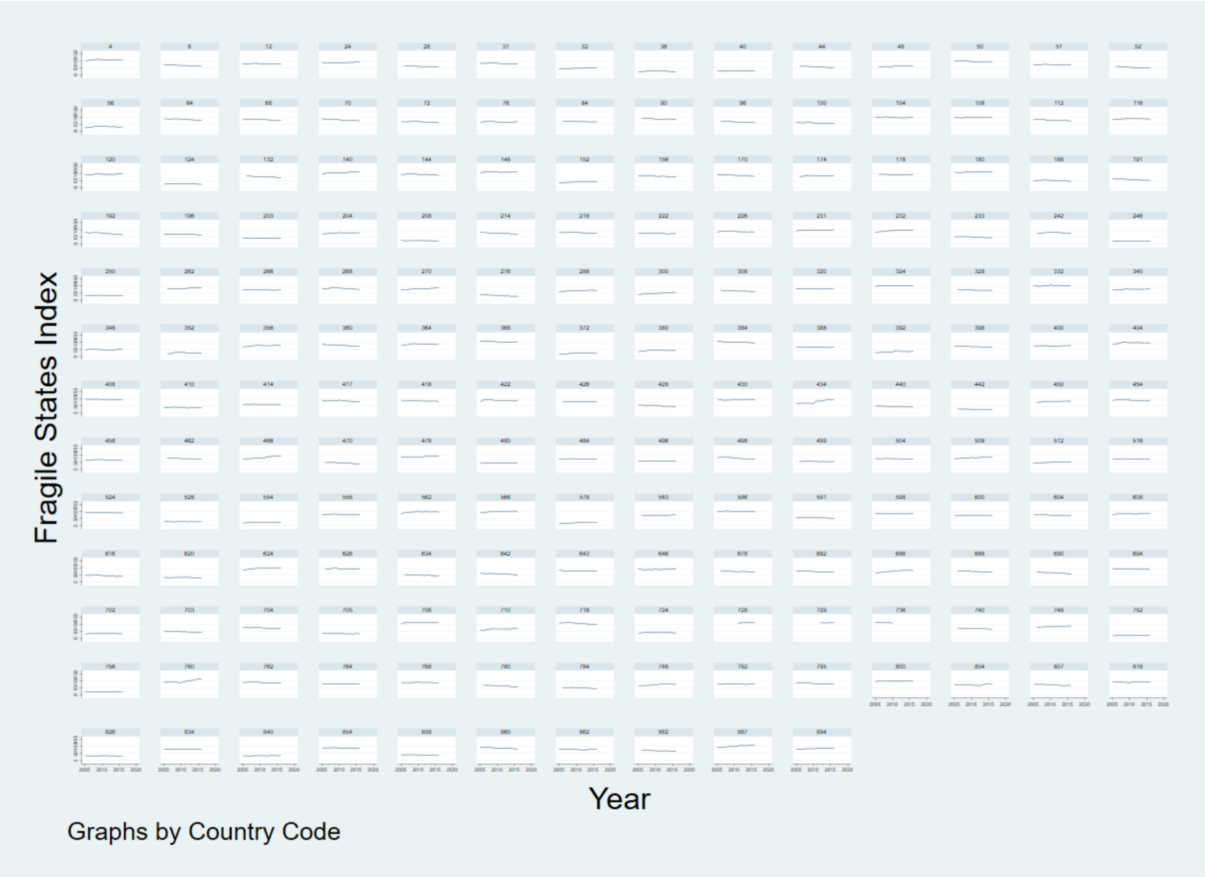


Figure 13 - ffp-fsi development across all states

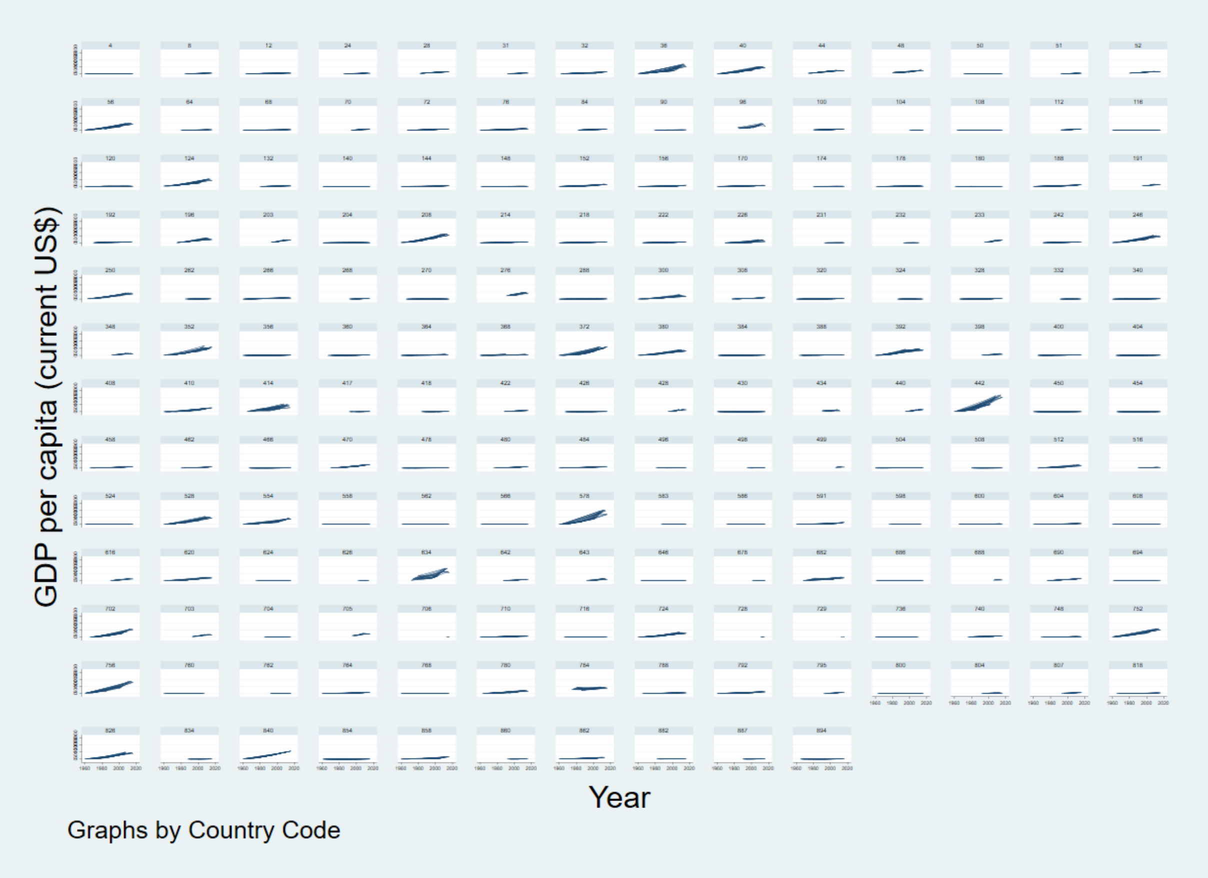


Figure 14 - wdi_gdpcapcur development across all states

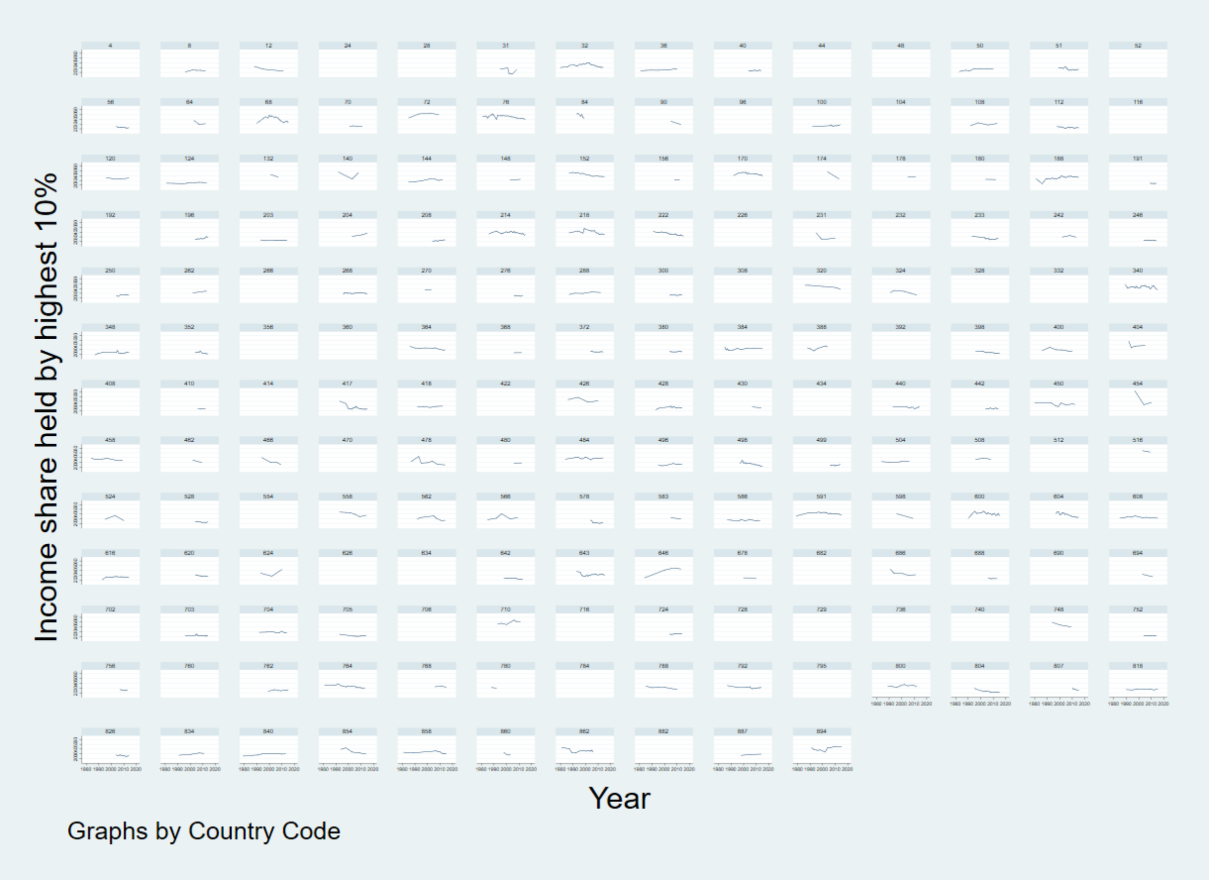


Figure 15 - wdi_incsh10h development across all states

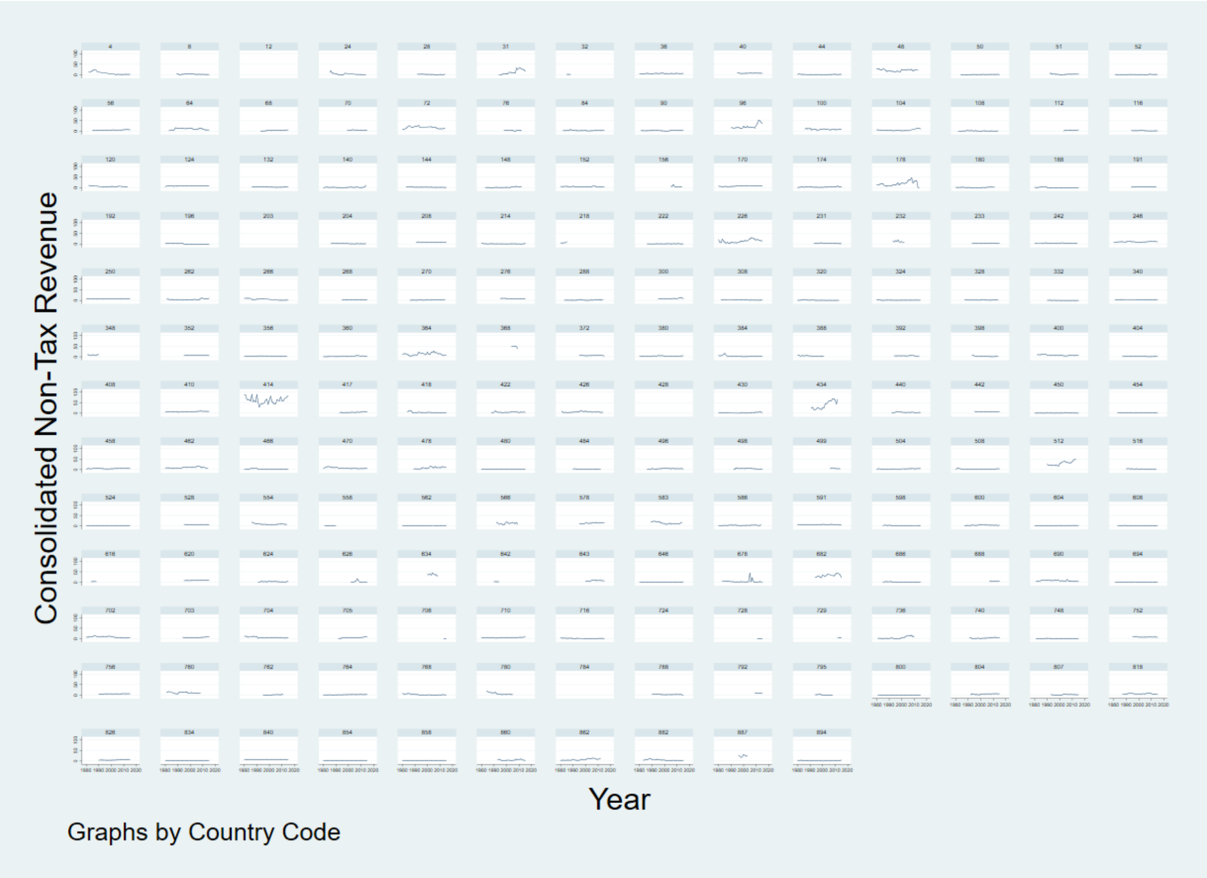


Figure 16 - ictd_nontax development across all states

STATA Outputs

To provide additional information on the estimations of the different models, the full STATA outputs are included below. The STATA commands generating the outputs are also included for clarity.

Explanatory variables model:

Model featuring explanatory variables only, no lags or interactions.

<u>STATA command:</u> mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k c.FailedNeighbour c.Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

```
Mixed-effects ML regression                     Number of obs      =       172
Group variable: ccode                           Number of groups   =        47

                                                Obs per group:
                                                              min =          1
                                                              avg =        3.7
                                                              max =          6

                                                Wald chi2(17)      =    141.62
Log likelihood = -418.26291                     Prob > chi2        =    0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	-.0770973	.5097671	-0.15	0.880	-1.076222	.9220278
trend2	.5032696	.2377048	2.12	0.034	.0373767	.9691625
trend3	-.0823126	.031673	-2.60	0.009	-.1443905	-.0202347
centered_incsh10h	.0075708	.1077948	0.07	0.944	-.2037031	.2188448
wdi_pop1k	.000123	.00005	2.46	0.014	.0000249	.000221
FailedNeighbour	25.27155	4.96883	5.09	0.000	15.53282	35.01028
Landlocked	5.754738	4.965312	1.16	0.246	-3.977094	15.48657
wdi_gdpcapcur	-.0000711	.0000368	-1.93	0.054	-.0001432	1.10e-06
wdi_gdpcapgr	.0049161	.0295593	0.17	0.868	-.053019	.0628512
fe_etfra	34.82202	9.398784	3.70	0.000	16.40075	53.2433
bmr_demdur	-.0763407	.0367321	-2.08	0.038	-.1483344	-.0043471
bdist	-.0114024	.0240297	-0.47	0.635	-.0584998	.0356949
capdist	-.009091	.0079795	-1.14	0.255	-.0247304	.0065485
ttime	.0099802	.0067866	1.47	0.141	-.0033212	.0232817
ictd_revres	-.1720665	.1053461	-1.63	0.102	-.3785411	.0344082
ictd_nontax	.0051002	.1316644	0.04	0.969	-.2529573	.2631576
wdi_taxrev	.2683068	.1455813	1.84	0.065	-.0170273	.5536409
_cons	32.83917	5.501386	5.97	0.000	22.05665	43.62169

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	1.632974	.6363507	.7608094	3.504955
var(trend2)	.0167638	.0124269	.0039209	.0716744
var(_cons)	232.1599	59.86031	140.0595	384.8236
cov(trend,trend2)	-.1319615	.0809709	-.2906615	.0267385
cov(trend,_cons)	-6.700924	6.365739	-19.17754	5.775695
cov(trend2,_cons)	.0329932	.8587694	-1.650164	1.71615
var(Residual)	.6921383	.1386173	.4674329	1.024865

```
LR test vs. linear model: chi2(6) = 370.90               Prob > chi2 = 0.0000
```

Interactions model:

Model including interactions in addition to the explanatory variables.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k

FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h|| ccode:trend trend2 ,mle covariance(unstructured)

```
Mixed-effects ML regression                     Number of obs     =        172
Group variable: ccode                           Number of groups  =         47

                                                Obs per group:
                                                              min =          1
                                                              avg =        3.7
                                                              max =          6

                                                Wald chi2(21)     =     166.39
Log likelihood = -414.64545                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	-.146749	.5068802	-0.29	0.772	-1.140216	.8467179
trend2	.5855837	.2467556	2.37	0.018	.1019516	1.069216
trend3	-.0959585	.0332564	-2.89	0.004	-.1611399	-.0307772
centered_incsh10h	.3235578	.1702323	1.90	0.057	-.0100913	.6572069
wdi_pop1k	.0001141	.0000473	2.41	0.016	.0000214	.0002068
FailedNeighbour	24.05854	4.733728	5.08	0.000	14.78061	33.33648
Landlocked	6.468208	4.691896	1.38	0.168	-2.727739	15.66416
wdi_gdpcapcur	-.000115	.0000574	-2.00	0.045	-.0002275	-2.46e-06
wdi_gdpcapgr	.0150822	.0305649	0.49	0.622	-.0448238	.0749882
fe_etfra	35.04359	8.90606	3.93	0.000	17.58803	52.49914
bmr_demdur	-.0640925	.0344984	-1.86	0.063	-.1317082	.0035232
bdist	-.010151	.0227382	-0.45	0.655	-.054717	.0344149
capdist	-.0096996	.0075137	-1.29	0.197	-.0244262	.0050269
ttime	.0103156	.0064409	1.60	0.109	-.0023083	.0229395
ictd_revres	-.1370605	.1076882	-1.27	0.203	-.3481254	.0740044
ictd_nontax	-.1245954	.1418104	-0.88	0.380	-.4025386	.1533479
wdi_taxrev	.3206809	.1442682	2.22	0.026	.0379205	.6034414
c.ictd_nontax#c.centered_incsh10h	-.0250353	.0210897	-1.19	0.235	-.0663703	.0162997
c.wdi_gdpcapcur#c.centered_incsh10h	-2.19e-06	5.78e-06	-0.38	0.705	-.0000135	9.14e-06
c.bdist#c.centered_incsh10h	-.0000544	.000579	-0.09	0.925	-.0011892	.0010805
c.FailedNeighbour#c.centered_incsh10h	-.6265954	.3426035	-1.83	0.067	-1.298086	.0448952
_cons	33.47396	5.383177	6.22	0.000	22.92312	44.02479

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	1.249625	.569088	.5118426	3.050864
var(trend2)	.0080046	.0128911	.0003408	.1879996
var(_cons)	220.0053	56.37419	133.1438	363.5341
cov(trend,trend2)	-.0773901	.0775605	-.2294059	.0746257
cov(trend,_cons)	-6.731165	5.524073	-17.55815	4.095819
cov(trend2,_cons)	-.0315277	.7272192	-1.456851	1.393796
var(Residual)	.753522	.1652659	.4902375	1.158205

```
LR test vs. linear model: chi2(6) = 342.54                Prob > chi2 = 0.0000
```

Full model:

The full model is the main model used in the thesis. It contains explanatory variables, interactions and lagged variables.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     276.08
Log likelihood = -194.65547                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.928285	.9800925	1.97	0.049	.0073393	3.849231
trend2	-.3431564	.3483858	-0.98	0.325	-1.02598	.3396673
trend3	.0318527	.0399913	0.80	0.426	-.0465287	.1102342
centered_incsh10h	.3741926	.188678	1.98	0.047	.0043904	.7439947
wdi_pop1k	-.000068	.000099	-0.69	0.492	-.0002619	.000126
FailedNeighbour	21.98966	9.526424	2.31	0.021	3.318212	40.66111
Landlocked	8.443465	4.139489	2.04	0.041	.3302152	16.55671
wdi_gdpcapcur	-.0000481	.0000687	-0.70	0.484	-.0001828	.0000865
wdi_gdpcapgr	-.0804998	.0227979	-3.53	0.000	-.1251828	-.0358168
fe_etfra	16.50858	9.575367	1.72	0.085	-2.258796	35.27595
bmr_demdur	-.2382053	.0396314	-6.01	0.000	-.3158814	-.1605291
bdist	.0098883	.0232227	0.43	0.670	-.0356274	.0554039
capdist	.0010394	.0113202	0.09	0.927	-.0211479	.0232267
ttime	-.0008107	.0169031	-0.05	0.962	-.0339401	.0323188
ictd_revres	.1159606	.1688722	0.69	0.492	-.2150227	.446944
ictd_nontax	-.3994362	.2468245	-1.62	0.106	-.8832034	.084331
wdi_taxrev	.3436291	.1501637	2.29	0.022	.0493137	.6379445
c.ictd_nontax#						
c.centered_incsh10h	.0222502	.0369326	0.60	0.547	-.0501365	.0946368
c.wdi_gdpcapcur#						
c.centered_incsh10h	-8.61e-06	7.86e-06	-1.10	0.273	-.000024	6.79e-06
c.bdist#						
c.centered_incsh10h	-.0015735	.0014647	-1.07	0.283	-.0044443	.0012973
c.FailedNeighbour#						
c.centered_incsh10h	2.320665	1.624023	1.43	0.153	-.8623618	5.503691
wdi_gdpcapgr						
L3.	-.2805135	.0751018	-3.74	0.000	-.4277104	-.1333167
centered_incsh10h						
L1.	-.0840067	.1406988	-0.60	0.550	-.3597713	.1917578
L2.	-.0258027	.1174169	-0.22	0.826	-.2559356	.2043303
ictd_revres						
L3.	.7131551	.1855323	3.84	0.000	.3495185	1.076792
ictd_nontax						
L1.	-.655918	.1734393	-3.78	0.000	-.9958527	-.3159833
L2.	-.4350106	.2429487	-1.79	0.073	-.9111812	.0411601
wdi_taxrev						
L1.	-.1390894	.146688	-0.95	0.343	-.4265927	.1484138
L2.	.1277729	.1800577	0.71	0.478	-.2251337	.4806794
L3.	-.3260539	.1911574	-1.71	0.088	-.7007155	.0486076
_cons	56.46234	7.407887	7.62	0.000	41.94315	70.98154

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	4.138205	1.70848	1.842421	9.294696
var(trend2)	.0718679	.034422	.0281087	.1837505
var(_cons)	72.82532	34.74317	28.58874	185.511
cov(trend,trend2)	-.4822094	.2316434	-.9362222	-.0281966
cov(trend,_cons)	.7303084	7.198228	-13.37796	14.83858
cov(trend2,_cons)	-.9480457	1.010032	-2.927671	1.03158
var(Residual)	.1977903	.0561848	.1133471	.3451434

LR test vs. linear model: chi2(6) = 177.51   Prob > chi2 = 0.0000

Log-transformed variables model:

This model replaces the variables in the full model with logarithmically transformed variables where applicable.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h ln_pop FailedNeighbour Landlocked ln_gdpcapcur wdi_gdpcapgr fe_etfra ln_demdur ln_bdist ln_capdist ln_ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     661.60
Log likelihood = -186.20345                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.489382	1.001176	1.49	0.137	-.472886	3.45165
trend2	.28764	.3861856	0.74	0.456	-.4692699	1.04455
trend3	-.0634164	.0461435	-1.37	0.169	-.1538559	.0270232
centered_incsh10h	.3139439	.2031013	1.55	0.122	-.0841273	.7120151
ln_pop	1.367786	1.792008	0.76	0.445	-2.144485	4.880057
FailedNeighbour	12.81147	7.238933	1.77	0.077	-1.376578	26.99952
Landlocked	3.40977	3.497298	0.97	0.330	-3.444809	10.26435
ln_gdpcapcur	-7.383634	1.354711	-5.45	0.000	-10.03882	-4.72845
wdi_gdpcapgr	.0223912	.0306876	0.73	0.466	-.0377555	.0825378
fe_etfra	9.385444	7.55985	1.24	0.214	-5.431589	24.20248
ln_demdur	-8.808515	2.037408	-4.32	0.000	-12.80176	-4.815268
ln_bdist	3.018612	2.496842	1.21	0.227	-1.875109	7.912332
ln_capdist	-2.278563	3.021866	-0.75	0.451	-8.201311	3.644184
ln_ttime	-.3342375	3.288802	-0.10	0.919	-6.780171	6.111696
ictd_revres	.1575932	.178488	0.88	0.377	-.1922368	.5074232
ictd_nontax	-.1925737	.2511401	-0.77	0.443	-.6847991	.2996518
wdi_taxrev	.380101	.1558773	2.44	0.015	.0745871	.685615
c.ictd_nontax#						
c.centered_incsh10h	.0256343	.0386179	0.66	0.507	-.0500554	.101324
c.wdi_gdpcapcur#						
c.centered_incsh10h	-.00001	4.53e-06	-2.21	0.027	-.0000189	-1.14e-06
c.bdist#						
c.centered_incsh10h	-.0002537	.0015892	-0.16	0.873	-.0033686	.0028611
c.FailedNeighbour#						
c.centered_incsh10h	.6502058	1.726055	0.38	0.706	-2.732799	4.033211
wdi_gdpcapgr						
L3.	-.2100556	.0814913	-2.58	0.010	-.3697757	-.0503356
centered_incsh10h						
L1.	-.1303074	.1265294	-1.03	0.303	-.3783005	.1176857
L2.	-.0525525	.115101	-0.46	0.648	-.2781463	.1730413
ictd_revres						
L3.	.6006354	.1972839	3.04	0.002	.213966	.9873047
ictd_nontax						
L1.	-.7664016	.1833184	-4.18	0.000	-1.125699	-.4071042
L2.	-.1941205	.2355465	-0.82	0.410	-.6557831	.2675422
wdi_taxrev						
L1.	.0306557	.1602102	0.19	0.848	-.2833506	.3446619
L2.	.1930527	.194923	0.99	0.322	-.1889893	.5750947
L3.	-.2983446	.1999175	-1.49	0.136	-.6901756	.0934865
_cons	129.0903	21.63185	5.97	0.000	86.6927	171.488

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	3.54087	2.037343	1.146427	10.93638
var(trend2)	.0840775	.0457524	.0289389	.2442741
var(_cons)	30.78507	18.39587	9.54328	99.30765
cov(trend,trend2)	-.5096061	.3036842	-1.104816	.0856041
cov(trend,_cons)	3.098535	4.841813	-6.391244	12.58831
cov(trend2,_cons)	-.9273494	.6945986	-2.288738	.4340389
var(Residual)	.2837421	.0938826	.1483488	.5427045

```
LR test vs. linear model: chi2(6) = 132.80                 Prob > chi2 = 0.0000
```

GINI Model:

In the GINI model the main explanatory variable is replaced with the GINI coefficient, wdi_gini.

STATA command: mixed ffp_fsi trend trend2 trend3 wdi_gini wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     265.03
Log likelihood = -195.00758                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.864931	.9815225	1.90	0.057	-.0588181	3.78868
trend2	-.3021587	.3507807	-0.86	0.389	-.9896762	.3853589
trend3	.0256982	.0404334	0.64	0.525	-.0535498	.1049462
wdi_gini	.3255096	.1654975	1.97	0.049	.0011404	.6498788
wdi_pop1k	-.0000428	.0000962	-0.44	0.657	-.0002313	.0001458
FailedNeighbour	20.66468	9.235501	2.24	0.025	2.563427	38.76593
Landlocked	8.295586	4.183935	1.98	0.047	.0952252	16.49595
wdi_gdpcapcur	-.0000546	.0000691	-0.79	0.430	-.0001901	.0000809
wdi_gdpcapgr	-.0711954	.0234127	-3.04	0.002	-.1170834	-.0253074
fe_etfra	19.39251	9.595578	2.02	0.043	.5855195	38.19949
bmr_demdur	-.233715	.0394216	-5.93	0.000	-.3109798	-.1564502
bdist	.0104256	.0232894	0.45	0.654	-.0352208	.0560721
capdist	-.001932	.0112749	-0.17	0.864	-.0240303	.0201663
ttime	.00323	.0162819	0.20	0.843	-.0286819	.0351419
ictd_revres	.1103194	.1721744	0.64	0.522	-.2271363	.4477751
ictd_nontax	-.448773	.2527154	-1.78	0.076	-.9440861	.0465401
wdi_taxrev	.4039765	.1530543	2.64	0.008	.1039956	.7039574
c.ictd_nontax#						
c.centered_incsh10h	.0130837	.0381624	0.34	0.732	-.0617133	.0878806
c.wdi_gdpcapcur#						
c.centered_incsh10h	-8.54e-06	7.93e-06	-1.08	0.281	-.0000241	7.00e-06
c.bdist#						
c.centered_incsh10h	-.0012071	.0014604	-0.83	0.408	-.0040695	.0016552
c.FailedNeighbour#						
c.centered_incsh10h	1.872077	1.612887	1.16	0.246	-1.289124	5.033278
wdi_gdpcapgr						
L3.	-.2703561	.0763274	-3.54	0.000	-.419955	-.1207572
centered_incsh10h						
L1.	-.1267339	.1422171	-0.89	0.373	-.4054742	.1520065
L2.	-.0110225	.1187153	-0.09	0.926	-.2437003	.2216553
ictd_revres						
L3.	.7310918	.185538	3.94	0.000	.367444	1.09474
ictd_nontax						
L1.	-.646664	.1735226	-3.73	0.000	-.9867621	-.3065659
L2.	-.4156369	.2428502	-1.71	0.087	-.8916145	.0603407
wdi_taxrev						
L1.	-.1078877	.1474745	-0.73	0.464	-.3969324	.181157
L2.	.1504099	.1800651	0.84	0.404	-.2025113	.5033311
L3.	-.3090415	.1916383	-1.61	0.107	-.6846456	.0665627
_cons	39.81925	9.311909	4.28	0.000	21.56824	58.07026

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	4.09996	1.701325	1.817874	9.246883
var(trend2)	.0661717	.0328821	.0249858	.1752479
var(_cons)	78.19514	38.53897	29.76184	205.447
cov(trend,trend2)	-.4598959	.2246973	-.9002945	-.0194973
cov(trend,_cons)	-1.767419	7.506989	-16.48085	12.94601
cov(trend2,_cons)	-.6633122	.9789529	-2.582025	1.2554
var(Residual)	.2062745	.0603633	.1162394	.3660478

```
LR test vs. linear model: chi2(6) = 179.15                Prob > chi2 = 0.0000
```

Alternative models:

The alternative models referenced in the thesis are here presented with both STATA outputs and commands.

Alternative model 1:

This model replaces only one explanatory variable, wdi_gdpcapcur, with its logarithmically transformed counterpart.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked ln_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.ln_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

BIC: 552.3406

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     673.98
Log likelihood = -188.29582                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.442681	1.060141	1.36	0.174	-.6351573	3.520519
trend2	.4617953	.4054793	1.14	0.255	-.3329296	1.25652
trend3	-.0947416	.0482574	-1.96	0.050	-.1893244	-.0001589
centered_incsh10h	.5208557	.2044727	2.55	0.011	.1200965	.9216148
wdi_pop1k	-.0000585	.0000629	-0.93	0.353	-.0001818	.0000649
FailedNeighbour	16.66937	5.784111	2.88	0.004	5.332718	28.00602
Landlocked	3.649731	2.877191	1.27	0.205	-1.98946	9.288922
ln_gdpcapcur	-9.250515	1.409696	-6.56	0.000	-12.01347	-6.487562
wdi_gdpcapgr	.0445557	.0320734	1.39	0.165	-.0183071	.1074185
fe_etfra	8.068469	6.672011	1.21	0.227	-5.008432	21.14537
bmr_demdur	-.1499137	.0337974	-4.44	0.000	-.2161554	-.083672
bdist	.0099827	.0166667	0.60	0.549	-.0226834	.0426488
capdist	.0109207	.0079946	1.37	0.172	-.0047483	.0265898
ttime	-.0245117	.0115893	-2.12	0.034	-.0472264	-.0017971
ictd_revres	.1287911	.1852484	0.70	0.487	-.234289	.4918713
ictd_nontax	-.0021338	.2705052	-0.01	0.994	-.5323142	.5280465
wdi_taxrev	.4126578	.1624457	2.54	0.011	.0942701	.7310454
c.ictd_nontax#						
c.centered_incsh10h	.0480651	.0402666	1.19	0.233	-.030856	.1269861
c.wdi_gdpcapcur#						
c.centered_incsh10h	-.0000153	4.82e-06	-3.17	0.002	-.0000247	-5.83e-06
c.bdist#						
c.centered_incsh10h	-.0014911	.0015005	-0.99	0.320	-.0044319	.0014498
c.FailedNeighbour#						
c.centered_incsh10h	2.232242	1.60675	1.39	0.165	-.91693	5.381414
wdi_gdpcapgr						
L3.	-.2718635	.0896052	-3.03	0.002	-.4474864	-.0962405
centered_incsh10h						
L1.	-.064803	.1385301	-0.47	0.640	-.336317	.2067109
L2.	.1052813	.1236827	0.85	0.395	-.1371322	.3476949
ictd_revres						
L3.	.6443254	.2009324	3.21	0.001	.2505051	1.038146
ictd_nontax						
L1.	-.7311157	.1877048	-3.90	0.000	-1.09901	-.3632209
L2.	-.2456906	.2371076	-1.04	0.300	-.710413	.2190317
wdi_taxrev						
L1.	-.0505594	.1633921	-0.31	0.757	-.3708021	.2696833
L2.	.1100376	.2079246	0.53	0.597	-.297487	.5175622
L3.	-.4026758	.209331	-1.92	0.054	-.8129571	.0076054
_cons	142.3871	14.1462	10.07	0.000	114.6611	170.1132

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	4.007025	1.957705	1.537985	10.4398
var(trend2)	.0872032	.0447997	.0318593	.2386867
var(_cons)	19.71325	12.0855	5.92811	65.55413
cov(trend,trend2)	-.5566481	.2932581	-1.131423	.0181271
cov(trend,_cons)	-1.598359	4.504373	-10.42677	7.230049
cov(trend2,_cons)	-.0320005	.8175023	-1.634276	1.570275
var(Residual)	.3021794	.0948121	.1633771	.5589058

```
LR test vs. linear model: chi2(6) = 137.50                Prob > chi2 = 0.0000
```

Alternative model 2 – FSI with ffp_ued subtracted

This model replaces the dependent variable with AlternativeFSI. AlternativeFSI ffp_fsi with the scores of ffp_ued subtracted.

<u>STATA command:</u>

generate AlternativeFSI = ffp_fsi-ffp_ued
mixed AlternativeFSI trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:trend trend2 ,mle covariance(unstructured)

Mixed-effects ML regression   Number of obs   =   102
Group variable: ccode   Number of groups   =   23

Obs per group:
min = 2
avg = 4.4
max = 6

Wald chi2(30) = 289.57
Log likelihood = -187.01614   Prob > chi2 = 0.0000

AlternativeFSI	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.680539	.8769243	1.92	0.055	-.0382011	3.399279
trend2	-.3109848	.3116409	-1.00	0.318	-.9217898	.2998201
trend3	.0330146	.03571	0.92	0.355	-.0369756	.1030049
centered_incsh10h	.3037256	.1690343	1.80	0.072	-.0275755	.6350268
wdi_pop1k	-.0000935	.0000986	-0.95	0.343	-.0002868	.0000999
FailedNeighbour	22.43179	9.720492	2.31	0.021	3.379979	41.48361
Landlocked	6.795631	3.870214	1.76	0.079	-.7898492	14.38111
wdi_gdpcapcur	-.0000119	.0000623	-0.19	0.848	-.000134	.0001102
wdi_gdpcapgr	-.0749127	.0204432	-3.66	0.000	-.1149807	-.0348448
fe_etfra	10.79316	9.133362	1.18	0.237	-7.107903	28.69422
bmr_demdur	-.216708	.0383627	-5.65	0.000	-.2918974	-.1415185
bdist	.0012063	.0220354	0.05	0.956	-.0419824	.044395
capdist	.0052749	.0108059	0.49	0.625	-.0159042	.026454
ttime	-.0013426	.0165037	-0.08	0.935	-.0336893	.031004
ictd_revres	.0736575	.1514036	0.49	0.627	-.223088	.370403
ictd_nontax	-.3941858	.221805	-1.78	0.076	-.8289155	.040544
wdi_taxrev	.2838984	.1340346	2.12	0.034	.0211955	.5466013
c.ictd_nontax#c.centered_incsh10h	.0218762	.0333544	0.66	0.512	-.0434972	.0872496
c.wdi_gdpcapcur#c.centered_incsh10h	-5.10e-06	7.06e-06	-0.72	0.470	-.0000189	8.74e-06
c.bdist#c.centered_incsh10h	-.002173	.0013034	-1.67	0.095	-.0047276	.0003816
c.FailedNeighbour#c.centered_incsh10h	3.037435	1.456182	2.09	0.037	.1833709	5.891499
wdi_gdpcapgr						
L3.	-.27665	.0666514	-4.15	0.000	-.4072844	-.1460157
centered_incsh10h						
L1.	-.092036	.1268242	-0.73	0.468	-.340607	.1565349
L2.	-.076554	.1042869	-0.73	0.463	-.2809526	.1278446
ictd_revres						
L3.	.6759945	.1683795	4.01	0.000	.3459767	1.006012
ictd_nontax						
L1.	-.653895	.1578114	-4.14	0.000	-.9631997	-.3445902
L2.	-.4233492	.2241037	-1.89	0.059	-.8625844	.015886
wdi_taxrev						
L1.	-.2150168	.1338517	-1.61	0.108	-.4773613	.0473278
L2.	.1555477	.1632253	0.95	0.341	-.1643679	.4754634
L3.	-.3428482	.1722967	-1.99	0.047	-.6805436	-.0051529
_cons	53.87978	6.982612	7.72	0.000	40.19412	67.56545

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Unstructured				
var(trend)	3.27924	1.367296	1.448306	7.42482
var(trend2)	.0642941	.0306106	.0252879	.1634669
var(_cons)	68.26668	37.00263	23.59563	197.5086
cov(trend,trend2)	-.3957826	.1952683	-.7785015	-.0130637
cov(trend,_cons)	4.505588	8.355495	-11.87088	20.88206
cov(trend2,_cons)	-1.388007	1.251841	-3.841571	1.065556
var(Residual)	.1577181	.0444632	.0907642	.2740617

LR test vs. linear model: chi2(6) = 190.58   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	102	.	-187.0161	38	450.0323	549.7812

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Alternative model 3:

Alternative model 3 replaces the covariance structure of the other models with the default identity to accommodate an AR(1) residual structure.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode: trend trend2, residuals(ar 1,t(year)) mle

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     254.14
Log likelihood = -202.07506                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	1.527923	.9156579	1.67	0.095	-.2667334	3.32258
trend2	-.1224882	.3683302	-0.33	0.739	-.8444021	.5994258
trend3	.009375	.0439702	0.21	0.831	-.076805	.095555
centered_incsh10h	.2301005	.2146211	1.07	0.284	-.190549	.6507501
wdi_pop1k	-.0000327	.0001066	-0.31	0.759	-.0002416	.0001761
FailedNeighbour	14.23021	9.936905	1.43	0.152	-5.245768	33.70618
Landlocked	12.92771	4.911505	2.63	0.008	3.301333	22.55408
wdi_gdpcapcur	-2.73e-06	.0000773	-0.04	0.972	-.0001542	.0001487
wdi_gdpcapgr	-.0663217	.027273	-2.43	0.015	-.1197757	-.0128677
fe_etfra	29.6434	10.52727	2.82	0.005	9.010325	50.27648
bmr_demdur	-.3577385	.0459181	-7.79	0.000	-.4477363	-.2677406
bdist	.000299	.0280965	0.01	0.992	-.0547692	.0553671
capdist	.0087753	.0135784	0.65	0.518	-.017838	.0353886
ttime	-.0073493	.0185253	-0.40	0.692	-.0436582	.0289595
ictd_revres	.1446314	.1832387	0.79	0.430	-.2145099	.5037727
ictd_nontax	-.5551903	.254966	-2.18	0.029	-1.054914	-.055466
wdi_taxrev	.45618	.1692454	2.70	0.007	.1244651	.7878949
c.ictd_nontax#c.centered_incsh10h	.0006454	.0415182	0.02	0.988	-.0807289	.0820196
c.wdi_gdpcapcur#c.centered_incsh10h	1.12e-07	8.68e-06	0.01	0.990	-.0000169	.0000171
c.bdist#c.centered_incsh10h	-.0014005	.0016143	-0.87	0.386	-.0045646	.0017635
c.FailedNeighbour#c.centered_incsh10h	1.701242	1.776982	0.96	0.338	-1.781579	5.184064
wdi_gdpcapgr						
L3.	-.2750368	.0854392	-3.22	0.001	-.4424944	-.1075791
centered_incsh10h						
L1.	-.0010713	.143675	-0.01	0.994	-.2826692	.2805266
L2.	-.0323276	.1322594	-0.24	0.807	-.2915512	.226896
ictd_revres						
L3.	.6075084	.2131772	2.85	0.004	.1896888	1.025328
ictd_nontax						
L1.	-.6715362	.1763862	-3.81	0.000	-1.017247	-.3258257
L2.	-.2941097	.2663546	-1.10	0.270	-.8161551	.2279356
wdi_taxrev						
L1.	.0196677	.1539454	0.13	0.898	-.2820596	.3213951
L2.	.1624711	.2108218	0.77	0.441	-.250732	.5756743
L3.	-.358777	.2120807	-1.69	0.091	-.7744475	.0568934
_cons	52.55798	7.724527	6.80	0.000	37.41818	67.69777

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Independent				
var(trend)	.78393	.3733132	.3082683	1.993543
var(trend2)	.0033209	.0096651	.0000111	.9966799
var(_cons)	.0036892	.	.	.
Residual: AR(1)				
rho	.9919906	.0039465	.9790169	.9969551
var(e)	52.87267	18.33203	26.798	104.3182

```
LR test vs. linear model: chi2(4) = 162.68                Prob > chi2 = 0.0000
```

BIC: 566.0242

Alternative model 4:

This model sacrifices the trend and trend2 from the random part, for a combination of an AR(1) residual structure and an unstructured covariance structure.

STATA command: mixed ffp_fsi trend trend2 trend3 centered_incsh10h wdi_pop1k FailedNeighbour Landlocked wdi_gdpcapcur wdi_gdpcapgr fe_etfra bmr_demdur bdist capdist ttime ictd_revres ictd_nontax wdi_taxrev c.ictd_nontax#c.centered_incsh10h c.wdi_gdpcapcur#c.centered_incsh10h c.bdist#c.centered_incsh10h c.FailedNeighbour#c.centered_incsh10h l3.wdi_gdpcapgr l.centered_incsh10h l2.centered_incsh10h l3.ictd_revres l.ictd_nontax l2.ictd_nontax l.wdi_taxrev l2.wdi_taxrev l3.wdi_taxrev|| ccode:, residuals(ar 1,t(year)) covariance(unstructured) mle

```
Mixed-effects ML regression                     Number of obs     =        102
Group variable: ccode                           Number of groups  =         23

                                                Obs per group:
                                                              min =          2
                                                              avg =        4.4
                                                              max =          6

                                                Wald chi2(30)     =     254.92
Log likelihood = -209.95813                     Prob > chi2       =     0.0000
```

ffp_fsi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trend	.4002942	1.103455	0.36	0.717	-1.762437	2.563026
trend2	.4296197	.4622697	0.93	0.353	-.4764123	1.335652
trend3	-.0608614	.0559558	-1.09	0.277	-.1705327	.0488098
centered_incsh10h	.4942455	.2730021	1.81	0.070	-.0408287	1.02932
wdi_pop1k	-.0000332	.0000971	-0.34	0.732	-.0002235	.0001571
FailedNeighbour	14.8492	8.868157	1.67	0.094	-2.532071	32.23047
Landlocked	10.64869	4.475407	2.38	0.017	1.877051	19.42032
wdi_gdpcapcur	-.0001464	.0000953	-1.54	0.125	-.0003332	.0000405
wdi_gdpcapgr	-.0350849	.0363019	-0.97	0.334	-.1062354	.0360656
fe_etfra	28.19564	9.860488	2.86	0.004	8.869438	47.52184
bmr_demdur	-.2860031	.0446784	-6.40	0.000	-.3735711	-.198435
bdist	.0054347	.0269494	0.20	0.840	-.0473851	.0582545
capdist	.0069035	.0127301	0.54	0.588	-.018047	.031854
ttime	-.0106116	.0175482	-0.60	0.545	-.0450055	.0237822
ictd_revres	.0859007	.2296273	0.37	0.708	-.3641606	.5359619
ictd_nontax	-.1364655	.3342395	-0.41	0.683	-.7915629	.5186319
wdi_taxrev	.4660875	.2110986	2.21	0.027	.0523419	.8798331
c.ictd_nontax#c.centered_incsh10h	.0430636	.052826	0.82	0.415	-.0604734	.1466006
c.wdi_gdpcapcur#c.centered_incsh10h	-.0000163	.0000103	-1.57	0.116	-.0000365	4.00e-06
c.bdist#c.centered_incsh10h	-.0015437	.0019481	-0.79	0.428	-.0053619	.0022745
c.FailedNeighbour#c.centered_incsh10h	1.775376	2.109974	0.84	0.400	-2.360097	5.910849
wdi_gdpcapgr						
L3.	-.2508718	.1074176	-2.34	0.020	-.4614064	-.0403373
centered_incsh10h						
L1.	.0075191	.1592525	0.05	0.962	-.3046101	.3196483
L2.	-.0018789	.1585046	-0.01	0.991	-.3125422	.3087844
ictd_revres						
L3.	.5084581	.2683264	1.89	0.058	-.017452	1.034368
ictd_nontax						
L1.	-.6796908	.2246049	-3.03	0.002	-1.119908	-.2394733
L2.	-.3386766	.2728651	-1.24	0.215	-.8734825	.1961292
wdi_taxrev						
L1.	.0532661	.1995869	0.27	0.790	-.3379171	.4444493
L2.	.070895	.2698864	0.26	0.793	-.4580727	.5998627
L3.	-.3917118	.2559275	-1.53	0.126	-.8933206	.1098969
_cons	53.98174	7.75	6.97	0.000	38.79202	69.17146

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
ccode: Identity				
var(_cons)	2.93e-13	.	.	.
Residual: AR(1)				
rho	.9806272	.0070562	.9605688	.9905312
var(e)	44.88999	14.02692	24.33166	82.81847

LR test vs. linear model: chi2(2) = 146.91   Prob > chi2 = 0.0000

BIC: 572.5404