Modelling Terrorist Attacks

Daan Koopman

Thesis Submitted to the
Delft Institute of Applied Mathematics
in Candidacy for the Degree of
Master of Science

Graduation Committee

Dr. P. Cirillo
Dr. R.J. Fokkink
Prof. Dr. C.W. Oosterlee

June 2018

TU Delft Delft University of Technology

Abstract

This thesis adds to quantitative literature on terrorism by examining the relationship between various annual country statistics and the number of terrorist attacks. In addition, it assesses the potential of forecasting terrorism. Combining an extensive review of literature from social science, with data analysis of the Global Terrorism Database, results in a specific selection of attacks on country-level, along with various factors that allegedly affect terrorism. It is shown that, under certain conditions, the countries can be fit by a nonhomogeneous Poisson process model with a Weibull baseline intensity and a piecewise constant covariates component. For three out of five countries, the in-sample results are satisfactory, providing a tentative answer to the general debate whether terrorism can be explained by root causes. Out-of-sample results for Afghanistan and Somalia are promising for future research and implementation in policy making.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 General Introduction

Terrorism is one of the biggest challenges facing the world today. Over the last decade, it has reached unprecedented heights with dramatic consequences. Not surprisingly, social science produced a vast amount of literature that elaborates this issue. The fact that there is no general consensus on the definition of terrorism, reflects the complexity of the phenomenon. Moreover, scholars are divided when it comes to explaining its causes. Quantitative research has emerged in recent years to add to the debate.

The emphasis of this study is combining social science study of terrorism, data analysis and mathematics, to use a theoretically motivated model with two clear goals. First, to examine the relationship between the number of terrorist attacks and various annual country statistics that represent demographic, economic, environmental and sociopolitical phenomena prevalent in society. Secondly, the potential of forecasting the number of attacks is assessed.

The initial chapter examines terrorism from a social science perspective. It covers a thorough evaluation of the definition of terrorism according to the used data source, as well as an extensive review of literature on causes of terrorism. As it turns out, scholars disagree whether simple factors can explain, let alone predict, terrorism. A general methodological approach to explaining terrorism by root causes is put forward, that proposes to focus on a specific selection of attacks.

Chapter 3 introduces the Global Terrorism Database (GTD), the main data source of this study. The GTD provides a wide array of information for domestic, transnational and international terrorist attacks around the world. The database has undergone several changes in data collection methodology, which are discussed as they had impact on the quality of the data. By means of general statistics and data visualisation, the possibilities within the database are explored. Analysing attacks throughout the world, leads to a final focus of modelling a specific selection of attacks for six countries that experience high frequencies of Islamic terrorism.

Modelling the number of attacks is done with a nonhomogeneous Poisson process model with a Weibull baseline intensity, introduced in chapter 4. By means of statistical tests and graphical evaluations, it is shown that under certain conditions, this model is appropriate for the country data. In hindsight, the poor results of modelling countries simultaneously in a proportional intensity framework, suggested to model countries individually. In that case, however, covariates are only able to improve the model if they are allowed to vary over time. Since country statistics are available on an annual level, the model is adjusted to allow piecewise constant covariates. The log-likelihood function, required for maximum likelihood estimation, is derived. Finally, both strengths and limitations of the model are discussed.

The annual country statistics that represent the alleged causes of terrorism are introduced in chapter 5. First, the correlation between terrorist activity and the country statistics over time is compared to literature from social science. Secondly, the statistics that have the potential to improve the model fit are selected. Sets of covariates are then presented to be implemented in the model.

The model results are presented in chapter 6. First, the general model selection and forecasting procedure is introduced. Both in-sample and out-of-sample results are then evaluated for three countries. When the out-of-sample results allow it, forecasts are made and interpreted.

The general conclusion 6 summarises the findings obtained over the course of the thesis and assesses to what extend the main goals are achieved. A tentative answer is given to the general debate whether root causes are able to explain terrorism. The potential of forecasting terrorism is discussed, along with the possibilities of implementation in policy making. Finally, opportunities for future research are presented.

1.2 Literature Study

There are a large number of empirical studies on the relationship between terrorist attacks and country characteristics. In pursuit of a comprehensive, theoretical motivation for the covariate data of the model, section 2.3 discusses a wide range of these studies. For an overview of quantitative studies on causes of terrorism is referred to Krieger and Meierrieks [139].

While quantitative studies on causes of terrorism are abundant, literature on forecasting terrorist attacks is very limited. The main literature found is that of Bakker et al. [121], which uses a Bayesian multi-level, Poisson log-normal mixture model to forecast attacks for all countries in the world over the period 1970-1996. They find that, though the model forecasts are reasonable, the results do not allow application in practice. One of the reasons posited, is the fact that country data can be scarce. In order to improve out-of-sample results, a different approach is suggested that estimates either each country, or a group of neighboring countries, individually.

Newman explores several methodologies to explaining terrorism by root causes and finds that a more specific selection of attacks has more potential for explaining and predicting terrorist activity [115]. Therefore, this study steps away from large cross-national samples and models a more specific selection of attacks.

The statistical model of this study is primarily based on Lawless [91]. He introduces a proportional intensity Poisson process model, to deal with the setting in which individuals experience recurrent events, where the data of each individual constitutes to a point process, along with covariates. Repeated events are assumed to follow a nonhomogeneous Poisson process with a multiplicative intensity function consisting of a baseline intensity function and a function of the covariates and regression coefficients. Both parametric and semiparametric approaches to model fitting are covered. The selected data will give reason to assume that the baseline intensity follows a Weibull model and therefore the parametric approach is pursued. The initial idea of modelling countries simultaneously was dropped after disappointing model results. For one individual, the proportionality assumption naturally holds resulting in a NHPP model with multiplicative intensity.

The brief preliminaries to this model in section 4.2, introduce the Poisson process and Poisson regression. A general introduction to probability models and stochastic processes can be found in Ross [136] and [135], respectively. An introduction to Poisson regression analysis is given in chapter five of [47] by Tai *et al.*

Chapter 2

A Social Science Understanding of Terrorism

Social science is essential to understanding terrorism. This chapter reviews literature from scholars and organisations on the definition and causes of contemporary terrorism. A brief review of the evolution of terrorism provides historical and political context. It will become apparent that terrorism is a dynamic and partially subjective concept; two reasons why the phenomenon is so complex.

This study relies on data from the Global Terrorism Database (GTD) and therefore has to adopt the definition according to which the attacks are included in the database. Section 3.3 thoroughly evaluates that definition by comparing it to the differing views of leadings international organisations.

Now the definition of terrorism for this study is clear, section 2.3 determines a methodological approach to explaining terrorism by underlying factors prevalent in society. Finally, an extensive overview of underlying causes of terrorism posited by scholars, is presented.

2.1 Evolution of modern terrorism

One of the most influential works in terrorism studies is "The Four Waves of Modern Terrorism" by David Rapoport [65]. It categorises the evolution of modern terrorism (1880-present) in four time periods defined as waves, in which terrorist groups are linked based on their ideology, motives and strategies. As Rapoport writes, an essential aspect of a wave is the international character, i.e. that there exists a common energy that transcends national boundaries and shapes the characteristics and mutual relationships of terrorist organisations. Furthermore, it is important to note that the waves represent main global trends; various terrorist groups exist and operate independent of the waves or exist longer than the wave they belong to. Many groups and conflicts are not explicitly mentioned, but their significance is not triviliazed in any way.

The terrorism data source used in this study includes terrorist attacks from 1970 through 2014, see chapter 3. This represents the last ten years of the third wave and the fourth wave up to 2004, the year Rapoport's work was published. A brief summary of the first three waves is found in appendix A. The emphasis is placed on the fourth wave, which is complemented to include key political events of recent years.

2.1.1 Religious wave (1980-Present)

Although the fourth wave is specifically characterised by religion, religious elements have also been important in previous waves. Rapoport explains this by noting that religious and ethnic identities often overlap. This is exemplified by conflicts such as the Armenian, Irish and Israeli-Palestinian conflicts. The concerning terrorist organisations had the aim to create secular states. In the fourth wave, this shifted from secular motivations towards more religious and theological ideologies [43]. Rapoport adds that religion "...started to supply justifications and organising principles for a state". However, scholarly positions on the relation between religion and contemporary terrorism diverge, as discussed in section 2.3.3.1.

Religious terrorism in the fourth wave is predominantly Islamic; over the period 1968 to 2005 alone, Islamic groups were responsible for 93.6% of all attacks coming from religious groups [90]. In addition, this section will show the political events that sparked the fourth wave originated in Islam. However, given the rising tension between the Muslim and non-Muslim population in the (Western) world, illustrated by the increasing popularity of far-right groups such as Front National in France, Partij Voor de Vrijheid in the Netherlands and Alternative für Deutschland in Germany, it needs to be emphasized that Muslims are also the group most victimised of terrorism [114]. When referring to Islam being the dominant religion in the fourth wave, it should be clear that the radical Islam of terrorists does not equal Islam of the vast majority of the Muslim population.

Rapoport identifies three decisive events in 1979 as political turning point, crucial for the launch of the religious wave. The first is the Iranian revolution in 1979. As Axworthy writes, demonstrations against the Shah Mohammad Reza over "living conditions, pay cuts, and the threat of unemployment fused with the general disillusionment and anger with the regime" [104]. Using the discontent in society, the grip of the clergy on the revolution kept increasing, with ayatollah Ruhollah Khomeini as key figure. The struggle of Khomeini against the Shah lasted many years and fits into the tradition of Iranian clerics who oppose the secular authority [99]. In 1979 Khomeini took over power in Iran, but his message of the liberation of Muslims from the control of the West and the Soviets, spread to Muslims, both Sunnis and Shiitis, worldwide [42]. This was amplified by a second important event, that 1979 was the year of a new Islamic century, where one tradition holds that a redeemer will come at the start of a new century.

The third decisive event was the invasion of the Soviet Union in Afghanistan at the end of 1979. In response, Palestinian Sunni Abdullah Yusuf Azzam issued a fatwa, *Defence of the Muslim Lands, the First Obligation after Iman* [32], where he proclaimed Muslims should wage jihad when Islamic territory gets occupied by infidels as the Western capitalists or the Eastern communists [73]. An estimated 20000 Muslims answered to the call to join the fight against the Soviet Union in Afghanistan [42]. Rapoport underscores that this was subsidized by the U.S. (with both arms and money), but he refrains to mention the equally significant contribution of Saudi Arabia [84]. This is important since Saudi Arabia kept financing terrorism until today [45]. The jihad forced the Soviets out by 1989: religion eliminated a secular superpower, which was the decisive event for the third wave to fade out. As Rapoport notes, countries in Central Asia formerly part of the Soviet Union - such as Uzbekistan, Tajikistan and Azerbaijan - became important suppliers for Islamic militants. Afghan veterans became major participtants in the upcoming and ongoing conflicts.

The role of the U.S. changed again. Called 'Great Satan' by Iran and regarded as main enemy of Al-Qaeda, the country became a target. This started with attacks by Al-Qaeda, again funded by Saudi Arabia [45], on U.S. military posts in the Middle East and Africa, with the aim to force U.S. military out of the Middle East. Al-Qaeda's leader Bin Laden shifted the attacks to the heart of the West, resulting in attacks on U.S. soil in 1993. Two attacks in Kenya and Tanzania in 1998 provoked repercussions against al-Qaeda. For the first time missiles were used against a terrorist organisation, with as consequence that bin Laden turned into a global celebrity. He inspired youth worldwide to join the fight for global Jihad, which resulted in attacks around the world [42].

In response to the September 11 attacks, the U.S. invaded Afghanistan at the end of 2001, with the aim to oust the Taliban from Afghanistan and to dismantle al-Qaeda. In March 2003 the U.S. invaded Iraq, despite the fact that there was no evidence of Iraq possessing weapons of mass destruction [89]. With this invasion the U.S. created the preconditions for radical Sunni groups, such as Islamic State of Iraq (ISI), to establish.

Iraq was a Sunni minority regime that surpressed the Shiite majority. The U.S. was working on replacing Saddam Hussein's secular state by democracy and since the Shiites were the majority, it was replaced by a predominantly Shiite administration [108]. This was led by Nouri al-Maliki, who formed a sectarian intolerant regime where hundreds of thousands of Sunnis lost their their jobs. The Islamic State in Iraq (ISI) used this to strengthen their movement, which was later led by Abu Bakr al-Baghdadi in 2010. The U.S. detained al-Baghdadi in Camp Bucca back in 2004, where he met salafists, generals and colonels of Hussein [42], who later became the framework of IS's effective armed forces.

Because of the civil war in Syria, ISI expanded to Syria in 2013, where the group became known as Islamic State of Iraq and Syria (ISIS). Section 2.3.3.3 covers the causes of the emergence of ISIS in more detail. On 29 June 2014, ISIS declared itself a caliphate and renamed itself to Islamic State (IS) [12].

In Afghanistan, the U.S. American and NATO Troops' attempt to bring stability could not prevent the Taliban from gaining strength. Especially since foreign combat troops left in 2014, the Taliban has rapidly increased control over territory.

Meanwhile, Islamic terrorism is becoming more dominant around the world. It is out of the scope of this study to discuss particular conflicts in detail, but the main events that led to uprise of the Islamic terrorism the world is facing today, have been covered. The fourth wave saw several distinctive features:

- Suicide bombing became the most deadly tactical innovation. Although this might primarlity be associated with Islamic terrorism because of the reward of paradise, the Sri Lankan [61] Liberation Tigers of Tamil Eelam, one of the exceptions of secular groups in the fourth wave, used it more than any other group.

- Since a religious community is much larger than a national group, groups from the fourth wave increased in size and the number of terrorist groups declined. The main exception here is the PLO, that tried to transform itself into a regular army in Lebanon, where it had around 25000 members.

- Islamic groups exist longer than groups from previous waves. The previous waves however, lasted about a generation (40 years). If history repeats itself, then the fourth wave ends around 2020. Unfortunately, Rapoport sees reason to believe that religious terrorism will last longer than its predecessors, as he writes: "The life cycle of its predecessors may mislead us. Each was inspired by a secular cause, and a striking characteristic of religious communities is how durable some are". As mentioned this was published in 2004. Our dataset reveals that, considering the uprising trends from figure 3.7, it is unlikely Islamic terrorism will die out in the upcoming years. Our model's forecasts can provide more insight on this matter.

  could give more insight on this matter.

2.1.2 Summarising remarks

Analysing the course of modern terrorism led to the following findings:

- The concept of terrorism changes over time and inevitably contains subjective elements.
- The first wave was triggered by failure of a democratic reform program, self-determination inspired the second, the existing surpressing, not truly democratic systems led to the third wave. Rapoport gives a worrisome account of the nature of the fourth wave as he writes: "The spirit of the fourth wave appears explicitly antidemocratic because the democratic idea is inconceivable without a significant measure of secularism".
- The GTD contains attacks from 1970 through 2014, thus the database primarily covers the religious wave, dominated by Islamic terrorism. This will turn out to be decisive for the focus of the study.
- Several preconditions for terrorism can be destilled from the analysis of the fourth wave. Discontent and resentment caused by harsh living conditions, inequality, corruption and unemployment, fuels the radical narrative. A clash of cultures and foreign interventions lead to hatred against modernisation. This study tries to incorporate such factors in the model to improve its accuracy. Section 2.3 provides an extensive analysis on proposed causes of terrorism.
- The first three waves lasted about a generation (40 years). If history repeats itself, the fourth wave ends around 2020. However, Rapoport reasoned religious terrorism will last longer than secular terrorism, which the data until 2014 seemed to agree with. The model's forecasts can hopefully provide a more definite answer.

2.2 Definition

To this day, there is no general consensus on the definition of terrorism. The essence revolves around an act of violence, with the intention to cause fear in order to achieve a certain goal. Formulating a definition has proven to be very complex, which resulted in a vast amount of literature of which a concise overview can be found in "The Numerous Federal Legal Definitions of Terrorism" [117]. A compelling reason for the complexity might be the changing nature of the term [44], as observed in section 2.1.

Perhaps more important is the judgemental nature of the word, as Cooper explains: "...it not only encompasses some event produced by human behavior but seeks to assign a value or quality to that behavior" [78]. Because of the pejorative nature of modern day terrorism, using the word terrorism, and therefore morally judging, has as consequence that "...if one party can successfully attach the label terrorist to its opponent, then it has indirectly persuaded others to adopt its moral viewpoint" [50]. In addition, there is the subjectivity of the term, since "...its meaning differs, depending on the individuals describing it, their motives, and when it is defined" [72]. The difficulty is to determine when the use of violence is legitimate, or when it is terrorism.

This study is bound to the definition according to which the Global Terrorism Database (GTD) records its attacks. The goal of this section is therefore not to formulate a comprehensive definition, but rather to get a better understanding of the strengths- and limitations of the GTD's definition, by comparing it to those from prominent international organisations.

First of all, the League of Arab States, governing the region most affected by terrorism, see table 3.1, depicts terrorism in the Arab Convention for the Suppression of Terrorism [6]:

> "Any act or threat of violence, whatever its motives or purposes, that occurs in the advancement of an individual or collective criminal agenda and seeking to sow panic among people, causing fear by harming them, or placing their lives, liberty or security in danger, or seeking to cause damage to the environment or to public or private installations or property or to occupying or seizing them, or seeking to jeopardize national resources"

Amnesty International criticized this definition, stating it "...can be subject to wide interpretation and abuse, and in fact does not satisfy the requirements of legality in international human rights and humanitarian law" [85]. The act or threat of violence leaves much room open to interpretation. First of all, Amnesty questions that the Convention does not define violence itself, noting "...harsh penalties for committing crimes of 'terrorism' under the pretext that the acts were 'violent' " could be imposed. Secondly, including the threat of violence in the definition could be used "...against people who are not accused of committing violence, but for their alleged affiliation with certain political opposition parties that use violence, since such affiliation may be seen as threat to commit an act of violence". Furthermore the inclusion of "damage to ... any public facility or public ... property" creates the possibility that conduct not violating international humanitarian law in non-international armed conflict, can be characterised as terrorism. When legitimate acts of war turn into crimes of terrorism, "...members of armed political groups will lose an important incentive to comply with international humanitarian law, since they will feel that whatever they do will be perceived as a crime of international concern". Finally, when characterized as terrorist, there is the risk that goverments will stop seeing the perpetrator as protected by international humanitarian law.

The current European Union approach can be found in the Council Common Position 2001/931/CFSP as well as the Council Framework Decision 2002/475/JHA, defining terrorism as an act

> *"...committed with the aim of 'seriously intimidating a population', 'unduly compelling a government or international organisation to perform or abstain from performing any act', or 'seriously destabilising or destroying the fundamental political, constitutional, economic or social structures of a country or an international organisation' "*

Similarly, this definition criticised for being too broad by several human rights organisations such as Human Rights Watch, European Network Against Racism, Amnesty International and others [86]. As they indicate, it permits states to criminalise public protests or other peaceful acts that they regard "seriously destabilise the fundamental political, constitutional, economic or social structures of a country or an international organisation" as terrorism, hence politically and socially degraded such protests or acts [117].

The General Assembly of the United Nations, the only organ where every 193 member states have equal representation (one nation, one vote), has thus far been unable to formulate an internationally agreed definition of terrorism. The UN acknowledged this is needed "...for consensus on completing comprehensive convention against it" [134], opening the way for global cooperation to prevent and combat terrorism. Its High-level Panel on Threats, Challenges and Change, concluded in 2004 that a consensus definition of the General Assembly should include the following elements [79]:

(a) recognition, in the preamble, that State use of force against civilians is regulated by the Geneva Conventions and other instruments, and, if of sufficient scale, constitutes a war crime by the persons concerned or a crime against humanity;

(b) restatement that acts under the 12 preceding anti-terrorism conventions are terrorism, and a declaration that they are a crime under international law; and restatement that terrorism in time of armed conflict is prohibited by the Geneva Conventions and Protocols;

(c) reference to the definitions contained in the 1999 International Convention for the Suppression of the Financing of Terrorism and Security Council resolution 1566 (2004);

(d) description of terrorism as "any action, in addition to actions already specified by the existing conventions on aspects of terrorism, the Geneva Conventions and Security Council resolution 1566 (2004),

> that is intended to cause death or serious bodily harm to civilians or non-combatants, when the purpose of such act, by its nature or context, is to intimidate a population, or to compel a Government or an international organization to do or to abstain from doing any act"

In (a), clear emphasis is placed on recognition that states' use of force against civilians is covered by international humanitarian law and does not need to be included in a definition of terrorism. The conventions referred to in (b) cover legal issues around terrorism, i.a. limiting state abuse and solidifying anti-terrorism cooperation. The Security Council resolution 1566 [59] named in (c), calls on states to take measure against acts similar to those described in (d). The description in (d), is similar to that of the EU, but with the addition of the intention to "...cause death or serious bodily harm to civilians or non-combatants" and the named conventions.

The GTD assesses acts of violence according to its own definition of terrorism. As stated in the GTD Codebook from June 2015 [15], the GTD defines a terrorist attack as

> ...the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.

In practice, this translates to definition 2.1

Definition 2.1 (GTD's definition of terrorism)**.** In order for an incident to be characterised as terrorism and therefore to be included in the GTD, it must satisfy the following two requirements

Requirement 1. The following three attributes must be satisfied:

Attribute 1 The incident must be intentional – the result of a conscious calculation on the part of a perpetrator.

Attribute 2 The incident must entail some level of violence or immediate threat of violence -including property violence, as well as violence against people.

Attribute 3 The perpetrators of the incidents must be sub-national actors. The database does not include acts of state terrorism

Requirement 2. At least two of the following three criteria must be met

Criterion 1 The act must be aimed at attaining a political, economic, religious, or social goal. In terms of economic goals, the exclusive pursuit of profit does not satisfy this criterion. It must involve the pursuit of more profound, systemic economic change.

Criterion 2 There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims. It is the act taken as a totality that is considered, irrespective if every individual involved in carrying out the act was aware of this intention. As long as any of the planners or decision-makers behind the attack intended to coerce, intimidate or publicize, the intentionality criterion is met.

Criterion 3 The action must be outside the context of legitimate warfare activities. That is, the act must be outside the parameters permitted by international humanitarian law (particularly the prohibition against deliberately targeting civilians or non-combatants).“

Attribute 1 is clear, the attack must be the result of a conscious calculation, hereby excluding spontaneous incidents. Attribute 2 does not require unlawful violence. Combining this with the fact that two out of three criterea suffice to meet requirement 2, reveals that the GTD incorporates acts legal according to international humanitarian law. An example of this are Palestinan resistance acts on Isreali military, which included in the GTD, see [76]. The inclusion of property violence involves the same risk. Recall Amnesty International critisised the Arab Convention for the Suppression of Terrorism for, warning "...members of armed political groups will lose an important incentive to comply with international humanitarian law, since they will feel that whatever they do will be perceived as a crime of international concern". Finally, Attribute 2 complements the Arab Convention with the more direct 'immediate' threat of violence, this is however still senstive to interpretation and needs careful application. Attribute 3 represents clear exclusion of state terrorism from the database, as in (a) from the UN's definition outline.

Coinciding with the EU's definition, Criterion 1 requires the agenda of the perpetrators to be political, economic, religious or social. The essense of criterion 2 is that the immediate victims are merely a mean to reach an end. The last criterion 3 is referring to the Geneva Conventions and other instruments, covered extensively in the outline definition of the UN, separating terrorism from legitimate warfare activities. As mentioned, requirement 2 allows acts legal according to international humanitarian law to be included in the database. Criterion 3 should be absolutely imperative, which is impeded by the loophole of requirement 2.

Concluding, definition 2.1 possesses the key elements from the definitions of the EU and the (outline) UN. With its construction however, it induces a loophole which allows acts legal according to international humanitarian law to be included in the database. Fortunately, the explicit attributes and criterea are represented as seperate fields in the database and can be thus filtered. The absence of state terrorism might give a distorted view of conflicts in countries as Syria, but state use of force is sufficiently covered by the Geneva conventions. The inclusion of - 'immediate threat' of violence is vague and requires clarification.

In the beginning of chapter 3, terrorism follows definition 2.1. In a later stage, a selection of the data is proposed that corresponds to definition 2.2, which remains binding for the rest of this study.

Definition 2.2 (Research' Definition of Terrorism)**.** An act is defined as terrorism when it is an intentional level- or immediate threat non-state violence (including property violence) not legal according to international humanitarian law, with the intention to coerce, intimidate, or convey some other message to a larger audience than the immediate victims, aimed at attaining a political, economic, religious or social goal.

2.3 Root cause approach

2.3.1 Introduction

There exists a vast amount of literature on causes of terrorism. Crenshaw distinguishes between *preconditions* -factors also called *root causes* that set the stage for terrorism in the long run- and *precipitants*, specific events that more directly trigger terrorism [105]. Newman explains: "...certain conditions provide a social environment and widespread grievances that, whem combined with certain precipitant factors, result in the emergence of terrorist organisation and terrorist acts" [115].

Scholars do however not agree on whether root causes are capable of explaining terrorism, arguing they fail to "...accomodate the richness and diversity of situations that breed terrorism" [128]. Furthermore, root causes do not show a direct cause and effect [115]. Phenomena associated with terrorism, such as lack of democracy, income inequality and surpression, are found in many countries around the world where terrorism does not thrive. Finally, the diversity of conflicts would lead to case specific sets of root causes, making it hard to treat terrorism in general form. Combining the former arguments suggests that root causes alone are not sufficient to explain, let alone predict, terrorism. Section 2.3.2 discusses a methodological approach to explaining terrorism by root causes.

2.3.2 Methodological approach

Newman examines the analytical potential of various root causes, by testing them as explanatory variables for different methodological approaches [115]. First, he evaluates if there is relationship between the frequency of terrorist attacks coming from a wide range of samples and several key root causes. The results are disappointing, there is no significant correlation found. Thus, for a wide range of samples, root causes are not sufficient in explaining, let alone predicting, terrorism. Newman attributes this to the notion that there are a large number of variables that contribute to the frequency of terrorist attacks; the root causes tested are only a small part of the larger picture.

Terrorist organisations do not always conduct terrorist acts in the societies from which they emerged. As second methodology, Newman suggests identifying terrorism organisations on basis of their nature and aims, the background of their leadership and supporters, and their social base. The aim is to group together those organisations with similar sets of root causes. Newman finds better results, especially when focusing on societies from which the most deadly terrorist organisations emerged and are based. He concludes that root causes analysis may be useful for explaining terrorism by ideological, ethno-nationalist and Islamist groups in developing countries in the Global South.

Categorising terrorist organisations by the aforementioned characteristics, immediately separates those coming from distinct waves discussed in section 2.1. Since the GTD covers attacks from 1970 through 2014, it is rational to focus on organisations from the fourth, religious wave. This wave is largely dominated by Islamic groups [90], for which Newman does find a correlation between their activity and specific factors.

Bakker *et al.* use a Bayesian multi-level, Poisson log-normal mixture model to forecast attacks for all countries in the world over the period 1970-1996 [121]. Although the out-of-sample results are reasonable, a different approach is suggested, that estimates either each country, or a group of neighboring countries, individually. The idea of modelling a more specific selection of attacks is in line with Newman. Focusing on Islamic groups seems therefore a valid option.

2.3.3 Root causes

This section discusses underlying causes of terrorism posited by scholars. It serves as the theoretical motivation for chapter 5, that finds country statistics to be examined for their relationship with terrorist activity.

As a start, the role of religion itself is considered. In the analysis, globalisation is posited as an important factor in the emergence of Islamic terrorism, which is discussed next. Finally, various demographic, economic, environmental, political and social factors are elaborated.

2.3.3.1 Religion & Globalisation

Religion

Religious terrorism is characterised as being carried out with the purpose of "...influencing or coercing governments and/or populations towards saliently religious goals" [81]. But what is the role of religion itself in comptemporary terrorism?

Scholars posited different views on the relationship between religion and contemporary terrorism. Okoro destills four different arguments [118]. The first, mostly plead for by scholars of Western orientation, claims religion lies at the root of modern-day terrorism. The other side of the spectrum exonorates religion, claiming that terrorism is the consequence of socio-political, economical and demographic phenomena of which religion is the victim. A third group follows this line, however, it does not consider religion as victim, but rather as a socio-spiritual phenomenon which fans the flames of- and supplies justifications for contemporary terror. Okoro advocates a fourth argument that claims globalisation causes socio-political conflict that religion translates into a moral conflict.

Statistical analysis from Brenda- and James Lutz shows globalisation resorts in increased levels of terrorism, but emphasises that no single cause explains terrorism [49]. A more reasonable argument therefore acknowledges the prominent role of globalisation, while in addition recognising the importance of case specific economic, environmental, political and social circumstances. The role religion plays is uncertain, but should not be trivialised and is therefore included in the analysis.

Globalisation

Globalisation provides a wide array of opportunities to countries around the world through "...economic liberalisation, foreign investments and capital flows, technological exchange as well as information flows" [2]. As a result, an increasing number of countries is becoming integrated into the global economy, seeing faster economic growth and poverty reduction. However, the benefits of globalisation have not been been equally dispersed between regions, between nations and even within nations. The gains of the West far outreach those of other regions, especially Africa [7]. Furthermore, globalisation in its current form exacerbates inequality [9][127], leads to exploitation of labour [8] and has severe impact on the environment [62][34]. Inequality, human rights violations and environment will be discussed as individual factors.

In addition, globalisation influences culture. The homogenising influence of local culture has clear positive effects, for example that it promotes the integration of societies and has provided millions of people with new opportunities [143]. At the same time, contact with different value systems in traditional societies and communities can lead to the perception or fear of cultural imperialism [115], especially when exposed to rapid modernisation. In this light, Islamic fundamentalism could be seen as a response to this cultural clash; a rise against "modernity -the secular, scientific, rational and commercial civilisation created by the Enlightenment as it is defined by both its virtues (freedom, democracy, tolerance and diversity) and its vices (inequality, hegemony, cultural imperialism and materialism)" [51]. This sentiment is amplified by foreign presence, interventions and especially invasions such as those in Afghanistan and Iraq by the United States. Together with "changes in social and economic structures of Muslim countries" and catalysed by the events discussed in the fourth wave of section 2.1.1, this led to the rise of radical Islam [36].

Causes of emergence of specific Islamic terrorist organisations can of course deviate from this general picture. Section 2.1.1 discussed the emergence of Al-Qaeda and Islamic State, and confirms root causes of Al-Qaeda are globalisation, modernisation, foreign presence and invasion. In case of ISI, the emphasis seems to be more on the regime change as a consequence of the foreign invasion by the U.S., causing unemployment, and a feeling of humiliation as well as resentment in the Sunni minority. Outside of the Middle East, the emergence of Boko Haram in Nigeria is rooted in colonialism, resulting in a clash of values between Westernised South versus traditional Muslim North, combined with elements such as severe inequality, exclusion of Muslims [56] and weather/climate extremes [20].

In Southeast Asia, Rabasa recognises both external and domestic sources of Islamic radicalism [36]. The underlying external source is the change that the Muslim world has undergone by globalisation. The subregion saw "deterioration of economic and social conditions after the 1997-98" and domestic elements, such as an ethnic and religious shift (Philippines), Islamic minority resentment (Philippines), exclusion of political and economic opportunity (Thailand) and weakening government control (Indonesia and the Philippines) [109]. This combined produces "...the specific characteristics of each of the national variants".

Concluding, globalisation is a key element in Islamic terrorism that, combined with other case specific factors, creates an environment in which terrorism can thrive.

2.3.3.2 Economic Factors

One of the most debated root causes is poverty. As a multifaceted concept, poverty covers more than comparing income to the defined poverty line. It includes different social, political and cultural perspectives, such as the need for the provision by a community of the basic social services necessary to prevent individuals from falling into poverty [144], as well as social inclusion, which assesses societal and institutional factors that play important roles in determing living standards [142]. Comprehensive definitions can be found in the UNDP's Human Development Report 1997 [5].

Scholars disagree on the relationship between poverty and terrorism. On the one hand, scholars argue that poverty creates desperation and resentment, which according to Töpfer, Executive Director of the United Nations Environment Programme (UNEP), "...can fan the flames of hate and ignite a belief that terrorism is the only solution to a community's or nation's ills" [98]. Furthermore, educational programs and other terrorist prevention measures are often absent in poor societies, leading even more to an environment where terrorist groups can flourish.

On the other hand, scholars point out that the active supporters of most terrorist organisations are not poor and uneducated [115]. Bjørgo notes that at individual level, "...terrorists are generally not drawn from the poorest segments of their societies", arguing their education and socio-economic background is (above) average [137].

Conflicting empirical studies contribute to the controversy. Krieger and Meierrieks review numerous large country-sample studies and conclude that there is only weak evidence poor economic development causes terrorism [139]. Note that since empirical studies cannot work with the inclusive definition of poverty, due to lack of data which forces to use measures such as GDP per capita, the term economic development is used instead of poverty. However, studies that examine numerous countries over decades, rather than focusing on particular groups or individuals, overlook patterns that exist in particular cases. Sterman explains "Indeed, it is quite likely there are multiple routes into terrorism, some of which might involve poverty and some of which might not. When this data is aggregated, the poverty-related routes become less visible, but that does not mean they don't exist" [64]. Piazza agrees that it is possible that short-term and geographically isolated conditions "...can distort the overall picture and mask more universal predictors of transnational terrorism".

Although questionable as root cause, poor economic development, or poverty, is incorporated in the statistical analysis.

Related to poverty, is economic inequality. Economists generally distinguish three metrics of economic inequality: income, wealth and consumption [69]. Research focuses mainly income inequality, primarily because this metric has the most comprehensive data available. Empirical literature supports a significant positive relationship between income inequality and terrorism. Krieger and Meierrieks (114 countries, 1985-2012) find higher levels of income inequality are associated with more terrorism and conclude "...frustration over the distribution of income within a society, resulting in terrorism to voice dissent and achieve a redistribution of wealth" [140]. Shahbaz *et al.* find this significance for domestic terrorism in Pakistan, for attacks between 1972 and 2010 [112]. Goldstein confirms income inequality has significant positive correlation with terrorist risk for a 105 country sample [100]. Ezcurra and Palacios conclude that high interregional inequality, i.e. income inequality across regions within a country, increases the number of domestic terrorist attacks (48 countries, 1990-2010) [122].

Unfortunately empirical literature on the relationship between wealth/consumption inequality and terrorism was not found. Prominent economist Thomas Piketty does claim economic inequality plays a big part in Islamic terrorism in the Middle East, but this is strongly critisised for lack of evidence [77].

Income inequality is interesting furthermore an interesting factor because the regions that experience high levels of Islamic terrorism are very unequal [102]. The factor is therefore included in the statistical anaylsis.

2.3.3.3 Environmental Factors

Most of the environmental factors that influence terrorism are either caused or exacerbated by climate change. Therefore, this section focuses on the effects of climate change and incorporates various additional environmental phenomena along the way.

Over the past decade, a large amount of literature on the relationship between climate change and conflict has emerged, of which a part focuses specifically on terrorism. Studies commonly assert that climate change acts as a threat multiplier, rather than as a root cause of terrorism [11][20]. One such study is a collaborative report by the German Federal Foreign Office and Adelphi published in 2016. It examines the links between climate change and various non-state armed groups (NSAGs), on basis of four case studies [20]. NSAGs are not necessarily terrorist groups and vice versa, but the four case studies include three major Islamic terrorist organisations (IS, Tabiban and Boko Haram), making the report a valid reference. Two case studies are discussed to understand how climate change affects terrorism.

The first examines how the impact of climate change facilitates the rise of Boko Haram in the region around Lake Chad. Located in the Sahel zone of West-Central Africa, Lake Chad is a freshwater lake bordered by Cameroon, Chad, Niger and Nigeria. In 2012, The UN Food and Agriculture Organisation estimated the lake provides livelhoods for a rapidly growing population of 30 million people [38]. The ethnic and religious divisions of the more than 70 ethnic identities living in the region, have been an important factor in the long history of violent conflict. The dramatic 90 percent decrease of the lake's surface over the last 50 years, contributed to resource scarcities that aggravated poverty, unemployment, social tension, violent conflict and population displacement. The UNEP estimated in 2008 that half of this reduction is caused by human water use, whereas the other half can be attributed to climate change [35].

Two conclusions can be drawn from the analysis of the conflict. First, climate change increases and intensities resource conflicts around Lake Chad, exacerbating "...existing fragility in areas where state presence and authority are low" allowing Boko Haram to thrive. Second, it deteriorates livelihoods, creating critical living conditions where people are vulnerable to terrorist recruitment. Although climate change plays an important role in the humanitarian crisis in the Lake Chad region, the case study emphasises other important drivers, such as marginalisation, identity and corruption.

The second case study examines the role of climate change in the emergence of ISIS in Syria. The conflict in Syria preceding ISIS, was the result of "...converging pressures including government failure" [20], sectarian grievances and a combination of extreme weather events (such as the mega-drought), water resource mismanagement and rapid population growth. This led to widespread crop failure and consequently mass urbanisation and high unemployment [55], which weakened the Assad regime, facilitating the rise of terrorist organisations such as ISIS and Al-Nusra. The important consequence was the vulnerability to terrorist recruitment by those affected of the livelihood insecurity and water scarcity. ISIS exploited this by providing clean water, economic perspective and various services to the population.

The case studies show two main mechanisms by which climate change facilitates the emergence- and growth of NSAGs. First, climate change is increasingly contributing to fragility by catalysing conflicts surrounding natural resources and increasing livelihood insecurity. Second, climate change deteriorates livelihoods, which creates opportunity for recruitment by terrorist organisations. The report therefore concludes climate change acts as a threat multiplier, for "...it interacts and converges with other existing risks and pressures in a given context and can increase the likelihood of fragility or violent conflict".

Concluding, the inclusion of an explanatory variable that represents climate change, or environmental phenomena caused by climate change, could shape the model and should be considered.

2.3.3.4 Demographic Factors

The population of countries that experience Islamic terrorism has increased rapidly in the 20th century. For determining the consequences of this phenomenon, the Middle East and North Africa (MENA) region is examined. Especially from the 1950s onward, the MENA region faced rapid population growth [119]. According to Coccia this "seems to be the basic for the source and evolution of terrorism". His study finds rapid population growth combined with "...acute environmental and socioeconomic stressors", can lead to terrorism, which corresponds to both case studies discussed in section 2.3.3.3.

Following this line of thought, the question is how rapid population growth contributes to terrorism. Homer-Dixon states one of terrorism's underlying factors in the Middle East is its dramatic increase in population, because it results in a large amount of urbanised and unemployed young men [138]. The United Nations assessed that in 2011, one in five people in the MENA region was aged between 15 and 24 [71]. Furthermore, the region experiences the highest youth unemployment rates in the world, in 2014 a dramatic 28.2% in the Middle East and 30.5% in North Africa [17]. Alienated and unemployed, young men with no prospects are vulnerable to terrorist recruitment. Empirical Research by Richardson (56 countries, 1980-2008) shows unemployment and population size are strongly correlated with increased terrorist activity [58]. Empirical Resrach by Bhatia and Ghanem shows that unemployed and underemployed, educated Arab youth is more likely to be radicalised.

As consequence of the rapid population growth and high unemployment, the MENA region is experiencing one of the highest urbanisation rates in the world [40]. Extreme weather events can also play an key role; the urbanisation of Syria discussed in section 2.3.3.3 was caused by extreme weather events in combination with poor water resource mismanagement and rapid population growth. Severe urbanisation can lead to social dislocation and uncertainty, which makes migrants particularly vulnerable to terrorist recruitment [97]. Empirical research by Campos and Gassebner (130 countries, 1968-2003) shows an increase in urbanisation rates leads to higher terrorist activity [116]. This confirms Homer-Dixon's argument and thus population growth, unemployment and urbanisation are included in the statistical analysis.

A demographic factor that one might link to terrorism is education. This comes from the misconception that (Islamic) terrorists are uneducated, impoverished individuals. Empirical research by Krueger and Maleckova shows little direct connection between education and participation in terrorism [39]. Testas finds higher education levels give rise to more transnational terrorism (37 sample Muslim countries). Education will therefore not be considered.

2.3.3.5 Sociopolitical Factors

Inherent to all human beings, human rights are basic rights that include, but are not limited to, equality, liberty, security and freedom of speech [18]. Large scale violations of human rights and fundamental freedoms have been proposed as underlying causes of terrorism [4]. Where higher levels of political rights provide ways for the population to engage in the political process and allow political opposition, restricted political rights can justify operating outside the norms of the political process [124]. Denial of political rights can contribute to a sense of injustice and resentment, an atmosphere in which people are vulnerable to terrorist recruitment.

Empirical research by Abadie (186 countries, 2003-2004) finds a nonlinear relationship between political rights and terrorist activity; countries with intermediate political freedom experience higher terrorist activity than both highly authoritarian regimes and countries with high political freedom such as legitimate democracies [30]. This suggests extreme denial of political rights manages to contain terrorism. Newman does however not find a clear increase in terrorist activity with decreasing political rights [115].

Although civil liberties show high correlation with political rights, there is a fundamental difference between the two [146]. Civil liberties include a wide range of liberties such as freedom of expression, association, education and religion. Political rights revolve around the domain of political participation and other democratic practices. Civil liberties violations lead to feelings of indignity, humiliation and injustice. Research by Krueger finds that countries with low levels of civil liberties "...are more likely to be the countries of origin of the perpetrators of terrorist attacks". Newman does not find a clear increase in terrorist activity for decreasing civil liberties [115].

For various reasons, corruption has increasingly been linked to terrorism. First of all, corrupt state institutions are less effective in fighting terrorism and are vulnerable to exploitation by terrorist groups [24]. Furthermore, security forces are weakened by corruption, because the money destined for the soldiers, weapons and supplies, gets diverted. Moreover, weapons and equipment get sold to the terrorist organisations they are meant to be used against. A 2014 report by the Special Inspector General for Afghanistan Reconstruction concluded that 43% of the ca. 200000 weapons donated by the United States and other NATO partners to the Afghan National Army since 2004, had been lost [107].

Secondly, corruption finances terrorist organisations by facilitating illicit trade. More than half of the worldwide income generated by drug trade was attributed to terrorist organisations ISIL, Taliban, Al-Qaeda, Al-Shabaab and Boko Haram [14].

Finally, corruption leads to feelings of injustice that make people vulnerable to terrorist recruitment.

Chapter 3

Global Terrorism Database

3.1 Introduction

The Global Terrorism Database (GTD), version June 2015, contains information on terrorist attacks from 1970 through 2014 [15]. This chapters starts with a concise historical review of the phases of different collection methodologies of the database. The logical next step to assess the definition according to which to the attacks are included, was covered in the previous chapter.

The GTD provides a wide array of information for domestic, transnational and international terrorist incidents around the world. By means of general statistics and data visualisation, section 3.3 demonstrates what information the database offers. Simultaneously, it quantifies the recent history of regional terrorism. The analysis shows three regions are particularly interesting from both a sociological and mathematical perspective.

It will turn out that the focus on Islamic terrorist organisations proposed in section 2.3.2, has great overlap with these regions. Section 3.4 analyses the corresponding data and determines the final focus of this study.

3.2 Collection Methodology

To understand how the GTD is constructed, it is necessary to review its history. The first version of the GTD originates from a terrorism database (containing attacks from 1970 to 1997) constructed by security and risk management company Pinkerton Global Intelligence Service (PGIS). In 2001, PGIS donated paper copies of their database to researchers at the University of Maryland. The Maryland team revised the data, correcting and adding additional information, and completed digitisation in 2005. Unfortunately, PGIS lost the data for 1993 in an office move, which has never been fully recovered. This is the reason that the GTD does not include the year 1993 [74].

In 2006, a cooperation between the National Consortium for the Study of Terrorism and Responses to Terrorism (START) and the Center for Terrorism and Intelligence Studies (CETIS), sponsored by the U.S. Department of Homeland Security, started to extend the GTD beyond 1997. The retrospective assembling of incidents by means of archival sources, meant changing the original data collection methodology, that documented terrorist events real-time. The extension was finished in 2008, the same year analysts of a different institute took over the data collection that START then included in the GTD. In 2012, START took over all data collection at the University of Maryland in 2012, which was the last methodological advancement in the data collection of the GTD.

The current data collection method combines automated and manual data collection [15]. Initial filters reduce the large amount of articles to a selection of potentially relevant articles. Natural language processing and machine learning techniques refine the results to a final selection, which the GTD team then manually reviews. The validity of each source document is assessed. In order for an incident to be included in the GTD, it must satisfy definition 2.1 and needs be documented by at least one high-quality independent source. In certain areas however, terrorism is not sufficiently documented by high-quality sources, which makes the GTD vulnerable to undervaluing terrorist activity in particular geographic areas.

But more troubling are the four different data collection methods, because it makes the GTD a questionable source for measuring trends [123]. The dramatic increase of worldwide attacks in 2012 over 2011, see figure 3.1, is partly the result of improved efficiency of the data collection process [15].

Nevertheless, the GTD is widely recognised as the most comprehensive open-source terrorism database. In a later stage, its limitations turn out to impact the model results, see section 6.3.

3.3 General Statistics

The Global Terrorism Database (GTD), version June 2015, covers 141966 terrorist attacks from 1970 through 2014. First of all, this includes 896 attacks (0.6%) for which no day and/or month is specified. These attacks are useless for this research and omitted in any further analysis.

In the period between 1970 and 2014, terrorism was prevalent all over the world, see 3.2. Iraq (16003), Pakistan (11485) and India (9040) immediately stand out for their troubling numbers of attacks. But over time, numerous countries struggled with problematic situations, especially Colombia, Peru, El Salvador, the Philippines and the United Kingdom.

The corresponding number of deaths show a slightly different image, see figure 3.3. Especially the African continent accounts for a larger global share, on the contrary to Western Europe. Indeed, this is confirmed by the average number of deaths per attack in figure 3.4. The highest averages in the world are prevalent around Central Africa, with extreme cases being South Sudan and Rwanda, showing averages above 20 deaths per attack. On the other hand, most countries in Europe and South America show averages below one.
Differences between regions become more apparent when aggregating the data, see table 3.1:

- Australasia & Oceania (0.2%), Central Asia (0.4%) and East Asia (0.5%) play an insignifcant role in global terrorism ($< 1\%$) and are omitted from further analysis.

- Relatively high average number of deaths per attack ratios are encountered in Central America & Caribbean (2.8), Middle East & North Africa (2.6) and especially Sub-Saharan Africa (4.7). On average, attacks in Sub-Saharan Africa are ca. 12 times deadlier as in Western Europe.

- Terrorist attacks can be connected, even though the events do not consititute a single incident (for example 9/11, the hijacks of the airplanes at different locations in the US). The significant regions show between 10-17% of such 'multiple incidents'.

- In case when the perpetrator group is unknown, events are coded with an unknown group name. In Middle East & North Africa (63.7%) and South Asia (50.4%), the majority of the attacks do not have an identified perpetrator. A focus on Islamic terrorist organisations that operate in these regions, thus means a loss of over half the original data. In Sub-Saharan Africa the perpetrator is identified in 37.9% of the cases. This needs to be taken into consideration when deciding on the focus of this research.

The dynamics of regional terrorism over time can be observed in figure 3.1. Insignicant regions are aggregated as 'Other' and the year 1993 is linearly interpolated. There seems to be a general turning point in 1998. Whether this has to do with the shift in data collection methodology discussed in 3.2 is unclear. There is a strong resemblance between neighbouring regions Middle East & North Africa and South Asia after this year. Although terrorism in these regions often originates in domestic conflict, there is a clear common denominator which supports Rapoport's wave theorem discussed in section 2.1.1.

The global trend is better visible when presenting the regions in a bar chart, see figure 3.7. The fact that terrorism has reached unprecedented heights explains why terrorism is one of the world's biggest problems today. Moreover, the global upward trend spanning the last decade indicates a future of increased instability and conflict.

Since the upsurge of terrorism in the new millennium, regions Middle East & North Africa, South Asia and Sub-Saharan Africa account for the vast majority of terrorist attacks and show similar trends. From both a sociological and mathematical perspective, modelling attacks from these regions in this timeframe therefore seems a legitimite choice.

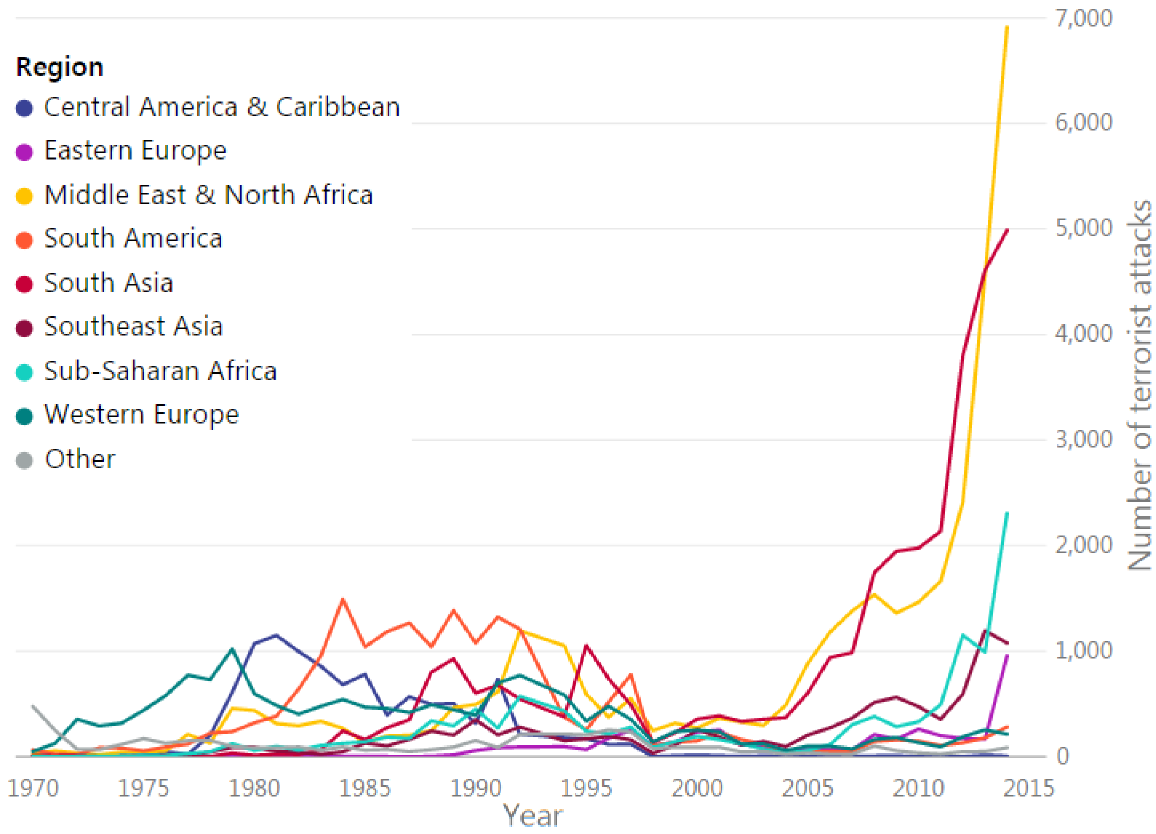

Fig. 3.1 Number of attacks over time per region

Recall that the second requirement of definition 2.1 contains three criteria, for which section argued the third is indispensable in order to prevent including incidents that are legal according to international humanitarian law. Futhermore, attacks that do not satisfy criterion 1 or 2 can be of a fundamentally different nature. This study therefore chooses to work with the 122382 attacks (86%) that satisfy all three criteria.

	Population[1] (millions)	Attacks	Global Share	Deaths	Injured	Average deaths	Multiple incident	Group unknown
Australasia & Oceania	40	234	0,2%	144	226	0.6	6.8%	59.0%
Central America & Caribbean	82	10252	7.3%	28597	8963	2.8	12.1%	35.4%
Central Asia	68	528	0.4%	957	1965	1.8	7.4%	79.2%
East Asia	1601	746	0.5%	997	9074	1.3	16.6%	58.0%
Eastern Europe	190	4196	3.0%	6405	10316	1.5	11.4%	70.1%
Middle East & North Africa	357	34305	24.3%	88156	157136	2.6	11.8%	63.7%
North America	565	3154	2.2%	4602	3733	1.5	15.0%	24.6%
South America	385	18313	13.0%	28283	16286	1.5	15.2%	27.8%
South Asia	1749	33165	23.5%	77087	112048	2.3	10.5%	50.4%
Southeast Asia	618	9265	6.6%	13469	22533	1.5	15.5%	49.7%
Sub-Saharan Africa	973	11393	8.1%	53893	36902	4.7	14.9%	37.9%
Western Europe	411	15519	11.0%	6208	16067	0.4	12.0%	28.6%
Global	6821	141070	100%	308798	395249	1.3	12.5%	46.3%

Table 3.1 General statistics regional terrorism

[1] World Population Sheet 2014 [13]

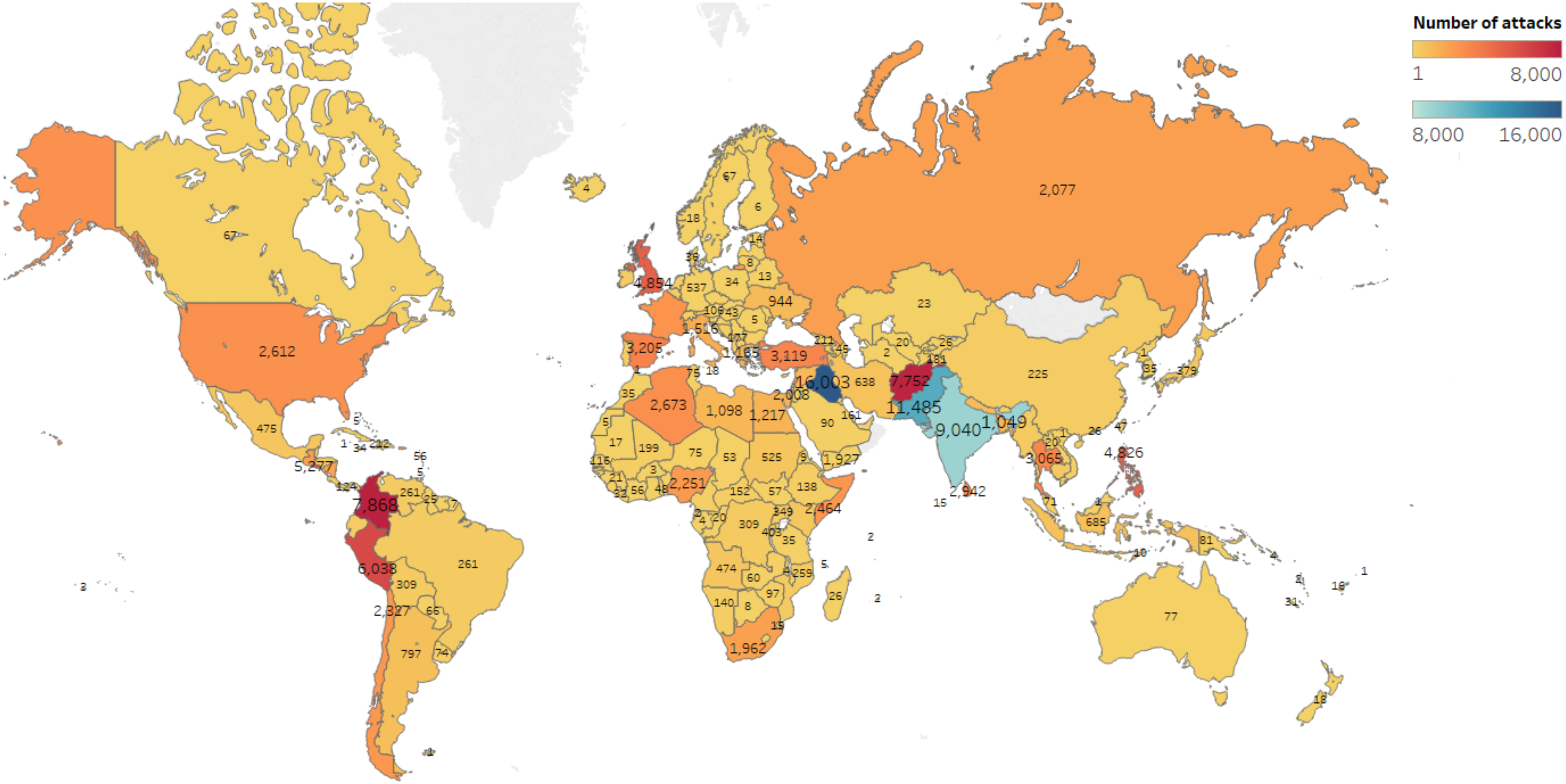

Fig. 3.2 Number of terrorist attacks per country (1970-2014)

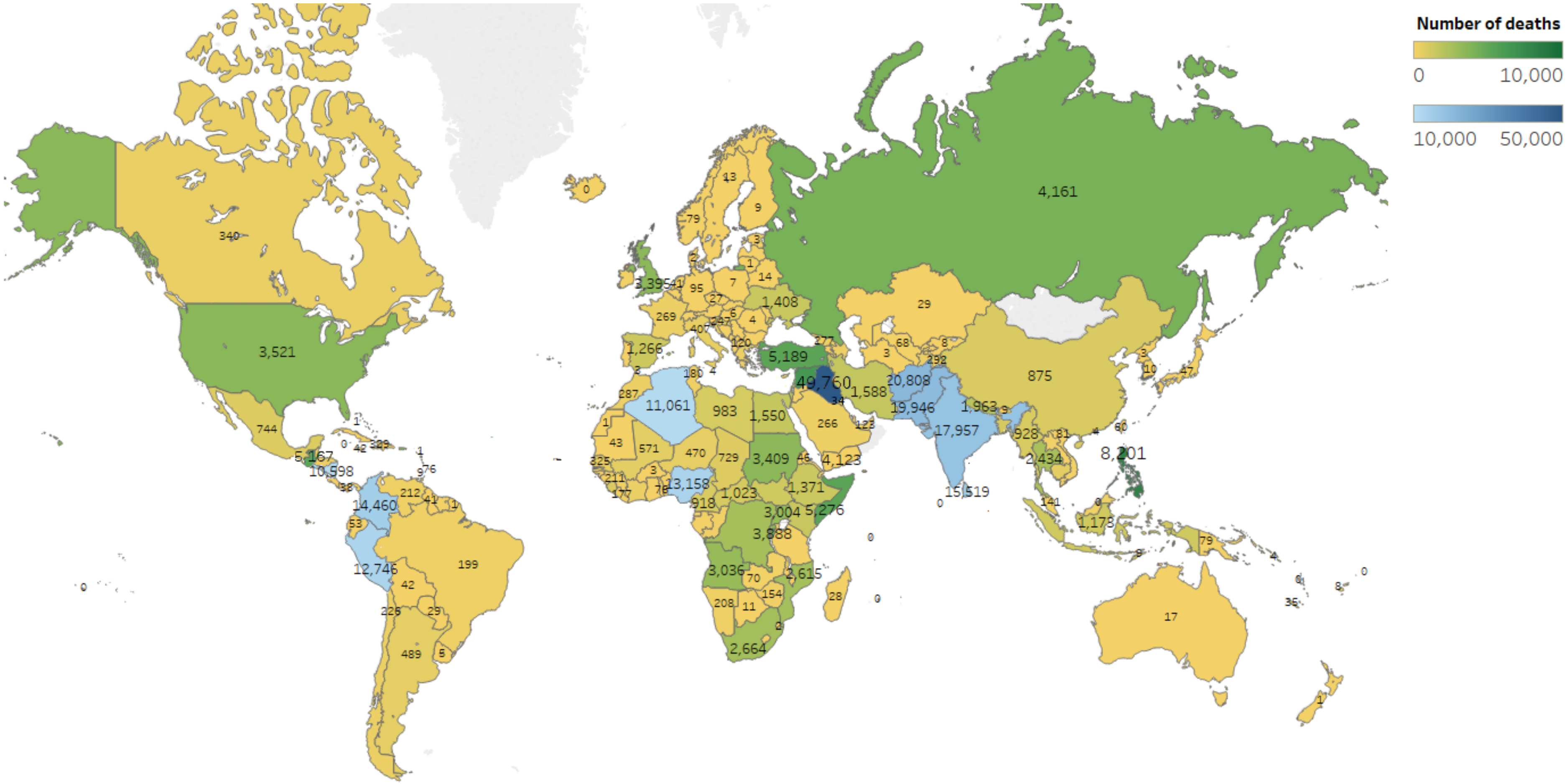

Fig. 3.3 Number of deaths per country (1970-2014)

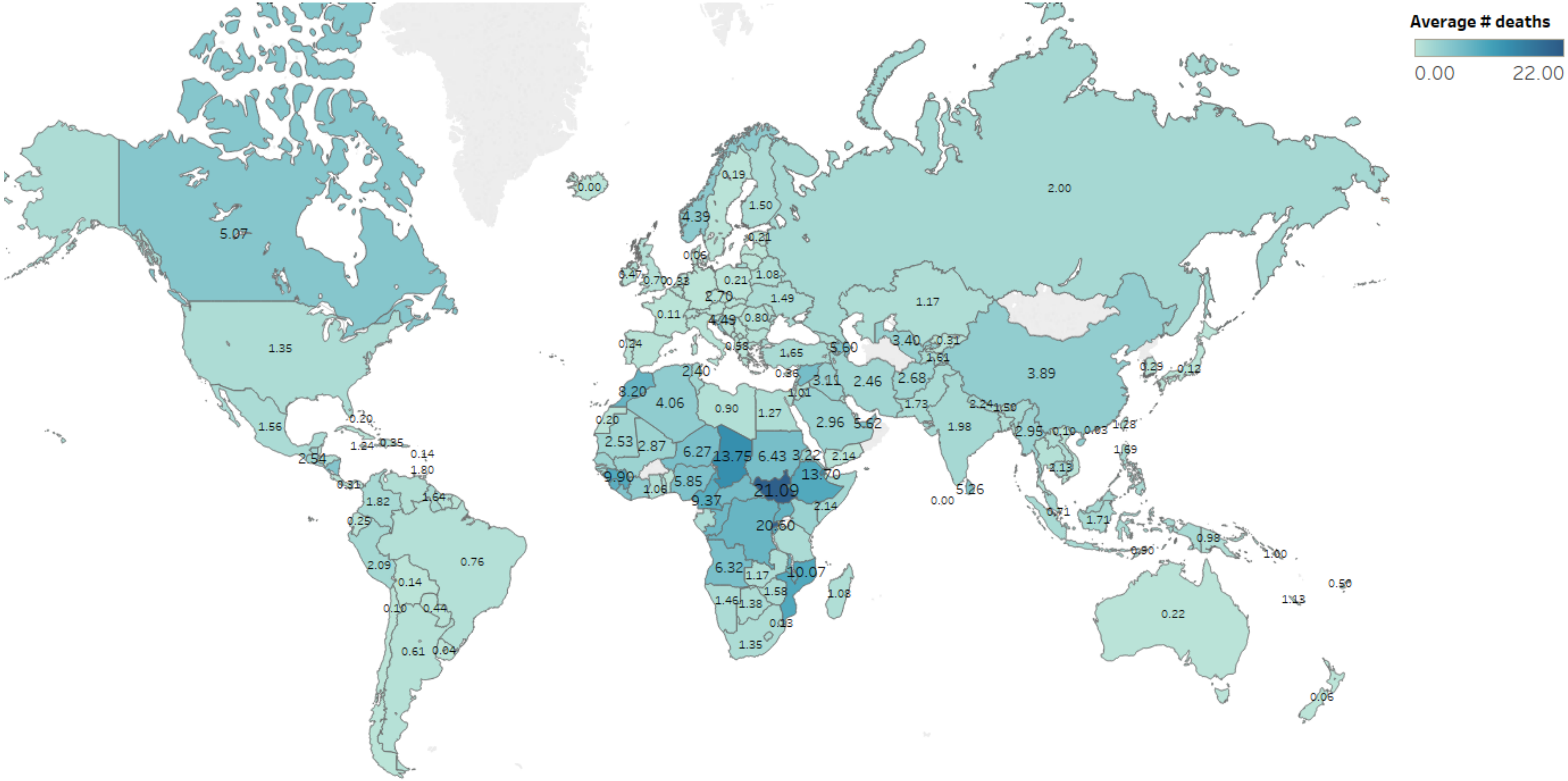

Fig. 3.4 Average number of deaths per attack by country (1970-2014)

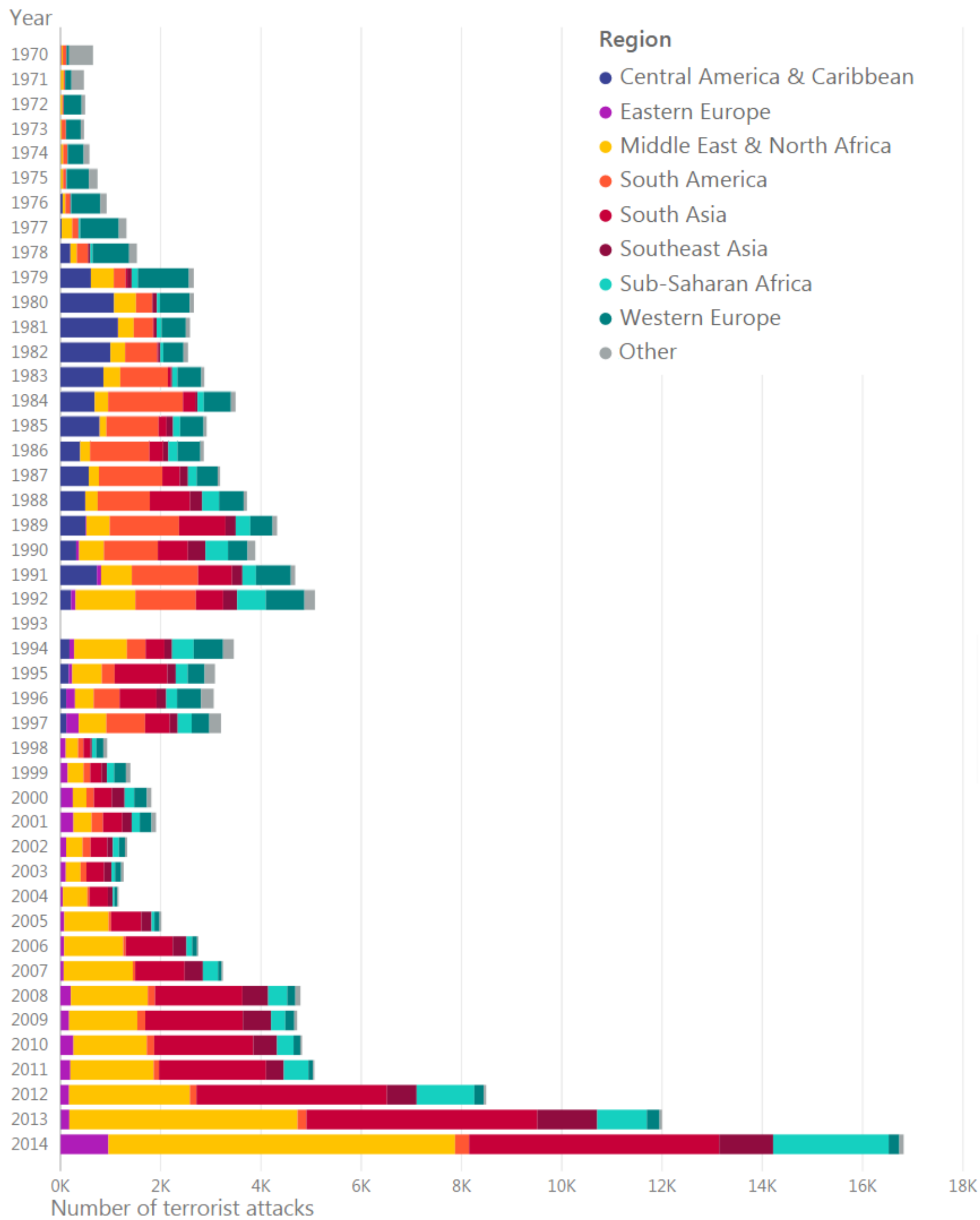

Fig. 3.5 Number of attacks per region

3.4 Focus of this study

General data analysis in the previous section suggested that modelling attacks from the regions Middle East & North Africa, South Asia and Sub-Saharan Africa between the period 1998-2014, is interesting from both a sociological and mathematical perspective. This shows great overlap with the proposed focus on Islamic terrorist organisations in section 2.3.2, which was based on empirical research by Newman that examined the analytical potential of root causes for different methodological approaches.

There is however a vast number of perpetrator groups in each region, see 3.6. Since the GTD does not provide a way to select the Islamic organisations, hundreds of groups need to be evaluated.

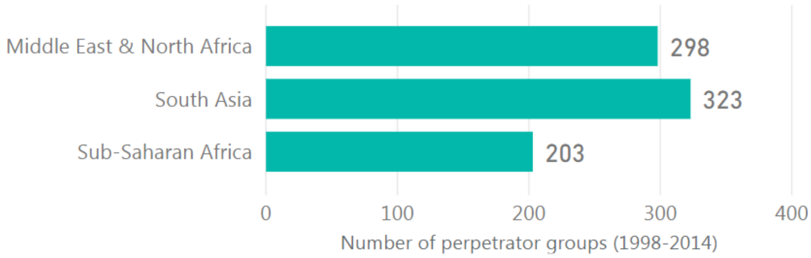

Fig. 3.6 Number of terrorist organisations by region for attacks between 1998-2014

As alternative, the major organisations are examined. Figure 3.7 shows the share of the seven largest groups of each region. The Taliban dominates South Asia claiming more than a third of the attacks. The second major Islamic organisation is Tehrik-i-Taliban in Pakistan. Still, non-Islamic groups such as communist parties and militant organisations are very prominent in South Asia.

In Sub-Saharan Africa, Islamic organisations Al Shabaab in Somalia and Boko Haram in Nigeria make up for ca. 60% of all attacks in Sub-Saharan Africa. Attacks in the Middle East & North Africa are more evenly dispersed, which reflects the wide spread conflict around the region. Except for the Kurdistan Workers Party (PKK), the main groups in the region are Islamic.

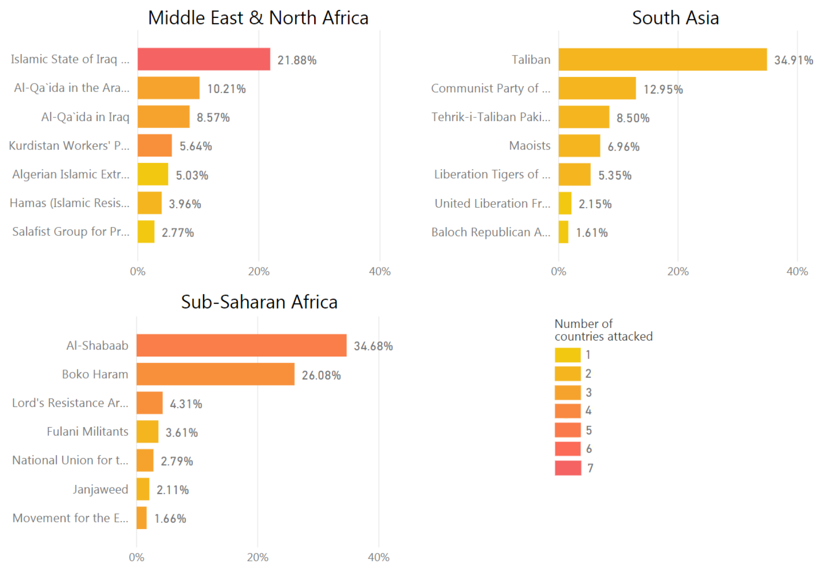

Fig. 3.7 Region shares of seven largest terrorist organisations (1998-2014). Colors represent the amount of different countries, within the region, each group attacked.

By means of a color scale, figure 3.7 in addition shows the amount of countries within the region which each group attacked. Note that this gives merely an indication of the international orientation of the groups, since a single attack outside the country of origin already impacts the color. Nevertheless, the figure suggests that the main groups in Middle East & North Africa and Sub-Saharan generally seem to operate more international than in South Asia. This could be a problem from a methodological perspective. The root causes covariates consist of country data and might not reflect the international character of organisations. The major Islamic organisations do however mainly operate in their country of origin: Islamic State of Iraq and the Levant (89.41% in Iraq), Al-Qa'ida in the Arabian Peninsula (98.63% in Yemen), Al-Qa'ida in Iraq (99.20% in Iraq), Taliban (98.84% in Afghanistan), Tehrik-i-Taliban (99.03%), Boko Haram (95.19% in Nigeria) and Al-Shabaab (84.37% in Somalia).

Apart from the Taliban and Al-Qa'ida in Iraq, the activity of these groups emerges after 2007, see figure 3.8. The absence of Islamic State of Iraq and the Levant (ISIL) before 2013 is explained by the fact that the group was first active as Islamic State of Iraq (ISI), for which attacks are first recorded in 2007 with a maximum of 57 attacks in 2009.

A timeframe of 2007-2014 does not allow out-of-sample validation, meaning the prediction accuracy of our model cannot be assessed. For our modelling purposes, exclusive focus on the major Islamic terrorist organisation is thus not suitable.

Country	Attacks	Unknown Perpetrator
Afghanistan	7651	42.32%
Iraq	15882	83.63%
Nigeria	2190	28.31%
Pakistan	9784	77.37%
Somalia	2318	30.54%
Yemen	1845	38.92%

Table 3.2 Number of attacks and percentage unidentified perpetrator for countries of interest.

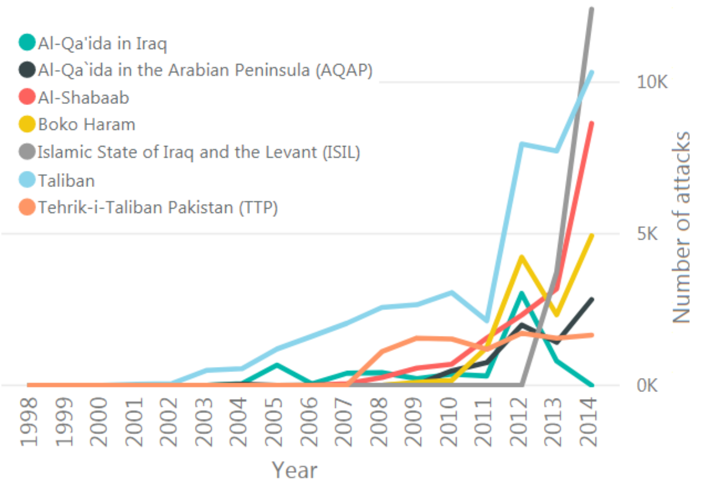

Fig. 3.8 Number of attacks by major Islamic terrorist organisations

The analysis is therefore shifted to the countries in which the Islamic organisations originate. To filter out attacks from a fundamentally different nature, such as those from a lone-actor, all attacks registered without an identified perpetrator are removed, see table 3.2. At the same time, this homogenises the data, since the countries facing the highest terrorist activity exhibit the largest perpetrator uncertainty.

A second data reduction is based on the loophole in GTD's definition of terrorism discussed in section 3.3. In order to satifsy the corrected definition 2.2, al three criteria of definition 2.1 need to be satisfiied. Attacks that do not satisfy either criteria 1 (56 attacks, 0.40%) or 2 (92 attacks, 0.65%) can be fundamentally different in nature. If criteria 3 is not met, an incident is legal according to international humanitarian law which conflicts with the definitions from the world's leading organisations. Remarkably, this is the case for 2124 out of 14080 attacks (15.09%). Finally, the 2272 attacks (16.14%) that do no satisfy all three criteria are removed, resulting in 11808 attacks constituting to the timeseries in figure 3.9. Considering the low activity in more distant years, the timeframe in further analysis is narrowed to 2000-2014.

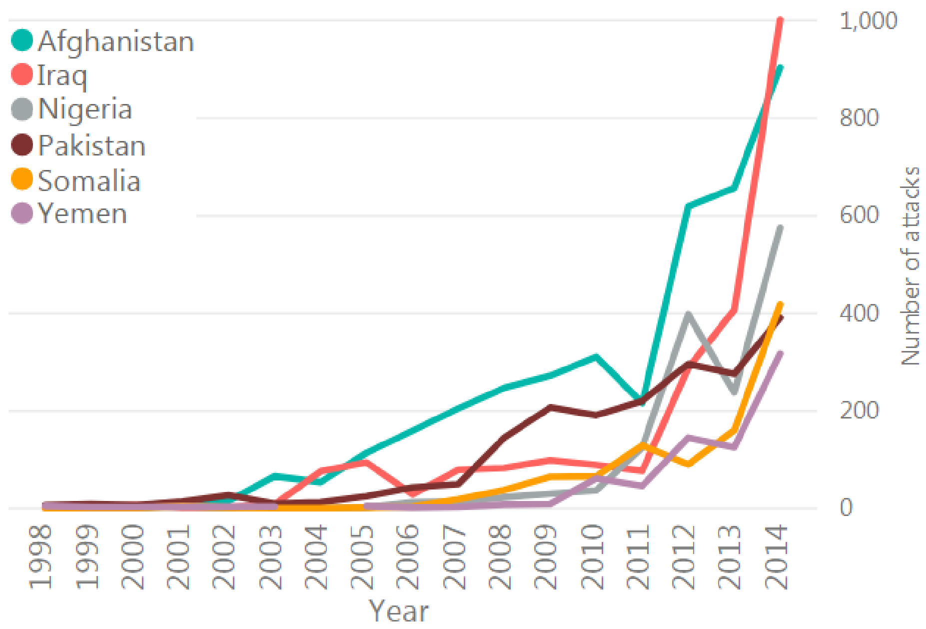

Fig. 3.9 Number of terrorist attacks by countries of interest for the final data selection

The GTD contains various additional information that could be incorporated in the model by separating the country data into different classes. From a sociological perspective, it would be particularly interesting to model terrorist attacks on different targets. Over the period 2000-2014, the six countries of interest were hit at 21 distinct categories, varying from civilians to government officials, police to military, business to private property and religious to educational institutions. Although the GTD allows multiple targets per attack, 94.6% contains one target. For attacks with multiple targets, the primary target is considered leading. This results in the distribution of table 3.3. The 21 targets are aggregated into five distinct, more general classes, found in the outer right column.

The distribution of the aggregated targets is shown in figure 3.10. Aggregated targets by country can be incorporated in the model by introducing dummy variables.

Target	Attacks	Aggregated Target
Airports & Aircraft	137	State, Government & Public Goods
Business	2669	Public Actors
Educational Institution	1418	State, Government & Public Goods
Food or Water Supply	63	State, Government & Public Goods
Government (Diplomatic)	374	State, Government & Public Goods
Government (General)	4652	State, Government & Public Goods
Journalists & Media	417	Public Actors
Maritime	54	State, Government & Public Goods
Military	1604	Military & Police
NGO	275	Public Actors
Other	99	Unknown & Other
Police	6892	Military & Police
Private Citizens & Property	10389	Private Citizens & Property
Religious Figures/Institutions	1276	Public Actors
Telecommunication	158	State, Government & Public Goods
Terrorists/Non-State Militia	988	Unknown & Other
Tourists	33	Unknown & Other
Transportation	923	State, Government & Public Goods
Unknown	1129	Unknown & Other
Utilities	906	State, Government & Public Goods
Violent Political Party	326	Public Actors

Table 3.3 Targets hit in six countries of interest, 2000-2014

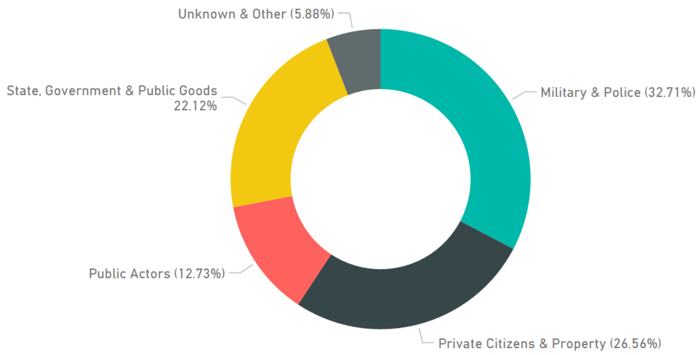

Fig. 3.10 Distribution of attacks over aggregated targets for the six countries of interest, 2000-2014

3.4.1 Summary

Concluding, general regional statistics turned the attention towards three regions -Middle East & North Africa, South Asia and Sub-Saharan Africa- for the period 1998-2014. The data showed great overlap with the proposed focus on Islamic terrorist organisations in section 2.3.2. However, a specific focus on Islamic terrorist organisation turned out to be infeasible for our modelling purposes. Instead, six countries were selected: Afghanistan, Iraq, Nigeria, Pakistan, Somalia and Yemen. A large amount of attacks without identified perpetrator were omitted, as a first filter of attacks from a fundamentally different nature. Similarly, attacks that did not satisfy the three criteria of definition 3.3 were removed, mainly to prevent the inclusion of incidents legal according to international humanitarian law. Finally, the timeframe was narrowed to 2000-2014. The 21 different targets hit in this period were aggregated to allow the modelling of targets by country.

Chapter 4

The Model

4.1 Introduction

The analysis of the GTD suggested to model the number of terrorist attacks from known perpetrators, satisfying all three criteria of definition 2.1, on aggregated targets of six countries over the period 2000-2014. Reviewing literature from social science brought forward candidates to possibly include in the model as covariates. The next step is to choose an appropriate model. Section 4.3 contains an in-depth analysis of the country data and provides the justification of using a nonhomogeneous Poisson process model with a Weibull baseline intensity function. The model is discussed into detail, including an adjustment that allows for time-dependent covariates, and derives the log-likelihood function needed for the estimation of parameters. This might however be meaningless to the reader unacquainted with mathematics. Therefore, this chapter starts with the basic theoretical background required to understand both the model and its justification. For basic concepts of probability models and an introduction to stochastic processes is referred to Ross, chapter two and five of [136] and chapter two of [135], respectively.

4.2 Preliminaries

4.2.1 Poisson process

The Poisson process is a widely used stochastic process, particularly for modelling situations where one counts the occurrences of completely random events. Its name derives from the fact that for a Poisson process with intensity λ, the number of events occurring in any interval of length t, is Poisson distributed with mean λt.

A stochastic process $\{X(t) : t \in T\}$ is a collection of random variables (defined on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is mathematically more precise). Since this study is interested in the number of terrorist over time, for each $t \in T$, $X(t)$ is simply a random variable observed at time t.

There are several, equivalent ways to define the Poisson process. A closer look at definition 4.1 reveals that observing the number of terrorist attacks over time yields a counting process. It is therefore convenient to define the Poisson process as special case of counting process.

Definition 4.1. A counting process $\{N(t), t \geq 0\}$, counting the number of occurrences of some type of event over time, satisfies the following conditions

(i) $N(t) \geqslant 0$

(ii) $N(t)$ is integer valued

(iii) If $s \leq t$ then $N(s) \leq N(t)$, i.e. $N(t)$ is non-decreasing

(iv) For $s < t$, $N(t) - N(s)$ represents the number of events occurred in interval $(s,t]$ (i.e. $N(t)$ is right-continuous)

Where $N(t)$ represents the number of events that have occurred up to and including time t, starting at zero. Equivalent notation often found in literature is N_t and $N(0,t)$. Thus $N(t) - N(s)$, the number of events that occurred in $(s,t]$, can also be written as $N(s,t)$

Before defining the Poisson process, there is one property of stochastic processes that needs to be introduced

Definition 4.2 (Independent increments)**.** Let $\{X(t), t \geq 0\}$ be a stochastic process. Then $X(t)$ is said to have independent increments if for $t_1 < t_2 < ... < t_n$, the increments $X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ are independent.

For counting process N_t having independent increments, means that the number of events that occurred in disjoint intervals are independent.

When the observed events of a Poisson process occur with a constant rate λ, the process is called a homogeneous Poisson process (HPP). However, as section 4.3.3.2 shows, the clear upward trend of figure 3.9 cannot be modelled with a HPP. Therefore a generalisation is needed that allows the intensity to vary over time; the nonhomogeneous Poisson process (NHPP). Nevertheless, the HPP is defined first, since, despite its simplicity, it provides intuition and shares many properties with the NHPP, despite its simplicity.

4.2.1.1 Homogeneous Poisson process

Definition 4.3 (Homogeneous Poisson process)**.** A counting process $\{N(t), t \geq 0\}$ is said to be a homogeneous Poisson process with constant intensity λ, if the following conditions are satisfied:

i) $N(0) = 0$

ii) $N(t)$ has independent increments

iii) The number of events in any time interval of length t is Poisson distributed with mean λt, that is:
$P\{N(t+s) - N(s) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ for all $s, t \geq 0$, $k = 0, 1, ...$

Note iii) means that the probability distribution of the number of events of an interval only depends on the length of that interval. Thus the HPP has stationary increments, as stated in definition 4.4:

Definition 4.4 (Stationary increments)**.** Let $\{X(t), t \geq 0\}$ be a stochastic process. Then $X(t)$ is said to have stationary increments if the probability distribution for $X_{t+s} - X_t$ is the same for all $s \in T$ such that $s + t \in T$

In addition, follows from iii) that the equidispersion property holds, i.e. for HPP $N(t)$ with rate λ, both the mean and the variance equal λt, see theorem 4.5.

Theorem 4.5 (Expected value and variance of a homogeneous Poisson process)**.** *The mean and variance of a HPP $N(t)$ with rate $\lambda > 0$, are given by*

$$\mathbb{E}\big[N(t)\big] = Var\big[N(t)\big] = \lambda t$$

A powerful property is that the sum of two independent HPPs is again a HPP, as illustrated in figure 4.1 and stated in theorem 4.6.

Theorem 4.6 (Superposition of two independent homogeneous Poisson processes)**.** *Let N_1 and N_2 be two HPPs, with intensities λ_1 and λ_2, respectively. Then $N_3 := N_2 + N_1$ is a HPP with intensity $\lambda_1 + \lambda_2$*

Another important group of properties of the Poisson process concerns the intervals between events. Theorem 4.7 shows that the interarrival times, the times between succeeding events, of the HPP are themselves independent and identically distributed exponential random variables. Figure 4.2 illustrates the concept of interarrival times. It is possible to show that if the interarrival times are independent and identically exponentially distributed, then counting process $\{N(t), T \geq 0\}$ is a Poisson process [96].

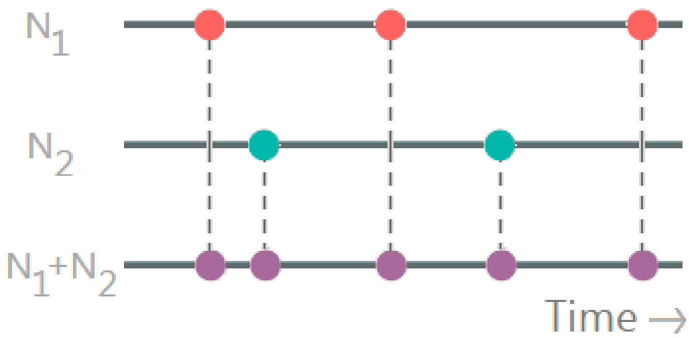

Fig. 4.1 Superposition of two HPPs is again a HPP

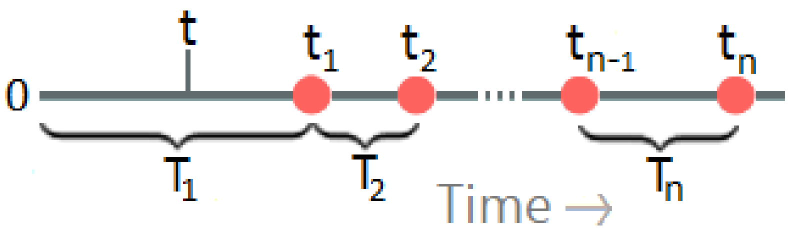

Fig. 4.2 Arrival times of a Poisson process. As theorem 4.7 shows, $T_1, T_2, ..., T_n \sim exp(\lambda)$ and are independent

Theorem 4.7 (Distribution interarrival times HPP)**.** *Let $N(t)$ be a HPP with rate λ and $T_1, T_2, ..., T_n$ be the inter-arrival times. Then $T_1, T_2, ..., T_n$ are independent and indentically distributed (i.i.d.) exponential random variables with rate λ.*

Proof. A nice proof using induction is found in [75].

□

From theorem 4.7 follows the joint probability density that events in the interval $(0, T]$ occur at times $t_1 \geq t_2 \geq ... \geq t_n$ is then given by

$$\lambda e^{-\lambda t_1} \lambda e^{-\lambda t_2} \dots \lambda e^{-\lambda t_n} e^{-\lambda(T-t_n)} = \lambda^n e^{-\lambda T} \tag{4.1}$$

where all terms on the left of the equality are p.d.f.'s of intervals, except the last term, which represents the probability that there is no event in $(t_n, T]$. The analogous result for a NHPP, see equation 4.7, determines the likelihood function of our final model as found in theorem 4.13.

To test homogeneity of the Poisson processes in figure 3.9, section 4.3.3.2 makes use of the fact that vector of increments of a HPP, given the total number of events, is multinomial distributed, more specifically $Mult(n,\pi)$. The relationship between the multinomial distribution and Poisson random variables is shown in theorem 4.8.

Theorem 4.8 (Conditional distribution of Poisson random variables)**.** *Let $X_1, X_2, ..., X_n$ be independent Poisson random variables with $X_i \sim Poisson(\lambda_i)$ for $i = 1, 2, ..., n$. Then the distribution of the vector $X = (X_1, X_2, ..., X_n)$ given their total number events $N = \sum_{i=1}^{n} X_i$, is Multinomial(n,π), i.e. $X \mid N \sim M_n(N, \pi)$, with $\pi = (\pi_1, \pi_2, ..., \pi_n)$ and $\pi_i = \frac{\lambda_i}{\lambda_1+\lambda_2+...+\lambda_n}$.*

Proof. Since the X_i are independent, the joint probability mass function for $x \in \mathbb{N}$ can be written as

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!}$$

Since the sum of Poisson random variables is again Poisson, $N = \sum_{i=1}^{n} X_i$ is also a Poisson random variable with intensity $\lambda_N = \sum_{i=1}^{n} \lambda_i$. Therefore

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid N = k) = \begin{cases} e^{-\lambda_N} \frac{\lambda_1{}^{x_1} \lambda_2{}^{x_2} ... \lambda_n{}^{x_n}}{x_1! x_2! ... x_n!} & \text{if } \sum_{i=1}^{n} x_i = k \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Hence by definition of the joint conditional probability mass function

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x} \mid N = k) &= \frac{P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid N = k)}{P(N = k)} \\ &= \frac{e^{-\lambda_N} \frac{\lambda_1{}^{x_1} \lambda_2{}^{x_2} ... \lambda_n{}^{x_n}}{x_1! x_2! ... x_n!}}{e^{-\lambda_N} \frac{\lambda_N{}^{k}}{k!}} if \sum_{i=1}^{n} x_i = k \\ &= \frac{k!}{x_1! x_2! ... x_n!} \left(\frac{\lambda_1}{\lambda_N}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_N}\right)^{x_2} ... \left(\frac{\lambda_n}{\lambda_N}\right)^{x_n} \end{aligned}$$

□

Note that the theorem is proven for independent Poisson random variables, not necessarily having equal intensities. For a HPP, $\lambda_1 = \lambda_2 = ... = \lambda_n$ and therefore $\pi_1 = \pi_2 = ... = \pi_n$, i.e the distribution of attacks in the intervals is the same, meaning that the expected number of events in each interval is the same. It is this, that is used to show that timeseries of countries as Iraq are nonhomogeneous Poisson. Indeed, considering figure 3.9, it is just not possible for these timeseries to have stationary increments.

Despite the simplicity of the HPP, it is widely used in queueing- and renewal theory. The HPP is however not able to model the clear trends in our data. A generalisation of the homogeneous Poisson process in needed that allows the intensity to vary over time. This generalisation is known as the nonhomogeneous Poisson process (NHPP).

4.2.1.2 Nonhomomogeneous Poisson process

The NHPP, when considered on the positive half-line, can also be defined as a counting process, see definition 4.9.

Definition 4.9 (Nonhomogeneous Poisson process)**.** A counting process $\{N(t), t \geq 0\}$ is said to be a NHPP with intensity $\lambda(t)$, if the following conditions are satisfied:

i) $N(0) = 0$

ii) $N(t)$ has independent increments

iii) $N(t) - N(s) \sim Poisson\big(\int_s^t \lambda(u)du\big)$

The integral from iii) is often written as intensity measure $\Lambda(s,t) = \int_s^t \lambda(u)du$. The probability of k events in interval $(s, t+s]$ is then

$$P\{N(t+s) - N(s) = k\} = e^{-\Lambda(s,t+s)} \frac{\Lambda(s,t+s)^k}{k!} \quad for\ all\ s,t \geq 0, k = 0,1,... \tag{4.3}$$

Since the probability distribution depends not only on the length of the interval, but also on the location of the interval (since $\lambda(t)$ is varies over time), the NHPP does not have stationary increments. Therefore, there is the possibility that events are more likely to occur at certain times than at others [135].

Note that $\lambda(t)$ is a deterministic function of time. When $\lambda(t) = \lambda$, i.e. constant intensity function, definition 4.9 equals the definition of the HPP. The HPP can therefore also be seen as a special case of the NHPP.

The equidispersion property from the Poisson random variable also holds for the NHPP, as stated in theorem 4.10. Where the mean (and variance) of a HPP is simply its intensity λ times the length of the interval, the mean (and variance) of the NHPP is obtained by integrating $\lambda(t)$ over the length of the interval.

Theorem 4.10 (Expected value and variance of a NHPP)**.** *The expected value and variance of a NHPP $\{N(t), t \geq 0\}$ with rate $\lambda(t)$, are given by*

$$\mathbb{E}\big[N(t)\big] = Var\big[N(t)\big] = \Lambda(0,t) = \int_0^t \lambda(u)du$$

The superposition property for the homogeneous Poisson process, see theorem 4.6 and figure 4.1, also holds for the NHPP, see theorem 4.11

Theorem 4.11 (Superposition of two independent inhomogeneous Poisson processes)**.** *Let N_1 and N_2 be two NHPPs, with intensities $\lambda_1(t)$ and $\lambda_2(t)$, respectively. Then $N_3 := N_2 + N_1$ is a NHPP with intensity $\lambda_1(t) + \lambda_2(t)$.*

The following theorem is the nonhomogeneous analog of theorem 4.7 and describes the distribution of the interarrival times of a NHPP.

Theorem 4.12 (Distribution interarrival times NHPP)**.** *Let $N(t)$ be a NHPP with rate λ_t and $T_1, T_2, ..., T_n$ be the inter-arrival times. Then for any n, the density of interarrival time T_n is*

$$f_{T_n}(t \mid t_{n-1}, t_{n-2}, ..., t_1) = \lambda(t + t_n) e^{-\int_{t_{n-1}}^{t+t_{n-1}} \lambda(u) du} \tag{4.4}$$

Proof. A good proof is found in [68]. □

The interarrival times of a NHPP are no longer identically distributed, but still have conditionally exponential distributions [68]. Theorem 4.12 gives the probability density of the interval to the next event and thus the probability that, starting from time t_i, the next event is in $(t_{i+1}, t_{i+1} + \Delta t]$, can be estimated by

$$\lambda(t_{i+1}) e^{-\int_{t_i}^{ti+1} \lambda(u) du} \Delta t + o(h) \tag{4.5}$$

As Cox writes [66], this is a generalisation of the simple exponential distribution of the intervals of the HPP, see theorem 4.7. Given that events occur at times $t_1 < t_2 < ... < t_n$, observing a NHPP from $[0, T]$, results in the likelihood function

$$\lambda(t_1) \exp\left\{-\int_0^{t_1} \lambda(u) du\right\} \lambda(t_2) \exp\left\{-\int_{t_1}^{t_2} \lambda(u) du\right\} \cdots$$
$$\lambda(t_n) \exp\left\{-\int_{t_{n-1}}^{t_n} \lambda(u) du\right\} \exp\left\{-\int_{t_n}^{T} \lambda(u) du\right\} \tag{4.6}$$
$$= \left\{\prod_{i=1}^{n} \lambda(t_i)\right\} \exp\left\{-\int_0^T \lambda(u) du\right\} \tag{4.7}$$

For this study, equation 4.7 is a crucial result since it determines the likelihood function of our model, see theorem 4.13. The likelihood function can be used to estimate the model's parameters. Since there are numerous country statistics at our disposal, the next step is to allow the intensity of the NHPP to dependent on a set of covariates. One way to model the relationship between a response variable and a set of predictor variables, is with regression analysis. Standard linear regression is however not the appropriate model for our data. Poisson regression, often used to model count data, is the logical next step. The model and its assumptions are presented, along with its strengths and limitations. With this in mind, Poisson regression can be generalised to allow a multiplicative intensity function that is both dependent on covariates and on a baseline function that handles the trend.

4.2.2 Poisson Regression

Let Y_i be the ith observation of the dependent variable ($i = \{1,\ldots,n\}$), x_{ij} the ith observation of the non-random jth covariate X_j, and β_j the unknown coefficient of the jth predictor with β_0 being the intercept. For a linear regression model, the normally distributed error terms imply that $Y_i \mid X_i$ is normally distributed. This is clearly violated in case that the dependent variable is Poisson distributed, since the Poisson distribution is truncated at zero and exhibits a positive skew that decreases as λ increases. As Hutchinson and Holtman show, an OLS model does not perform well under these conditions [111] . In addition, using OLS risks violating the homoscedasticity assumption that the variance of the error terms is constant and does not depend on the predictor variables [52]. As consequence of such violations, the standard errors of the parameters will be estimated incorrectly, meaning it is not clear whether the effects of the covariates are statistically significant or not.

O'Hara and Kotze demonstrate that transforming the data is not an valid option to prevent violations in the assumptions [126]. A log-transform of count data performs poorly and can lead to impossible predictions such as negative counts of individuals. Therefore the normality of the errors and homoscedasticity assumption need to be generalised. This results in a generalised linear model (GLM), in which the dependent variable can to take on any distribution of the exponential family (which includes the Poisson distribution) and several options for heterescedascity are included.

GLMs are a broad class of models that thus allow to work with more complicated kinds of data than the linear models (LM). Assumptions can be chosen that better match the actual distributions in the data, rather than forcing the data to match certain assumptions, as is the case with the LM. To understand the differences between the two models, it helps to look at their components:

(i) The **random component** refers to the probability distribution of the response variables Y_i. For the LM, the Y_is are normally distributed, where the mean of the Y_i are allowed to differ, but their variance σ^2 is fixed (homoscedasticity). The GLM allows the Y_i to take any distribution from the exponential family and includes several options for heteroscedasticity.

(ii) The **systematic component** specifies the predictor variables for both LMs and GLMs as a combination of linear predictors, e.g.

$$\eta_i := \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{p1} + \varepsilon_i \tag{4.8}$$

(iii) **Link function** $g(\mu)$ specifies the relation between the random and systematic components of the model. It transforms the response variable so that it can be modeled as a linear function of predictors [120]. For linear regression, this link is the identify function, $\eta = g(\mathbb{E}\left[Y_i\right]) = \mathbb{E}\left[Y_i\right]$. However, when the response variable is not assumed to be normally distributed, non-linear relationships between the response and predictor variables can arise. A transformation of the expected value of the response $\mathbb{E}\left[Y_i\right]$ is then needed, i.e. $g(\mathbb{E}\left[Y_i\right]) = \eta$, where g is monotonic and differentiable [31].

A special case arises when response variable Y_i is a count variable that is assumed to follow a Poisson distribution with mean λ_i. When expected value $\mathbb{E}\left[Y_i\right]$ is transformed using the log link function, $g(x) = log(x)$, the relationship between the transformed mean of the response and the predictors is linear, as in equation 4.9. This model is called the Poisson regression model. Note that the assumption of uncorrelated errors and weak exogeneity of the covariates from the linear model still hold for the Poisson regression model, but the error terms are no longer assumed to be normally distributed.

$$log(E(Y_i)) = \eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \tag{4.9}$$

Equation 4.9 is often written as

$$\mathbb{E}\left[Y_i\right] = \lambda_i = e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}} \tag{4.10}$$

Where $\exp(\beta_0)$ is the intercept of the model. The interpretation of the coefficients differs from the linear regression model. In the linear model, a unit change in x_{ij} simply leads to a change of β_j. However, a unit change in predictor $x_i j$ of the Poisson regression model, leads to a change in mean of $\mathbb{E}\left[Y_i\right]\beta_j$, since

$$\frac{\partial \mathbb{E}\left[Y_i\right]}{\partial x_i j} = e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}} \beta_j = \mathbb{E}\left[Y_i\right]\beta_j \tag{4.11}$$

So an additive change in predictor variable has a multiplicative effect on the mean. For count data, the effects of predictors are often multiplicative instead of additive and thus the log link is advantageous. Another advantage of the log link is that the exponential relationship between the transformed mean of the response and the predictors assures that a count will not be negative.

Since its reponse variables follow a Poisson distribution, the Poisson regression model inherits the highly restrictive equidispersion assumption that requires the mean to equal the variance. When the observed variance of the data is larger than was assumed, the data is said to be overdispersed. As a result, the calculated standard errors are artificially smaller than the true standard errors [111]. Empirically, count data often exhitibs overdispersion. When this is deemed problematic, a standardised correction for overdispersion can be applied, or one can switch to a Poisson mixture model such as the negative binomial model.

Now that Poisson regression is introduced, the model can be generalised to allow a multiplicative intensity function that depends on a baseline trend function and time-dependent covariates.

4.3 Nonhomogeneous Poisson Process model with Weibull baseline intensity

4.3.1 Introduction to model assumptions and theory

This study is interested in modelling a phenomenon in which a country experiences terrorist attacks over time, influenced by underlying factors that are presumed to affect the rate at which these attacks occur. In such a setting, it might be appropriate to consider the attacks as the realisation of a Poisson process.

However, there are phenomena that can conflict with the Poisson assumptions. First of all, the large amount of attacks the countries experienced in recent years could be problematic, because the Poisson distribution models the occurrence of a rare event. Secondly, considering that the attacks a country faces are planned by terrorist organisations (the analysis of section 3.4 shows mostly by one leading organisation), it is likely that after one attack has occurred, there is an increased probability of a subsequent attack occurring within a close time period. This would mean that the data is clustered, making the underlying process more variable than the equidispersed Poisson process [131].

The attack data of the countries therefore needs to be examined, possibly applying applying techniques from extreme value theory, that can help against against clustering and violation of the Poisson hypothesis. A nonhomogeneous Poisson process can then be used, such that, conditionally on the removal of the trend and clustering, attacks are independent. The intensity function, besides being a function of the time since the process began, is adjusted to incorporate the country statistics that represent underlying causes of terrorism as reviewed in section 2.3.

Lawless proposes a Proportional Intensity (PI) model for the setting in which individuals experience repeated events and data on each individual consist of an observed point process, along with covariates [91]. The PI model derives from the Proportional Hazard (PH) model, widely used in survival analysis to assess the importance of covariates in the survival times of individuals through the hazard function (the instantaneous rate of occurrence of the event)

[83]. The PI model extends PH to the case of repeated events, where the interarrival times between attacks contribute to the likelihood [106].

In hindsight, it turns out that the model's ability to explain terrorism is not good enough to model countries simultaneously. Moreover, the parsimonious model of the countries contain different covariate sets. Considering the similar dynamics in terrorist activity, Rapoport was right by recognising a common energy that transcends national boundaries and shapes the characteristics and mutual relationships of terrorist organisations [65]. However, people are vulnerable to terrorist recruitment because of a country specific mixture of circumstances. Although there will be common denominators, it appears that the complexity of the modelled phenomenon does not allow generalisation.

Countries are therefore modelled individually. As section 3.4 proposed, aggregated targets could be incorporated. Unfortunately, the Weibull baseline assumption turned out to be inappropriate for some of the targets. In the end, this study models plain, country data individually, for which the proportionality assumption naturally holds. In the specification of the model, the indices are still kept in the proportional intensity model framework, but note that our model is implemented for one individual.

The NHPP model with multiplicative intensity function for individual i with covariate vector $\mathbf{x_i}$, assumes a NHPP $N_i(t)$ with intensity function

$$\lambda_i(t) = \lambda_0(t) g(\mathbf{x_i}; \boldsymbol{\beta}) \tag{4.12}$$

for $i = 1,2,...,m$. For the rest of this study $m = 1$. Function $g(\mathbf{x_i}; \boldsymbol{\beta})$ is a positive-valued function $g : \mathbb{R}^p \to \mathbb{R}^+$ of the $(p \times 1)$ covariate vector $\mathbf{x_i}$ of individual i, and the $(p \times 1)$ unknown regression coefficients vector $\boldsymbol{\beta}$. The baseline intensity function $\lambda_0(t)$ corresponds to $\lambda_i(t \mid \mathbf{x_i} = 0)$ [113]. Conventional is to use the exponential form

$$g(\mathbf{x_i}; \boldsymbol{\beta}) = \exp(\mathbf{x_i} \boldsymbol{\beta}) \tag{4.13}$$

which provides flexilibility and assures a non-negative intensity function [106]. There are now different ways the model can be solved. The Cox semiparametric approach can be used, that allows to estimate the regression coefficients without specifying the baseline intensity function $\lambda_0(t)$. When reason exists to assume that the baseline intensity follows a particular form, a parametric approach can be pursued. The strong increasing trends in terrorist activity indeed suggest a Weibull model as baseline intensity might be appropriate, such that $\lambda_0(t) = \nu \delta t^{\delta - 1}$ [91]. Parameter $\delta > 0$ is called the shape parameter, that strictly increases (decreases) the intensity for $\delta > 1$ ($\delta < 1$). Note that a constant baseline intensity $\delta = 1$ reduces the model to the standard Poisson regression model of section 4.2.2. Following

Lawless, ν is included in the regression function $e^{\mathbf{x_i}\boldsymbol{\beta}}$ as intercept by defining $\nu = e^{\beta_0}$ and letting covariate $x_0 = 1$ [91]. This results in

$$\lambda_i(t) = \delta t^{\delta-1}\exp(\mathbf{x_i};\boldsymbol{\beta}) \tag{4.14}$$

The cumulated intensity function thus becomes

$$\begin{aligned} \Lambda_i(t) &= \int_0^t \lambda_i(u)du \\ &= t^{\delta}\exp(\mathbf{x_i};\boldsymbol{\beta}) \end{aligned} \tag{4.15}$$

The annual covariate data can be incorporated by letting the path of each covariate follow a step function. Let $\mathbf{x_{iy_t}}$ represent the covariate vector of unit i for year y_t, the year t is situated. With the exponential form for $g(\mathbf{x_{iy_t}};\boldsymbol{\beta})$, analogous to equation 4.13, the model is then defined by

$$\lambda_i(t \mid \mathbf{x_{iy_t}}) = \lambda_{i,0}(t)g(\mathbf{x_{iy_t}};\boldsymbol{\beta}) = \delta t^{\delta-1}x\exp(\mathbf{x_{iy_t}}\boldsymbol{\beta}) \tag{4.16}$$

This nonhomogeneous Poisson process model with Weibull baseline intensity function and piecewise constant covariates is referred to as PPWBPC model in further analysis. Because the covariates are piecewise constant, the cumulative intensity function can then be segmented into a sum of integrals

$$\begin{aligned} \Lambda_{x_i}(t) &= \int_0^t \lambda_{x_i}(u)du \\ &= \int_0^t \delta u^{\delta-1}e^{\mathbf{x_{iy_u}}\boldsymbol{\beta}}du \\ &= e^{\mathbf{x_{i1}}\beta}(1^{\delta}-0^{\delta}) + e^{\mathbf{x_{i2}}\beta}(2^{\delta}-1^{\delta}) + \ldots + e^{\mathbf{x_{iy_t}}\beta}(t^{\delta}-(y_t-1)^{\delta}) \\ &= e^{\mathbf{x_{iy_t}}\beta}(t^{\delta}-(y_{t-1})^{\delta}) + \sum_{k=1}^{y_{t-1}} e^{\mathbf{x_{ik}}\beta}(k^{\delta}-(k-1)^{\delta}) \end{aligned} \tag{4.17}$$

Since $N_i(t)$ is a Poisson process, the cumulative intensity $\Lambda_i(t)$ represents the expected number of events up to time t. In order to justify the use of the PPWBPC model, the following assumptions need to be verified

- Fitting a nonhomogeneous Poisson process

  1. *Analysing occurrences of attacks*
     The first step is to analyse the occurrence of attacks in more detail. The large number of attacks countries experienced in more recent years, as well as clustering due to dependency structures of attacks, are the main reasons of possible conflict

with the Poisson model's assumptions. To overcome this problem, a concept from extreme value theory is applied that reduces the amount of attacks and clustering. The goal is to find the right balance between reducing the amount of attacks such that the Poisson hypothesis holds and remaining an amount that preserves nonhomogeneity.

2. *Testing homogeneity*
   Because the previous step involved reducing the amount of attacks, there is the possibility that the occurrence of attacks gets too rare for some countries. The second step is therefore to test for homogeneity for the reduced country data, hoping to reject the null hypothesis. This is done by a multinomial test, which is based on the result of theorem 4.8 that states for a homogeneous Poisson process, conditioned on the total number of attacks, the joint distribution of attacks over the different intervals is multinomial with equal probabilities. The results show that the reduced country data is clearly not homogeneous. This does however not show that the NHPP model is appropriate for our data. The next step is therefore to perform a statistical test to verify if each country's attack data is consistent with a NHPP.

3. *Testing NHPP*
   A conditional-uniform test based on the Kolmogorov-Smirnov statistic is performed to verify whether NHPP model is appropriate. The country processes are divided into subintervals for which the intensity can be assumed approximately constant. The test can then exploit the conditioning property of a HPP to test whether the arrivals are uniformly distributed within their interval. The test is applied for different levels of data reduction.

4. *Dispersion*
   Finally, the equidispersion property of the Poisson model is checked to prevent increased probability of type I errors.

- Weibull baseline intensity
  Equation 4.36 allows a graphical evaluation of the validity of a Weibull model, by plotting the logarithm of the logarithm of the survival function, against the logarithm of the data. A Weibull baseline intensity function is appropriate when the plotted data is roughly linear.

These assumptions are verified in section 4.3.3. The parameters of the model with piecewiese constant covariates can be estimated with maximum likelihood estimation. Section 4.3.2

derives the joint log-likelihood function of the model. Finally, the strengths and limitations of the model are discussed in section 4.3.4.

4.3.2 Maximum Likelihood Estimation

The section derives the joint log-likelihood function and its derivates of the PPWBPC model, required for the estimation and evaluation of the regression coefficients with maximum likelihood estimation (MLE). Let individual i with given covariate vector $\mathbf{x_i}$ experience events according to a nonhomogeneous Poisson process with intensity λ_i. From theorem 4.12 and equation 4.5 follows that the intervals between successive events are independently distributed and that the probability that, starting from t_{ij}, the next event occurs in $(t_{i(j+1)}, t_{i(j+1)} + \Delta t]$, equals

$$\lambda_i(t_{i(j+1)}) \exp\left\{ -\int_{t_{ij}}^{t_{i(j+1)}} \lambda_i(u) du \right\} \Delta t + o(\Delta t) \tag{4.18}$$

The likelihood for unit of analysis i can then easily be derived. From equation 4.18 follows the joint likelihood function for unit of analysis i

Proposition 4.13 (Joint likelihood for unit of analysis i)**.** The joint likelihood for unit of analysis i observed from $[0,T]$, given attack times $0 = t_0 < t_{i1} < t_{i2} < ... < t_{in_i} \leq T$ and covariate vector $\boldsymbol{x}_i$, is defined as

$$L_i(\delta, \beta) = \prod_{j=1}^{n_i} \lambda_i(t_{ij}) \exp\left\{ -\Lambda_i(T) \right\} \tag{4.19}$$

Proof. Using equation 4.18 and independence of observations (note that the last exponential calculates the probability of no event between t_{in_i} and T)

$$\begin{aligned} L_i(\delta, \beta) =& \lambda_i(t_{i1}) \exp\left\{ -\int_0^{t_{i1}} \lambda_i(u) du \right\} \cdot \lambda_i(t_{i2}) \exp\left\{ -\int_{t_{i1}}^{t_{i2}} \lambda_i(u) du \right\} \dots \\ & \lambda_i(t_{in_i}) \exp\left\{ -\int_{t_{in_{i-1}}}^{t_{in_i}} \lambda_i(u) du \right\} \cdot \exp\left\{ -\int_{t_{in_i}}^{T} \lambda_i(u) du \right\} \\ =& \prod_{j=1}^{n_i} \lambda_i(t_{ij}) \exp\left\{ -\int_0^T \lambda_i(u) du \right\} \\ =& \prod_{j=1}^{n_i} \lambda_i(t_{ij}) \exp\left\{ -\Lambda_i(T) \right\} \end{aligned}$$

□

The joint likelihood function of the PPWBPC model is then easily derived

Theorem 4.14 (Joint likelihood of the PPWBPC model)**.** *Consider the conditions of theorem 4.13 for all $i = 1, 2, ..., m$. Then the joint likelihood function is defined as*

$$L(\delta,\beta) = \prod_{j=1}^{n_i} L_i(\delta,\boldsymbol{\beta}) = \prod_{i=1}^{m}\left\{\prod_{j=1}^{n_i}\lambda_i(t_{ij})\right\}\exp\{-\Lambda_i(T)\} \tag{4.20}$$

For various reasons such as computational convenience, numerical stability and Fisher information (see equation), it is more convenient to work with the logarithm of the likelihood. This is possible since the logarithm is a monotonic transformation preserving order. Let $n = \sum_{i=1}^{m} n_i$ be the total number of attacks. The joint log-likelihood is derived in theorem 4.15.

Theorem 4.15 (Joint log-likelihood of the PPWBPC model)**.** *Consider the conditions of theorem 4.13 for all $i = 1, 2, ..., m$. Then the joint log-likelihood function is defined as*

$$\begin{aligned} l(\delta,\beta) = n\log(\delta) + (\delta-1)\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log(t_{ij}) \\ + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathbf{x_{ij}}\beta - \sum_{i=1}^{m}\sum_{k=1}^{T}\exp(\mathbf{x_{ik}}\beta)\left(k^{\delta} - (k-1)^{\delta}\right) \end{aligned} \tag{4.21}$$

Proof.

$$\begin{aligned} l(\delta,\beta) &= \log\left(\prod_{i=1}^{m}\left\{\prod_{j=1}^{n_i}\lambda_{x_i}(t_{ij})\right\}\exp\{-\Lambda_i(T)\}\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\lambda_{x_{iy_{t_{ij}}}}(t_{ij})\right) + \sum_{i=1}^{m}\log\left(\exp\{-\Lambda_i(T)\}\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\delta {t_{ij}}^{\delta-1}\exp(\mathbf{x_{iy_{t_{ij}}}}\beta)\right) - \sum_{i=1}^{m}\Lambda_i(T) \end{aligned} \tag{4.22}$$

The left term can easily be split into three smaller terms

$$\begin{aligned} \sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\delta {t_{ij}}^{\delta-1}\exp(\mathbf{x_{iy_{t_{ij}}}}\beta)\right) &= n\log(\delta) + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left({t_{ij}}^{\delta-1}\right) + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\exp(\mathbf{x_{iy_{t_{ij}}}}\beta)\right) \\ &= n\log(\delta) + (\delta-1)\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log(t_{ij}) + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathbf{x_{iy_{t_{ij}}}}\beta \end{aligned}$$

The right term can be simplified by making use of equation 4.17, segmenting the cumulative intensity into a sum of integrals where the exponentials are known

$$\begin{aligned}\Lambda_i(T) &= \int_0^T \delta u^{\delta-1} \exp(\mathbf{x_{iy_u}}\beta) du \\ &= \exp(\mathbf{x_{i1}}\beta) \int_0^1 \delta u^{\delta-1} du + \exp(\mathbf{x_{i2}}\beta) \int_1^2 \delta u^{\delta-1} du + \ldots + \exp(\mathbf{x_{iT}}\beta) \int_{T-1}^T \delta u^{\delta-1} du \\ &= \sum_{k=1}^T \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta}) \left(k^\delta - (k-1)^\delta\right)\end{aligned} \tag{4.23}$$

The right term of equation 4.22 thus becomes

$$\sum_{i=1}^m -\Lambda_i(T) = -\sum_{i=1}^m \sum_{k=1}^T \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta}) \left(k^\delta - (k-1)^\delta\right) \tag{4.24}$$

□

For notational convenience, let $\theta = (\delta, \beta_1, \beta_2, \ldots, \beta_p)$ be the $\big((p+1)\times 1\big)$ vector of regression parameters. The maximum likelihood estimate is obtained by maximising the joint log-likelihood function of theorem 4.15 and reflects the parameter values that are most likely to have produced the data

$$\hat{\theta}_{MLE} = \underset{\delta,\beta}{\arg\max}\, l(\delta, \boldsymbol{\beta}) = \underset{\delta,\beta}{\arg\min} -l(\delta, \boldsymbol{\beta}) \tag{4.25}$$

The quality of the parameter estimates can be derived from the observed Fisher information, which is obtained by taking the negative of the second derivative of the log-likelihood function. The derivatives can be derived by simple calculus, resulting in equations 4.26 - 4.30.

$$\frac{\partial l}{\partial \beta_r} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij;r} - \sum_{i=1}^{m}\sum_{k=1}^{T} x_{ij;r}\exp(\boldsymbol{x}_{ik}\boldsymbol{\beta})\left(k^{\delta} - (k-1)^{\delta}\right) \tag{4.26}$$

$$\frac{\partial l}{\partial \delta} = \frac{n}{\delta} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log(t_{ij}) - \sum_{i=1}^{m}\sum_{k=1}^{T} \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta})\left(k^{\delta}\log(k) - (k-1)^{\delta}\log(k-1)\right) \tag{4.27}$$

$$\frac{\partial^2 l}{\partial \beta_r \partial \beta_s} = -\sum_{i=1}^{m}\sum_{k=1}^{T} x_{ij;r}x_{ij;s}\exp(\boldsymbol{x}_{ik}\boldsymbol{\beta})\left(k^{\delta} - (k-1)^{\delta}\right) \tag{4.28}$$

$$\frac{\partial^2 l}{\partial \delta^2} = -\frac{n}{\delta^2} - \sum_{k=1}^{T} \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta})\left(k^{\delta}\log(k)^2 - (k-1)^{\delta}\log(k-1)^2\right) \tag{4.29}$$

$$\frac{\partial^2 l}{\partial \delta \partial \beta_r} = -\sum_{i=1}^{m}\sum_{k=1}^{T} x_{ij;r}\exp(\boldsymbol{x}_{ik}\boldsymbol{\beta})\left(k^{\delta}\log(k) - (k-1)^{\delta}\log(k-1)\right) \tag{4.30}$$

The regression coefficients can now be estimated by standard convex optimisation techniques and their significance can be evaluated by utilising the Fisher information.

4.3.3 Verifying assumptions

4.3.3.1 Analysing the occurrences of attacks

The large number of attacks countries experienced in more recent years, as well as clustering due to dependency structures of attacks, are the main reasons of possible conflict with the Poisson model's assumptions. Figure 4.3, shows the daily number of that shows the daily number of attacks for Iraq and Yemen, the two opposite extremes in terrorist activity, in December 2014. Notice how, especially for Iraq, the occurrence of attacks is not a rare event and that attacks occur in intermittent bursts over multiple days.

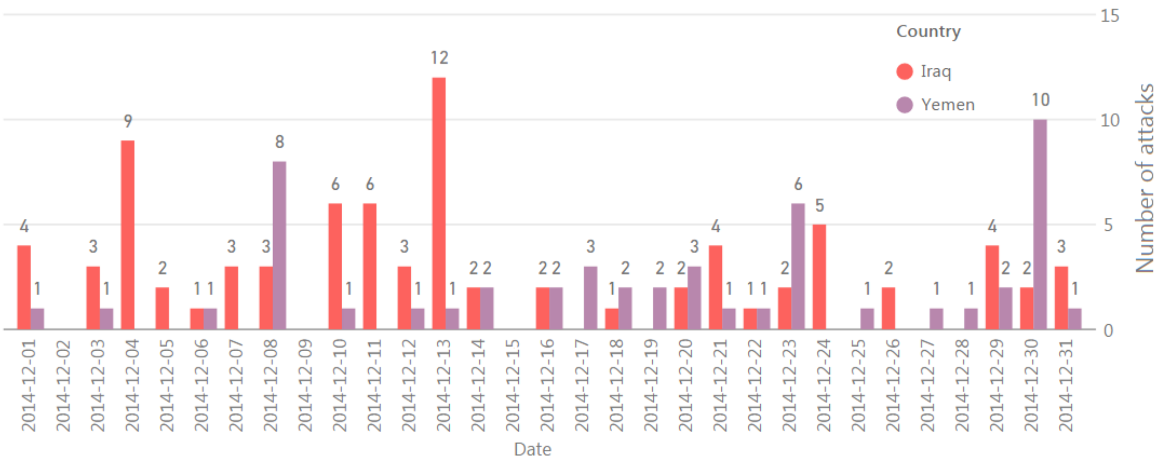

Fig. 4.3 Daily number of attacks December 2014. The occurrences of attacks can, especially for Iraq, not be considered as a rare event. In addition, there is clear clustering of attacks.

There are different ways this problem can be overcome. Recall the GTD includes information of the amount of deaths per attack. For the proposed data selection, this information is available for 10812 attacks, covering 95.9% of the total 11279 attacks. Figure 4.4 shows the distribution of lethality of terrorist attacks. A natural choice is to model (extreme) events which generate more deaths than a certain threshold x.

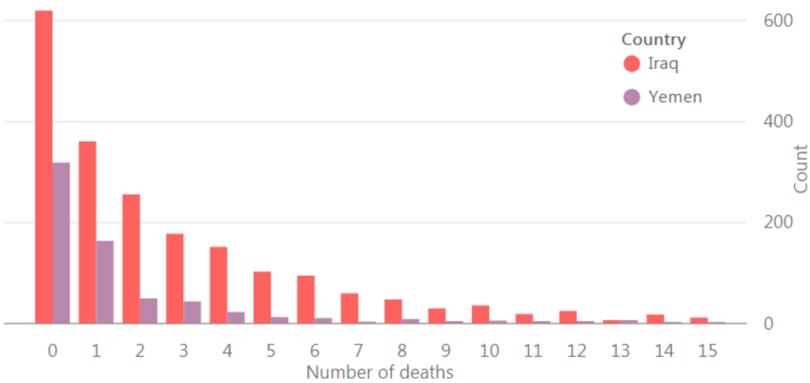

Fig. 4.4 Histogram number of deaths for Iraq and Yemen

Applying threshold x, the goal is to find the right balance between reducing the amount of attacks such that the Poisson hypothesis holds and remaining an amount that preserves nonhomogeneity. A proper value for x is a priori not known and needs to be determined by performing statistical tests that assess to what extent the Poisson model fits our data, see section 4.3.3.3 and 4.3.3.4. This section will provide a rough estimation of appropriate thresholds by graphical evaluation.

Certainly, the 3413 events (31.6% from the total 10812) that generate no deaths will be removed, leaving 7399 attacks. Note Clauset and Woodard show that the distribution of lethality of a similar group of terrorist attacks follows the Pareto distribution [33]. Indeed, our data (7399 attacks for the six countries of interest) can be fit by a Pareto distribution with scale parameter $x_m = 9.844$ and shape parameter $\theta = 2.558$ (for CDF $F(x) = 1 - (\frac{x_m}{x})^{\theta}$). Since $\theta = 2.558$ this distribution has a finite mean, but not a finite variance.

Figure 4.5 shows the decrease in attacks for increasing thresholds in 2013 and 2014, the years facing the most terrorist activity. When the attacks in these two years satisfy the Poisson hypothesis and show limited clustering, it is safe to assume that this holds for the previous years. Roughly estimated, because there 365 (366 for leap years) days in a year, reducing the country data to 200-300 attacks per year will probably be sufficient. The bottom three countries are therefore expected to be reasonably fit by a NHPP for $0 \leq x \leq 4$. The top three countries require a higher threshold somewhere between $4 \leq x \leq 8$. Notice that for this range of threshold, the relation between the countries changes. Nigeria is now the country showing the highest terrorist activity in 2014, exceeding Iraq and Afghanistan. This is explained by the lethality of attacks in Nigeria, see the world map of figure 3.4.

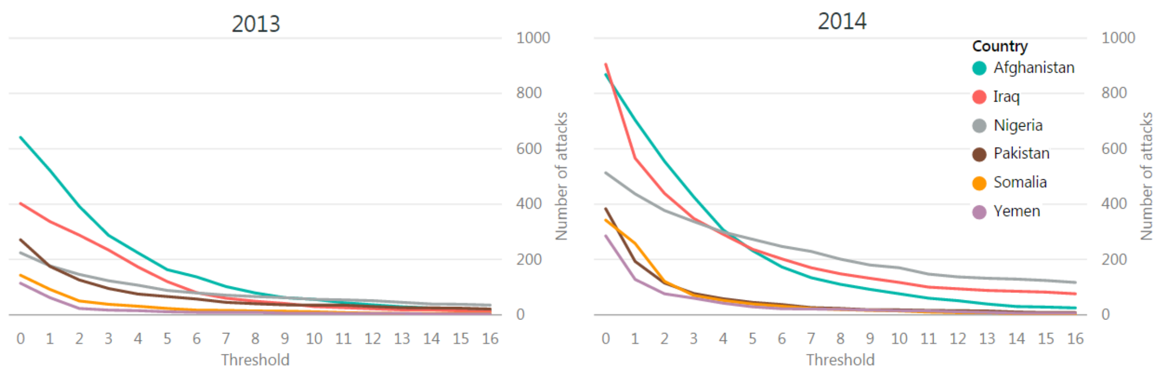

Fig. 4.5 Number of attacks by country, with number of casualties x as distriminant quantity. Thresholds will need to be applied for the NHPP model to be appropriate.

To gain more insight into clustering of attacks for reduced country data, figure 4.6 shows the daily number of attacks in Nigeria for $x = \{4, 8\}$, over the first four months of 2014. For $x = 4$ there are periods with heavy clustering, as illustrated by the month March, showing five days with five to six attacks per day. Applying higher thresholds reduces this considerably and seems to enforce the attack data to satisfy the Poisson assumptions.

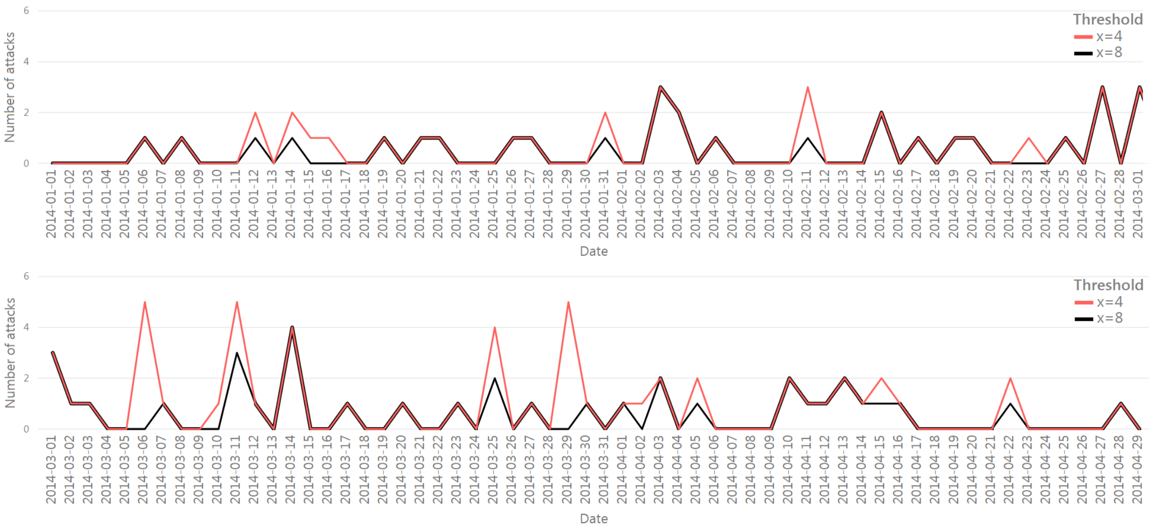

Fig. 4.6 Daily number of attacks Nigeria, for threshold x=4 and x=8

At the same time, the procedure causes major changes in the data. For $x = 4$, the data is reduced by 7603 attacks (70.4%), leaving 3209 attacks. Moreover, $x = 8$ removes 9777 attacks (90.4%) leaving just 1035 attacks for the six countries between 2000 and 2014. As figure 4.7 shows, the strong increasing trends present before applying thresholds (see figure 4.8), tend to disappear for the lower three countries Pakistan, Somalia and Yemen. The PPWBPC model however assumes that each country has attacks occur according to a NHPP. Since the HPP is a special case of the NHPP (constant intensity), the tests used in section 4.3.3.3 and 4.3.3.4 to see if the countries can be fit by NHPPs, also apply to the HPP. The next step is therefore to verify that applying the proposed thresholds preserves nonhomogeneity.

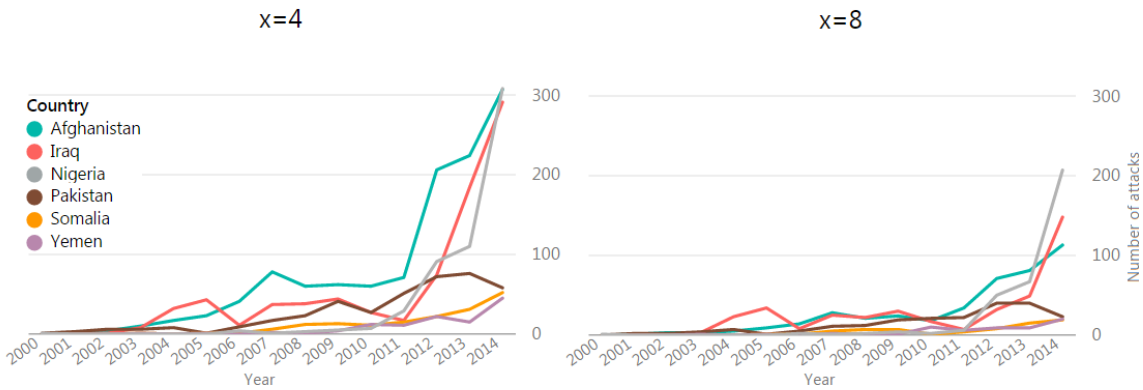

Fig. 4.7 Number of attacks by country, with at least 4 (left) or 8 (right) deaths per attack

The difference of the trends of the upper three countries (Afghanistan, Iraq and Nigeria) compared to the lower three (Pakistan, Somalia and Yemen) that arises after removing less lethal attacks, hints that the two groups of countries should rather be modelled individually. Besides, the lower three countries do probably not require a threshold above $x = 4$ for the NHPP model to be appropriate. In a later stage, it will be shown that the countries are better modelled individually.

Concluding, taking the number of casualties as discriminant quantity provides a way to improve the countries fit by NHPPs, without losing relevance from a sociological perspective. Rough estimation of appropriate thresholds show Afghanistan, Iraq and Nigeria require a threshold somewhere between of $4 \leq x \leq 8$, while Pakistan, Somalia and Yemen can go with the more liberal $0 \leq x \leq 4$. Applying thresholds does however considerably reduce the data. The next step is therefore to check if for the proposed thresholds, the occurrences of events are not so rare that the country processes are homogeneous.

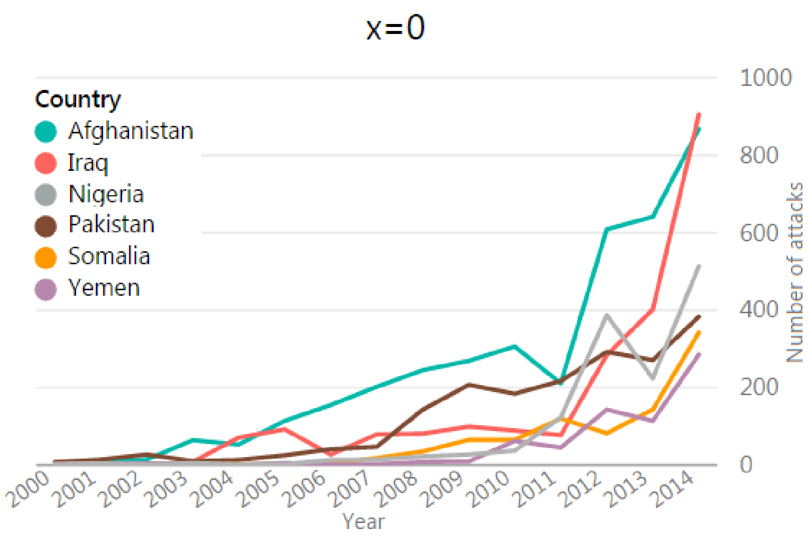

Fig. 4.8 Number of attacks by country, no threshold

4.3.3.2 Testing homogeneity

Figures 4.7 and 4.8 show that in all three threshold cases, $x = \{0, 4, 8\}$, the main three country are clearly not homogeneous Poisson. The same holds for the lower three countries when no threshold is applied. But what about Somalia and Yemen, for $x = 4$ and $x = 8$?

This can by tested with the multinomial test, which is based on the following result of theorem 4.8: for a homogeneous Poisson process, the joint distribution of attacks over the different intervals, conditioned on the total number of attacks, is multinomial with equal probabilities. In other words, the aim is to test the hypothesis H_0 : the data come from a multinomial distribution where the cells have equal probability of observing an attack, against the alternative hypothesis H_1 : the data come from a multinomial distribution where the cells have unequal probabilities of experiencing events. If Somalia and Yemen are homogeneous Poisson, the null hypothesis should not be rejected.

The timeframe is split up in seven disjoint intervals, 2001-2002, 2003-2004, ... , 2013-2014, where the year 2000 is omitted. Figure 4.10 shows the observed number of attacks for each country, for the three threshold cases. Comparing the trends between countries for $x = 8$, it becomes clear that the most plausible candidates for a HPP are Somalia and Yemen. If these countries are not homogeneous Poisson, then neither are the others (Pakistan is a bit less certain and needs to be checked). Likewise, if the countries are not homogeneous Poisson for $x = 8$, then for sure they are not for lower thresholds.

Therefore, the Pearson's chi-squared test is used to check whether the observed disitributions of Somalia and Yemen, $x = 8$, differ from a multinomial distribution with equal probabilities. The result are clear, the null hypothesis is rejected, i.e. Somalia and Yemen are not homogeneous Poisson for $x = 8$. The results of the chi-squared test are given in table 4.1, the probabilities of all possible events are given in figure 4.9.

Table 4.1 Results chi-squared test Somalia and Yemen, threshold $x = 8$. The countries are clearly not homogeneous Poisson.

	Somalia $x = 8$			Yemen $x = 8$		
Interval	Expected	Observed	$\frac{(x_i - E_i)^2}{E_i}$	Expected	Observed	$\frac{(x_i - E_i)^2}{E_i}$
2001-2002	67/7	0	9.57	60/7	0	8.57
2003-2004	67/7	0	9.57	60/7	1	6.69
2005-2006	67/7	2	5.98	60/7	1	6.69
2007-2008	67/7	12	0.62	60/7	3	3.62
2009-2010	67/7	8	0.26	60/7	11	0.69
2011-2012	67/7	12	0.62	60/7	15	4.82
2013-2014	67/7	33	57.34	60/7	29	48.67
χ^2			83.97			79.77
p-value			$1.11 * 10^{-16}$			$8.89 * 10^{-16}$

This means all six countries are not homogeneous Poisson for all thresholds $x \leq 8$. (Pakistan for $x = 8$ has $\chi^2 = 870.89$ and p-value below machine precision, i.e. clearly not homogeneous Poisson). Now it is sure that the country processes are nonhomogeneous, the step is to assess to what extend each country's attack data is consistent with a NHPP.

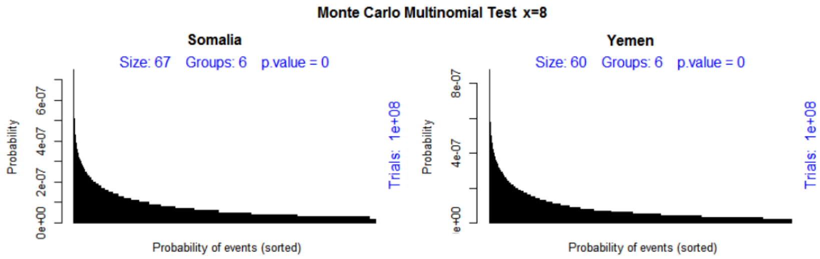

Fig. 4.9 Probabilities of all possible events for Somalia and Yemen, $x = 8$

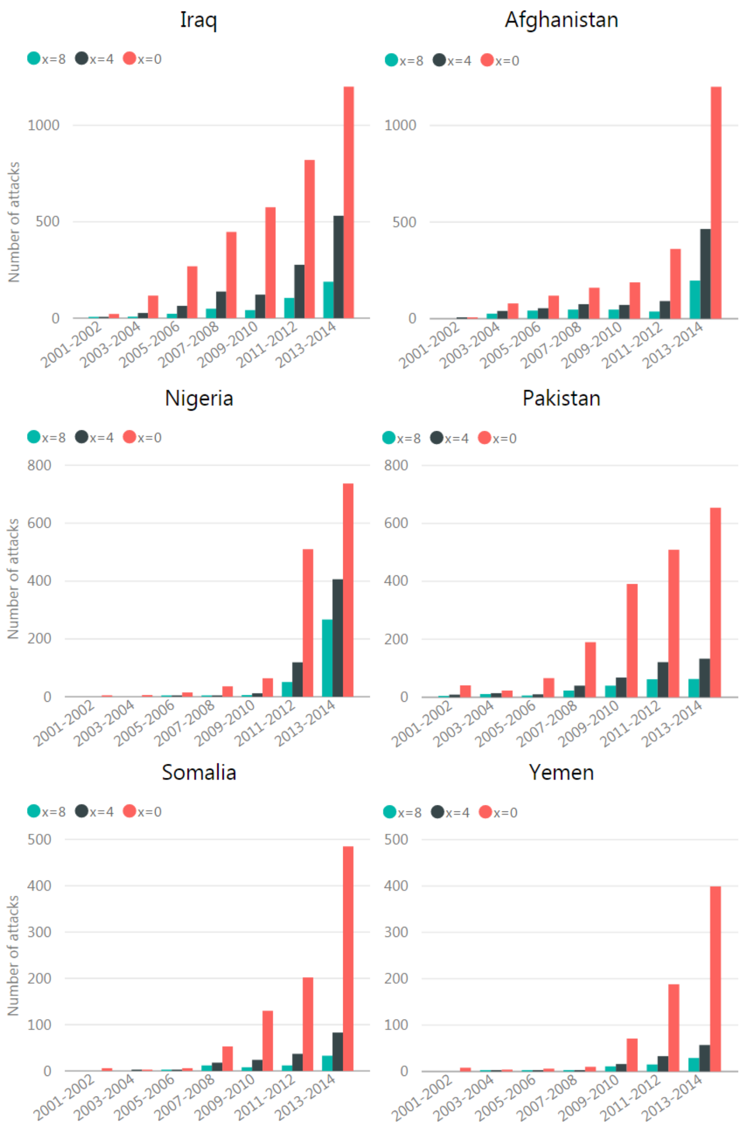

Fig. 4.10 Number of attacks observed per cell by country, for thresholds x=0,4,8

4.3.3.3 Testing NHPP

Section 4.3.3.2 used the conditioning property of a HPP, that states that conditioned on the total number of attacks, the joint distribution of attacks over the different intervals is multinomial with equal probabilities. Kim and Whitt review various tests that exploit this principle by transforming the data and applying a Kolmogorov-Smirnov (KS) test [130]. The KS test is a nonparametric test that can be used for goodness of fit of the empirical distribution function of a sample and a hypothetical distribution. For an i.i.d. sequence of random variables $X_1,\ldots,X_n$, the empirical cumulative distribution function (ecdf) is defined as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \tag{4.31}$$

which represents the proportion of observations less than or equal to some x. Kolmogorov and Smirnov defined the following statistic that quantifies the distance between the empirical distribution and the hypothetical distribution

Definition 4.16 (Kolmogorov-Smirnov Statistic)**.**

$$D_n = \sup_x\{\mid F_n(x) - F(x) \mid\} \tag{4.32}$$

If $F(x)$ is continuous the distribution of D_n does not depend on F [130]. In particular, Kolmogorov showed that if F is continous, then under the null hypothesis $\sqrt{n}D_n$ converges to the Kolmogorov-Smirnov distribution

$$F_K(t) = \frac{\sqrt{2\pi}}{x}\sum_{k\geq 1} e^{-(2k-1)^2\pi^2/(8x^2)} \tag{4.33}$$

The Kolmorgorov-Smirnov test is then constructed by using the critical values of the Kolmogorov-Smirnov distribution [57], rejecting the null hypothesis at significance level α if

$$\sqrt{n}D_n > k_{1-\alpha}$$

where $k_{1-\alpha}$ is the $(1-\alpha)$ quantile of the Kolmogorov distribution.

The essence of the conditional uniform (CU) KS test used in this section, is that it makes use of the conditional distribution property of the arrival times of the HPP, by assuming the NHPP can be approximated by a piecewise constant NHPP. The NHPP on $[0,t]$ is divided in K subintervals (possibly of equal length), such that within each subinterval, the intensity is approximately constant. Consider the ith subinterval $\Delta_i = [L_{i-1}, L_i]$ with n_i arrivals, where

$L_{i-1} \leq t_{i1} < \ldots < t_{in_i} \leq L_i$. Under the null hypothesis of a NHPP (assuming the intensity is approximately constant within each interval), the conditional property of the arrival times of a HPP gives that the arrivals in subinterval Δ_i are random variables distributed as the order statistics of i.i.d. random variables uniformly distributed over $[L_{i-1}, L_i]$ [130]. Dividing the arrivals by the length of subinterval Δ_i, results in n_i random variables which are distributed as the order statistics of i.i.d. random variables uniformly distributed over $[0,1]$ [135]. Doing this for all subintervals, the data can then be combined into a single ordered sequence of i.i.d. random variables uniformly distributed on $[0,1]$. The empirical cdf of this sequence is calculated by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \mathbb{1}_{\{\frac{t_{ij}}{L_i - L_{i-1}} \leq x\}} \tag{4.34}$$

where $n = \sum_{i=1}^{K} n_i$ is the total number of attacks, K the number of subintervals and $x \in [0,1]$. Then the KS test can be performed with uniform cdf $F(x) = x$, testing the null hypothesis $H_0 : F(x) = \hat{F}_n(x)$, i.e. the subintervals are HPP and the entire process is a NHPP.

Since the attack times are rounded to days, there exist many interarrivals of length zero. These do not occur for a NHPP and therefore the CU KS test might unfairly reject the null hypothesis [131]. Brown solves the negative effect of data rounding by adding independent uniform noise to each observation [101]. Terrorist attacks are however more likely to be comitted at times of high human activity. This is be approximated by a normal distribution around 12:00 p.m. normalised between 0:00 a.m. and 23:59 p.m. (note a more accurate approximation of reality would have a distribution skewed to the right). The CU KS test is performed for 1000 samples, for a variety of thresholds.

The big advantage of the CU KS test is that the method is independent of the intensity of the Poisson process. Moreover, the test allows easy and direct calculation of goodness of fit for 1000 samples of arrival data, for different thresholds. However, in hindsight, considering the fact that the country data is fit by a NHPP with a Weibull baseline intensity, this might cause confusion. Although the covariates will naturally be constant over certain intervals because they are piecewise constant, the Weibull component $\delta t^{\delta-1}$ will only be constant when $\Delta_i \to 0$. Furthermore, Kim and Whitt found that the CU KS test of NHPP has lower power than other tests considered, letting alternatives pass too easily.

A better option would have been to use a standard tests based on the Cramer-von-Mises or KS test as found in [103]. Another possibility is to use one of the Monte Carlo goodness-of-fit tests for parametric NHPPs by Lindqvist and Rannestad [48]. Moreover, they introduce a specific goodness-of-fit test for a power-law NHPP that could be useful for a future publication of this study.

Nevertheless, the results for the CU KS test are presented, as they are convincing and in line with section 4.3.3.4, that checks the dispersion of the country data for different thresholds. After testing the data, it turns out that monthly subintervals are sufficient for the intensity to be approximately constant. Performing the test for 1000 samples of attack data (adding the independent uniform noise to each attack), for thresholds $0 \leq x \leq 15$, results in the p-values of figure 4.12.

There is a clear difference between the countries experiencing relatively high terrorist activity (Afghanistan, Iraq and Nigeria) and low amounts of terrorism (Pakistan, Somalia and Yemen). As section 4.3.3.1 predicted, the upper three countries do not satisfy the NHPP null hypothesis when no threshold is applied. Table 4.2 shows the test results of appropriate thresholds. For Afghanistan (p-value mean 0.52, minimum 0.24) and Iraq (p-value mean 0.50, minimum 0.28) a threshold of $x = 4$ clearly suffices. Nigeria requires a much larger threshold of at least $x = 13$ (p-value mean 0.32, minimum 0.21), to be on the safe side. Figure 4.11 shows the CU KS tests results for a larger range of thresholds. Observe that the attacks in Nigeria start to be consistent with the Poisson assumptions for very high thresholds, caused by the fact that attack in Nigeria are highly lethal (on average 9 deaths per attack).

Country	**Threshold**	μ	σ^2	**min**	**max**
Afghanistan	$x = 4$	0.52	$6.88 * 10^{-3}$	0.24	0.77
Iraq	$x = 4$	0.50	$7.41 * 10^{-3}$	0.28	0.76
Nigeria	$x = 13$	0.32	$1.59 * 10^{-3}$	0.21	0.45
Pakistan	$x = 2$	0.41	$4.36 * 10^{-3}$	0.22	0.63
Somalia	$x = 2$	0.64	$5.39 * 10^{-3}$	0.43	0.85
Yemen	$x = 2$	0.62	$4.40 * 10^{-3}$	0.40	0.80

Table 4.2 CU KS tests p-values statistics for reasonable thresholds

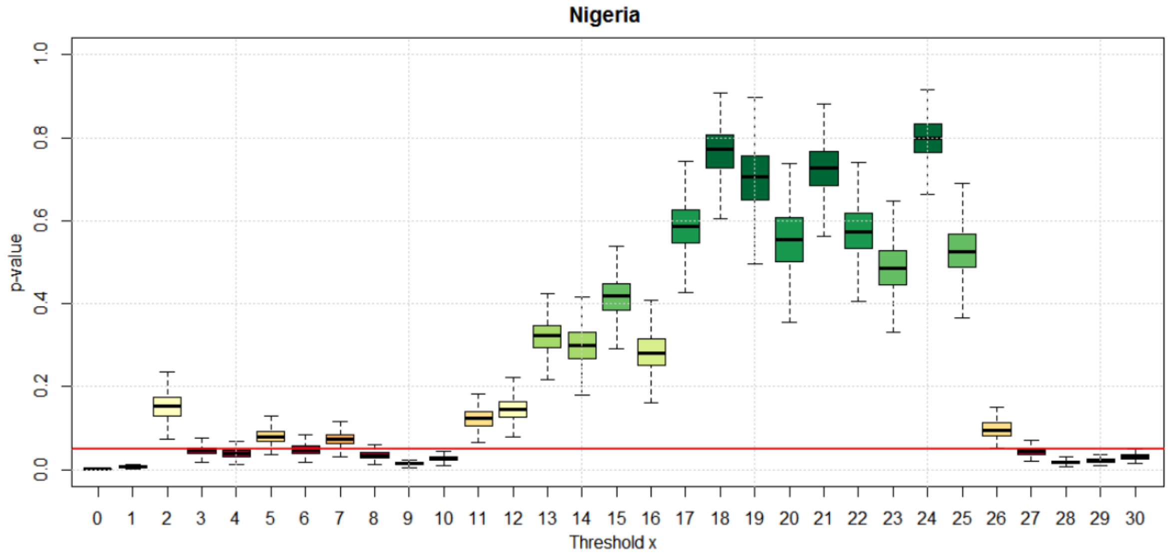

Fig. 4.11 CU KS test for Nigeria. Because of high lethality of attacks, Nigeria requires relatively very high thresholds in order for a NHPP to be appropriate.

Pakistan (p-value mean 0.41, minimum 0.22) and even more convincingly Somalia (p-value mean 0.64, minimum 0.43) and Yemen (p-value mean 0.62, minimum 0.40), satisfy the NHPP null hypothesis when applying a threshold of $x = 2$. The higher thresholds for the lower three countries should be ignored, since these result in point processes that suit extreme value theory rather than a NHPP model.

Summarising, the attacks of the countries were shown to follow the NHPP null hypothesis when removing attacks that do not generate at least a country-specific number of x deaths. With exception of Nigeria, all countries are consistent with a NHPP when applying threshold $x = 4$. Another possibility is to apply $x = 4$ to Afghanistan and Iraq, whereas Pakistan, Somalia and Yemen are modelled with $x = 2$. Nigeria is a special case with an abundance of extreme attacks and requires at least $x = 13$. In section 4.3.3.4, the results are compared to the dispersion that the data of the countries exhibits to solidify the justification of using a NHPP model.

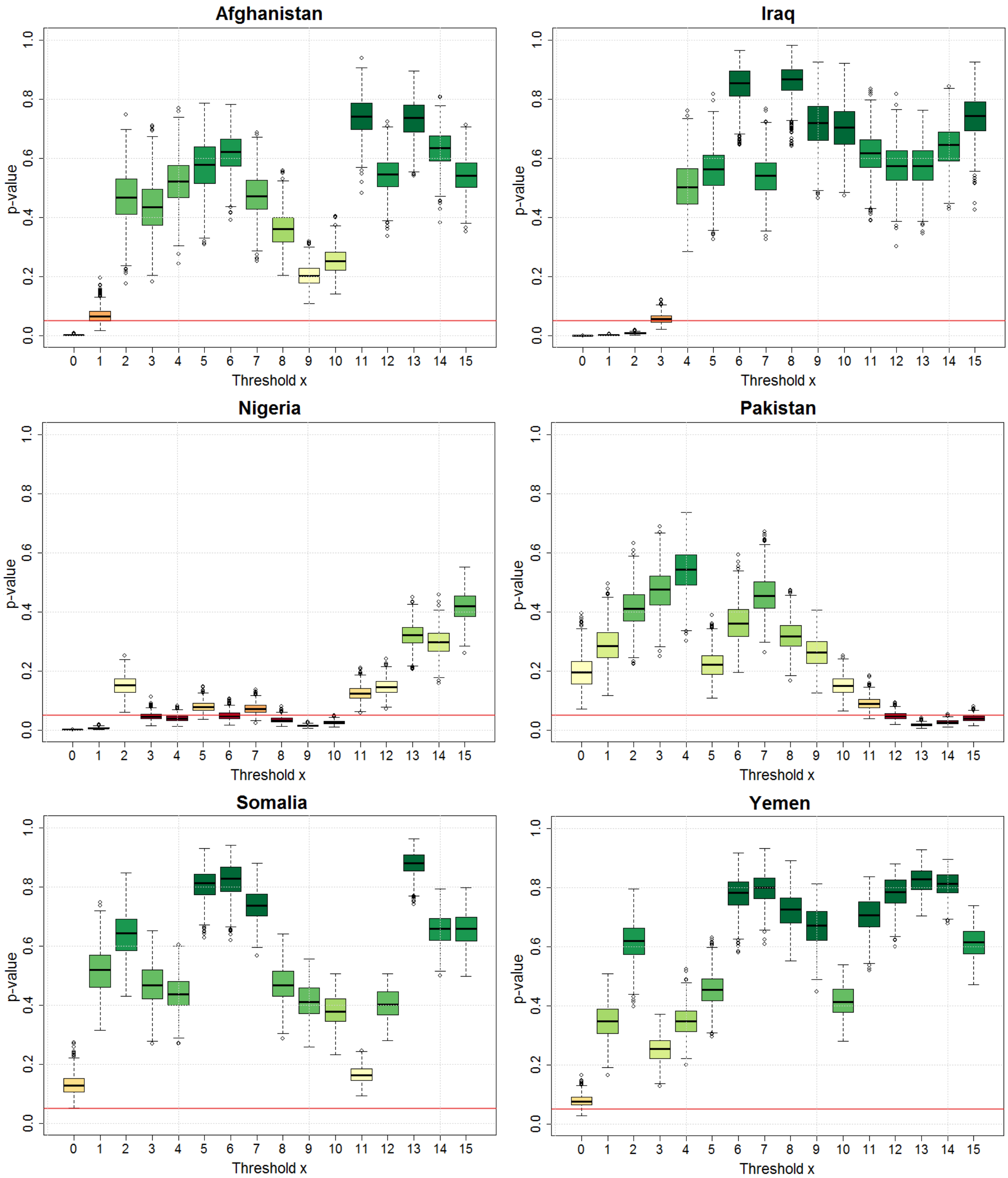

Fig. 4.12 CU KS test results for different thresholds. As expected, Afghanistan, Iraq and Nigeria require a higher threshold than Pakistan, Somalia and Yemen in order to satisfy the NHPP null hypothesis. Nigeria is a however special case, for the country experiences high amounts of extreme attacks as compared to the other countries.

4.3.3.4 Dispersion

Section 4.2.1 discussed the restrictive equidispersion property of the Poisson process. In practice, count data rarely exhibits equidispersion, often because the observed variance is larger than the assumed variance, a phenomenon called overdispersion [92]. Indeed, the clustering of attacks observed in section 4.3.3.1 indicates overdispersion. Although regression coefficients of the PPWBPC model are still reliable in case of overdispersion, the estimated standard errors are artificially smaller than the true standard errors [111]. Not accounting for overdispersion might therefore increase the probability of type I errors.

An indication of dispersion of the country data is given in figure 4.13, where the index of dispersion $D = \frac{\sigma^2}{\mu}$ is given over time for different thresholds (number of deaths as discriminant quantity). The results more or less are in line with CU KS test results of section 4.3.3.3. Pakistan, Somalia and Yemen show acceptable overdispersion for a threshold of $x = 2$, with a variance of at most 1.4 times the mean, see table 4.3. The proposed threshold for Afghanistan, $x = 4$, exhibits a reasonable dispersion index below the common heuristic 1.6.

Country	**Threshold**	$<$ **2009**	$<$ **2010**	$<$ **2011**	$<$ **2012**	$<$ **2013**	$<$ **2014**	$<$ **2015**
Afghanistan	$x = 4$	1.39	1.42	1.45	1.48	1.50	1.51	1.53
Iraq	$x = 4$	2.61	2.63	2.65	2.67	2.68	2.69	2.70
	$x = 8$	1.52	1.54	1.55	1.56	1.56	1.57	1.57
Nigeria	$x = 13$	2.16	2.17	2.17	2.18	2.18	2.19	2.19
	$x = 24$	1.52	1.52	1.52	1.53	1.53	1.43	1.53
Pakistan	$x = 2$	1.16	1.18	1.20	1.21	1.23	1.24	1.25
Somalia	$x = 2$	1.15	1.16	1.16	1.17	1.17	1.18	1.18
Yemen	$x = 2$	1.36	1.37	1.38	1.38	1.38	1.39	1.39

Table 4.3 Index of dispersion for different time intervals, e.g. $<$ 2009 reflects the dispersion calculated in the interval 01/01/2001 - 31/12/2009. Results for Pakistan, Somalia, Yemen and Afghanistan are in line with the CU KS test results from section 4.3.3.3. Iraq requires a higher threshold of $x = 8$. The NHPP model turns out not to be appropriate for Nigeria.

On the other hand, dispersion for Nigeria is problematic. Even for $x = 13$ the variance is more than twice the mean. Applying extreme thresholds such as $x = 24$ does results in more acceptable dispersion below 1.6, but leads to a situation that might rather suit extreme value theory than a NHPP model. Modelling Nigeria by the PPWBPC model is therefore not pursued. Finally, Iraq requires a higher threshold $x = 8$ than the $x = 4$ suggested by the CU KS test, to get the dispersion index below 1.60 instead of 2.70.

Concluding, the CU KS test results from section 4.3.3.3 combined with the dispersion the countries exhibit, show that removing attacks with excesses in deaths for Afghanistan (x=4), Iraq (x=8), Pakistan (x=2), Somalia (x=2) and Yemen (x=2), is effective against clustering and yields country point processes consistent with NHPPs. The attack data of Nigeria is not appropriate for the PPWBPC model.

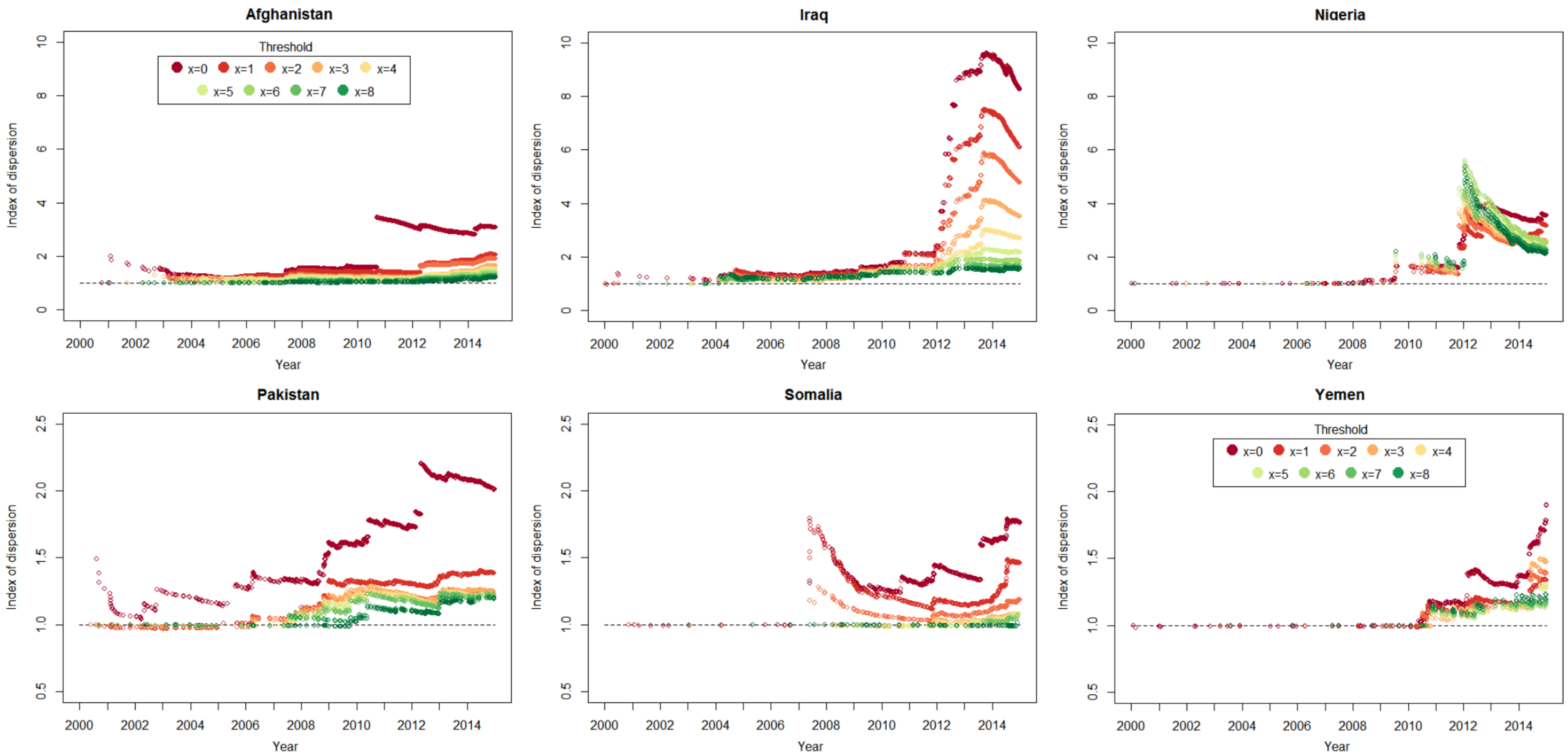

Fig. 4.13 Index of dispersion over time for different thresholds. Note the difference in scaling between the upper and lower three countries. The results are reasonably in line with the CU KS test results of section 4.3.3.3; Pakistan, Somalia and Yemen show acceptable dispersion for $x = 2$, for Afghanistan $x = 4$ suffices. Iraq and especially Nigeria exhibit significant overdispersion and larger data reductions are required.

4.3.3.5 Weibull baseline intensity

Prior to fitting a parametric model, the validity of assuming a specific baseline intensity function should be analysed [145]. One way to justify the use of a Weibull model as baseline involves the survival function. Recall the survival function for a Weibull distribution with hazard $\lambda_i(\delta,\nu,t) = \nu\delta t^{\delta-1}$, ($\delta$ and ν being the scale and shape parameters, respectively), is defined as

$$S_i(t) = 1 - F_i(t) = \exp(-\int_0^t \lambda_i(\delta,\nu,u)du) = \exp(-\nu t^{\delta}) \tag{4.35}$$

Taking the logarithm, multiplying by -1 and then taking the logarithm again, results in

$$\log(-\log(1-F_i(t))) = \log(\nu) + \delta\log(t) \tag{4.36}$$

The relation of equation 4.36 allows a graphical evaluation of the validity of a Weibull model, by plotting $y = \log(-\log(1-\hat{F}_i(t)))$ versus $x = \log(\nu) + \delta\log(t)$. Nelson [60] suggests to estimate the empirical cumulative distribution function $\hat{F}_i(t)$ by

$$\hat{F}_i(t_{ij}) = \frac{k-0.3}{n_i+0.4} \tag{4.37}$$

with k the rank of the j-th attack and n_i the total number of attacks of unit i. A Weibull baseline intensity function is appropriate when the plotted data is roughly linear. In addition, when lines of different units exhibit approximately the same slope, the proportional intensity assumption is reasonable [95].

Section 4.3.3 showed that the country attack data of Pakistan, Somalia and Yemen are consistent with NHPPs after removing attacks that generated less than two deaths. For Afghanistan, a threshold $x = 4$ was proposed. Figure 4.14 shows the Weibull plots for both thresholds including these countries.

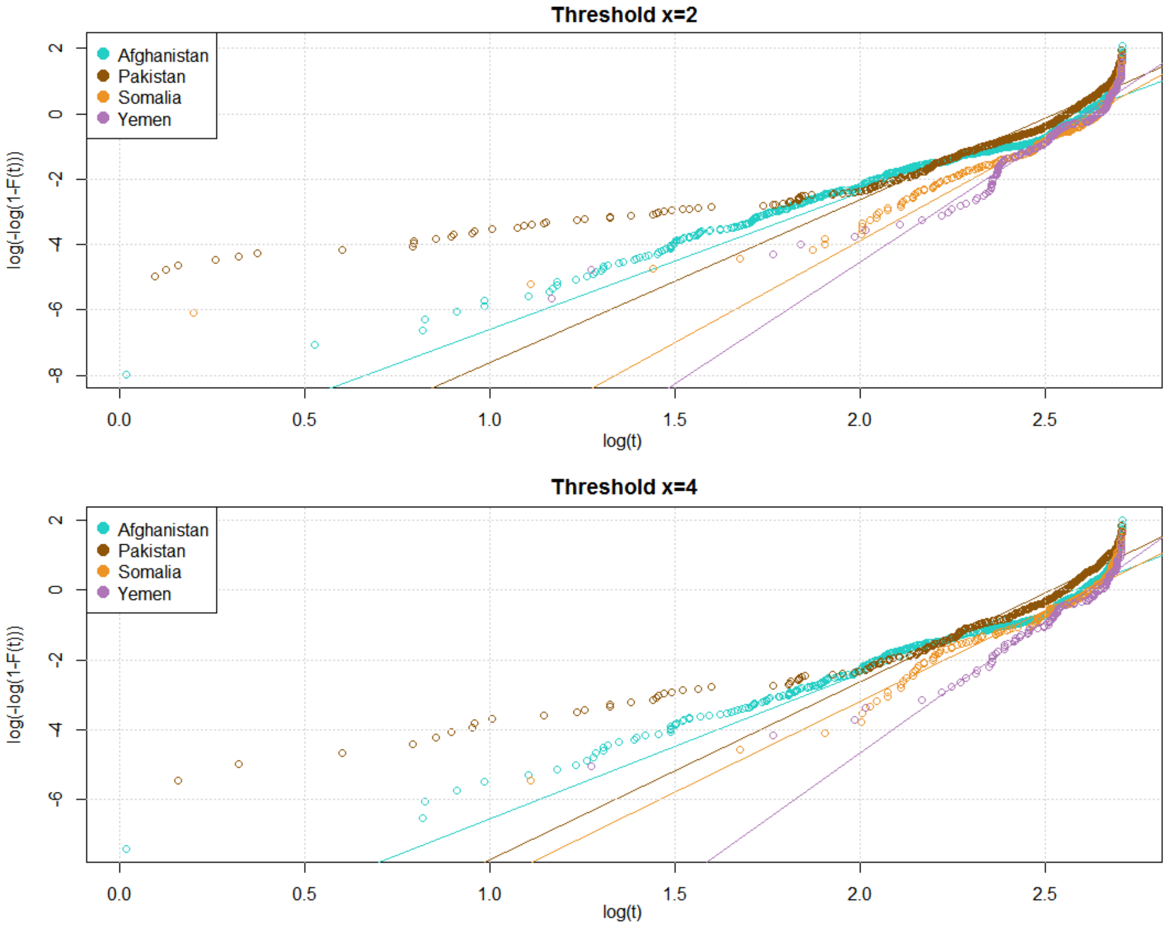

Fig. 4.14 Weibull plots including four countries that require similar thresholds. The country data seems consistent with a Weibull baseline model.

The Weibull baseline assumption seems decent for all four countries, even though Afghanistan does show slight divergence from a straight line in the tails. For Iraq, the divergence from the straight line is more significant, see figure 4.15. Model results in section 6.4 will show that the attacks indeed show quite some divergence from a Weibull model. Note that Nigeria's attack data is consistent with the Weibull model. Note that a rough graphical evaluation suggests that the piecewise proportionality assumptions is reasonble, except for Nigeria and Iraq.

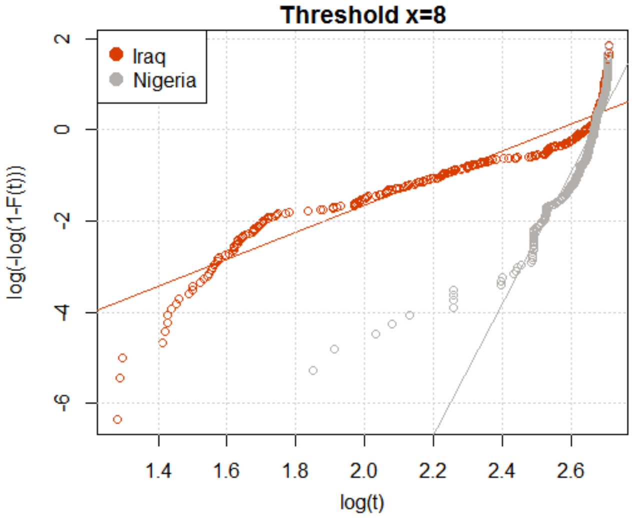

Fig. 4.15 Weibull plot for Iraq and Nigeria with a threshold x=8.

Section 3.4 suggested to choose aggregated targets by country as unit of analysis. The Weibull plots distinguishing these targets are shown in figures 4.16 and 4.17. The most visible illustration of conflict is target Military & Police. Nevertheless, Private Citizens & Property, State/Government/Public Goods and Public actors seem to satisfy the assumption for Afghanistan and Pakistan. For Iraq, Somalia and Yemen the distinction between targets does not seem reasonable. A rough graphical evaluation suggests that piecewise proportionality between the targets hold, with expection of Military & Police. However, this study chooses to model indvidual countries without distinguishing in targets, leading to more accurate model results.

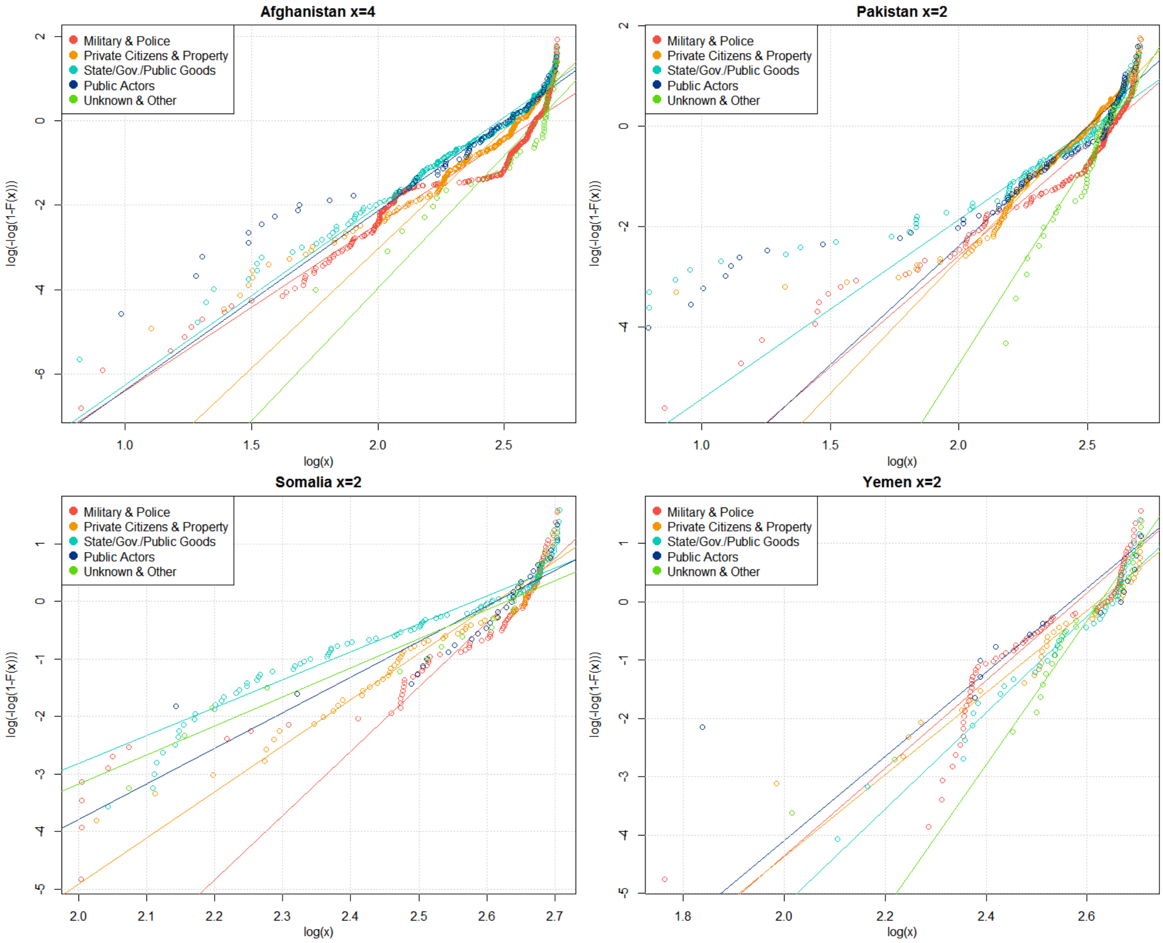

Fig. 4.16 Weibull plots individual countries for proposed aggregated targets. Biggest conflict with the Weibull assumption is target Military & Police. Private Citizens & Property, State/Government/Public Goods and Public actors seem to satisfy the assumption for Afghanistan and Pakistan.

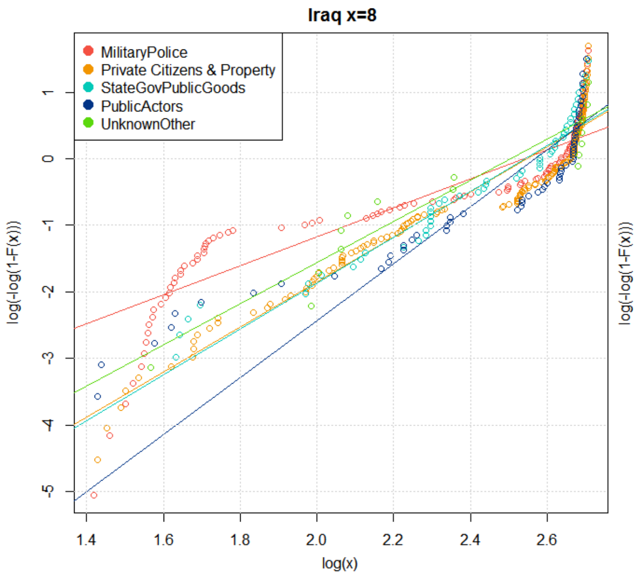

Fig. 4.17 Weibull plot Iraq for proposed aggregated targets. A Weibull model for the aggregated targets does not seem reasonable.

4.3.4 Strengths and limitations model

The PPWBPC model has both strengths and limitations that need to be clear in order to evaluate the performance of this particular model in a later stage.

A limitation to the use of a Poisson process, is that its assumptions conflict with the nature of the modelled phenomenon. Backed by the fact that the data is clustered; after one attack occurred, there is an increased probability of a subsequent attack occuring within a short period of time. To overcome this dependence of attacks, the data needed to be filtered to fit the rather restrictive Poisson assumptions. Especially for Iraq, with a threshold of x=8, this has impact on the practical applicability of the results in policy making.

On the other hand, using a Poisson process comes with the advantage that the model is easy to fit and the interpretation of parameters is straightforward. In addition, the confidence intervals are easily calculated from the fitted intensities. Regression coefficients can be estimated with both parametric and non-parametric approaches. The strong increasing trends in terrorist activity were shown to be reasonably consistent with the Weibull model, with some uncertainty for Iraq. This study therefore prefers the efficiency of the parametric model over the flexibility of the non-parametric approach [93].

The major drawback is that the Weibull distribution forces a monotonically increasing (or decreasing) baseline intensity. This means that the model is unable to handle a change in trend and will only produce reasonable results, if the increasing trend in terrorist activity persists. From a sociological point of view, the monotonic increasing baseline is unrealistic, because it constrains forecasts to eventually annihilate entire populations. Forecasting is therefore limited to a few years. Since the past is used as a predictor of the future, the model will naturally suffer from historical bias.

Generalising to piecewise constant covariates, utilising the annual country statistics, comes with the advantage that the accuracy of the model greatly increases. Still, the way they are implemented now, is that the covariate jumps in the beginning of a new year, while in reality, there is a smooth transition. Another limitation of our model is that the regression coefficients stay constant over time, forcing the effects of covariates to be the same at all points in time [93]. This might not be a reasonable assumption when modelling terrorist attacks. How people cope with demographic, economic, environmental and sociopolitical circumstances is dependent on a variety of factors and unlikely to stay constant over time. However, allowing the regression coefficients to vary over time, comes with the cost of losing analytical simplicity and interpretability of the results.

Finally, the amount of covariates that have the potential to improve the model's performance is limited. First of all, chapter 5 shows that annual country data can be scarce and their accuracy uncertain. Secondly, a part of the covariates will turn out to be strongly correlated and cannot be implemented simultaneously. Thirdly, modelling countries individually means that piecewise-constant covariates whose values constitute to a linear or exponential curve, cannot improve the model fit. The baseline model can only be improved by covariates that show dynamics corresponding to the behavior of terrorist activity. The fact that country statistics can be scarce and inaccurate is a big limitation to the general method of explaining terrorism by country statistics.

Chapter 5

Annual country statistics

This chapter starts with a general overview of the annual country statistics that represent underlying causes of terrorism expected to affect the rate at which attacks occur. The data is elaborated in section 5.2. Correlations and additional correspondences between the data and the number of attacks over time are discussed in section 5.2. The findings are compared to theory from social science. Finally, section 5.4 determines country-specific sets of covariates to be included in the model.

5.1 Overview

Since the countries are modelled individually, factors can only improve the model if they vary over time. This substantially limits the amount of potential covariates. Table 5.1 provides a general overview of the annual country statistics that were acquired to represent underlying causes of terrorism. Since the data can be scarce, the table includes the availability. Missing data for Somalia is substituted by the average of neighboring countries Djibouti and Ethiopia. The validity of this substitution should be examined by comparing the economic, political and sociological situations in the countries, this is however not pursued.

Elaborating on table 5.1, there was first of all no annual religious data found on country-level. The same holds for income inequality; the commonly used GINI index is too sparse for the countries of interest.

Poverty is tried to be captured by GDP per capita, or alternatively by the UN's Human Development Index. However, increasing GDP per capita does not necessarily imply reduction of poverty. A positive correlation between GDP per capita and terrorist activity might rather reflect increasing income inequality.

The Human Development Index measures a broader perspective of development, but is only available for the years 2000 and 2010-2014. Moreover, data for Somalia is missing. The country data representing the economic factors seems to have little potential in improving the model's results.

Demographic, environmental and sociopolitical data is fortunately abundant. Annual country statistics on unemployment, population growth and urbanisation are included. Drought is captured by annual precipitation data. Temperature is included in order to try to capture a more general influence of climate change.

Corruption is represented by one of the World Bank's governance indicators, named control of corruption. A second governance indicator, Political Stability and Absence of Violence/Terrorism, is included as extra statistic to examine whether there is a relationship between the stability of governments and terrorist activity.

Section	Root Cause	Represented by	Availability
Religion & Globalisation	Religion	-	-
	Globalisation	KOF Globalisation Index	Somalia missing, other countries 2000-2014
Economic factors	Poverty	GDP per capita	Somalia missing, Afghanistan missing 2000-2001
		Human Development Index	Somalia missing, other countries 2000, 2010-2014
	Income inequality	(GDP per capita)	-
Environmental factors	Drought	Precipitation	All countries; 2000-2014
	Climate change	Temperature	All countries; 2000-2014
Demographic factors	Unemployment	Unemployment rate	All countries; 2000-2014
	Rapid population growth	Population growth	All countries; 2000-2014
	Urbanisation	Urbanisation rate	All countries; 2000-2014
Sociopolitical factors	Political Rights	Political rights index	All countries; 2000-2014
	Civil Liberties	Civil liberties index	All countries; 2000-2014
	Corruption	Control of corruption	All countries; 2000, 2002-2014
	-	Political Stability & Abs...	All countries; 2000, 2002-2014

Table 5.1 Representation causes of terrorism by annual country data.

5.2 Covariate data

This section introduces the data of the five country statistics for which section 5.4 shows, they have the potential to improve the model fit. Numerous other factors were considered, they are elaborated in appendix B. Note that although the civil rights index is not included as covariate, the index is introduced in conjuction with political rights, because the indices are rated according to the same metric.

5.2.1 KOF Globalisation Index

The KOF Globalisation Index is a measure for the economic, social and political dimensions of the globalisation of countries [129]. Economic globalisation captures long distance flows of goods, capital and services [147]. The social dimension focuses on the spread of people information and culture. Political globalisation measures the diffusion of government policies internationally.

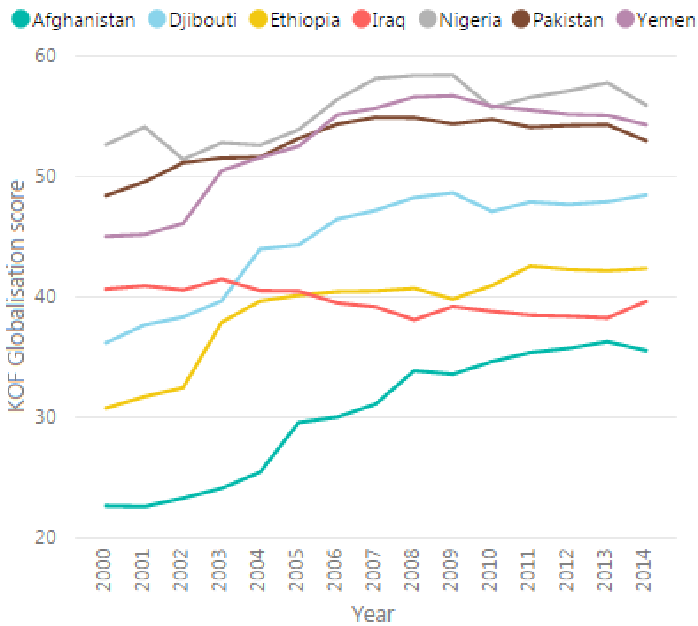

Fig. 5.1 KOF Globalisation Index

Data on Somalia is missing and is substituted by the average of Djibouti and Ethiopia (see section 5.1). Except for Iraq, that faces a general decrease, globalisation seems to stagnate around the year 2009, see figure 5.1. The countries are expected to show positive correlation between globalisation and terrorist activity.

5.2.2 Political rights & Civil liberties

NGO Freedom House publishes an annual report that assesses the civil liberties and political rights in countries worldwide. The statistics are available from 1972 through 2017. Their performance is evaluated on the individual level, rather than goverment performance, and can be affected by both state and nonstate actors. Both indices are assigned an integer rating between 1 and 7, with 1 representing the greatest- and 7 the smallest degree of freedom [21].

For political rights, a score of 1 represents a wide range of political rights, that include free and fair elections, whereas a 7 is appointed to countries that exercise few or no political rights because of government oppression, occasionally in combination with civil war. Similarly, a civil liberties rating of 1 exhibits a wide range of civil liberties that include freedom of expression, association, education and religion. The worst score 7 is given to countries where governments or nonstate actors "...allow virtually no freedom of expression or association, do not protect the rights of detainees and prisoners, and often control most economic activity".

Civil rights show high correlation with political rights. Freedom house explains this by pointing out that politically oppressive states typically do not allow a well-developed civil society. Conversely, it is hard to maintain political freedoms in the absence of civil liberties. The difference between the two indices is therefore rarely larger than two.

Section 2.3.3.5 cited research that found a nonlinear relationship between political rights and terrorist activity; countries with intermediate political freedom experience higher terrorist activity than both highly authoritarian regimes and countries with high political freedom such as legitimate democracies [30]. The direction of correlation between the indices and the number of attacks is therefore country-specific.

Fig. 5.2 Political rights and Civil liberties indices

5.2.3 Political Stability and Absence of Violence/Terrorism

Political Stability and Absence of Violence/Terrorism is one of six Worldwide Governance Indicators by the World Bank. It represents the likelihood of political instability and/or politically motivated violence, including terrorism [23]. Although terrorism is included, the emphasis lies on the political situation in a country, which includes, but is not limited to, social unrest, violent demonstrations, goverment stability and ethnic tensions.

The World Bank uses estimates based on a weighted average of the sources available for each country [10]. The weights are proportional to the reliability of each source. This results in estimates of which the expected value equals zero and standard deviation equals one, implying most estimates score between -2.5 and 2.5, where a higher score represents a better outcome.

Figure 5.3 shows the index over time, where the missing data for 2001 is linearly interpolated. A negative correlation between the index and terrorist activity is expected.

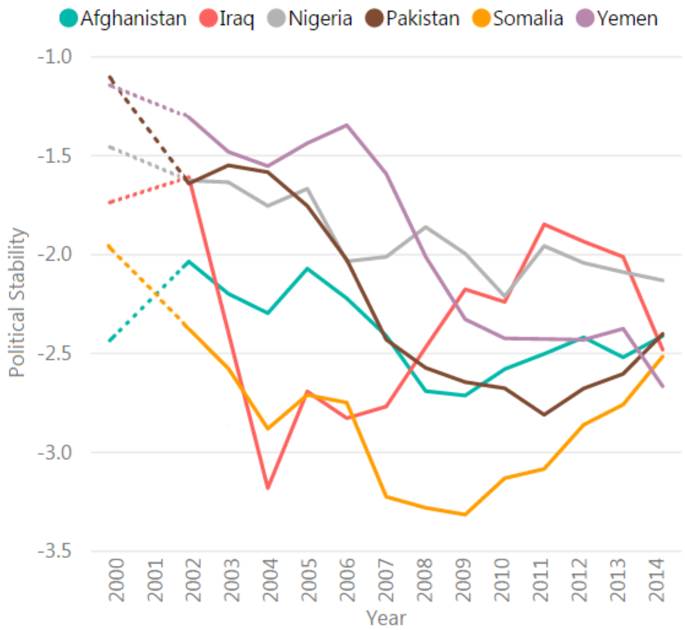

Fig. 5.3 Political Stability and Absence of Violence/Terrorism

5.2.4 Control of corruption

A second of the World Bank's governance indicators is control of corruption, that "...captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as capture of the state by elites and private interests" [46]. The index is rated in the same way as the political stability index of section 5.2.3. Data for 2001 is again missing and linearly interpolated, see figure 5.4. Increasing control of corruption is expected to decrease terrorist activity.

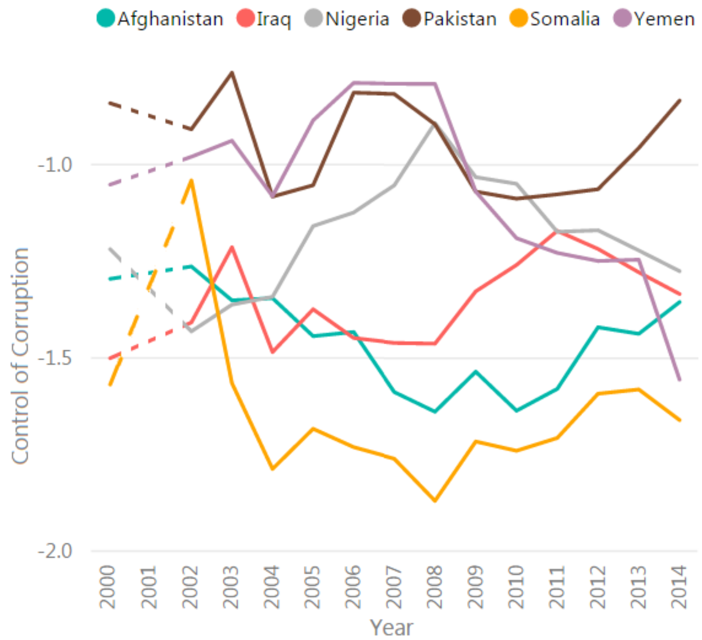

Fig. 5.4 Control of corruption

5.2.5 Unemployment Rate

The International Labour Organisation obtains annual unemployment rates by estimating the share of labor force without work, but available for and seeking employment [25]. The data is harmonised to account for differences in data sources, methodology and other factors.

The data exhibits several patterns that are in line with the analysis of sections 2.1.1 and 2.3.3.4. Iraq shows the upsurge in unemployment claimed to be due to the regime change enforced by the United Stated. Similarly, the invasion of Afghanistan in 2001 is likely to be the cause of the unemployment shift in 2003. The countries are expected to exhibit a positive relationship between unemployment and terrorist activity.

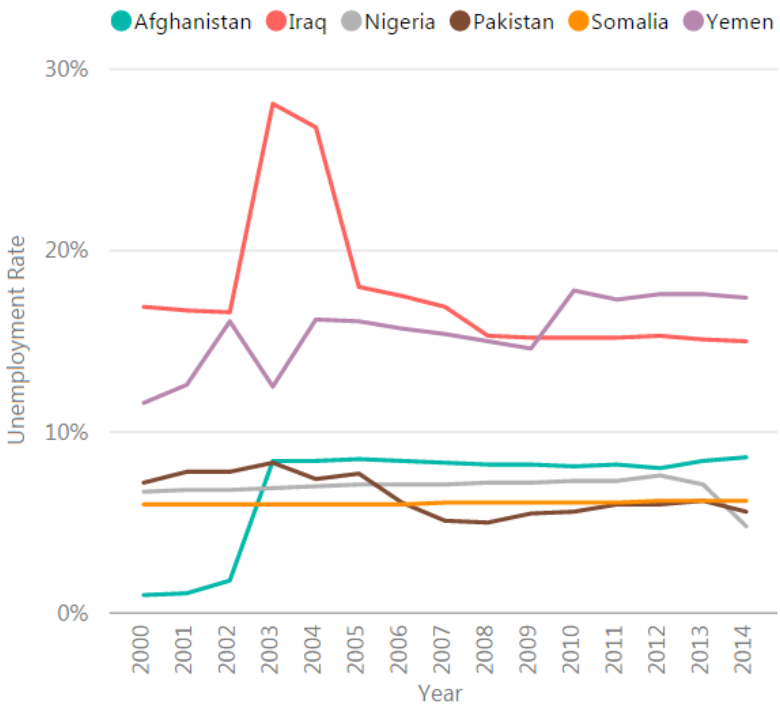

Fig. 5.5 Annual unemployment rate

5.3 Correlation between terrorist attacks and country statistics

The analysis of the bivariate correlations between the annual country statistics and terrorist activity is based on figures C.1 - C.6. Each figure covers one country and contains the correlation between the number of attacks (no threshold applied) and the country data, as well as the intercorrelation that helps to select covariates in section 5.4. More specifically, in the upper triangular, Spearman's rank correlation coefficient, possibly tie corrected, and the corresponding statistic significance are given. For a general overview, different levels of correlation are represented by different colors (dark blue strong positive relationship, dark purple strong negative relationship) and the significance is indicated by asterisks. The lower triangular contains the corresponding scatter plots, including linear model fits and 95% confidence intervals. Note that for Pakistan and Somalia, the correlations regarding civil liberties are absent because the index is constant over time.

Table 5.2 compare the direction of each significant relationship with the direction proposed by literature, where green (red) cells reflect significant correlations that (do not) correspond to theory. There are just five directions that conflict with social science theory. In the analysis that follows, scatterplots are examined and the timeseries of number of attacks (figure 4.8) and annual country statistics (section 5.2) are compared for striking correspondences in the data. The interesting findings are discussed. Although examining scatterplots and correlations can be useful, it does not take into account that there might be a natural trend in occurrence of attacks rooted in human psychology and sociology. The deviations from this trend could then be explained by the underlying factors alleged to impact terrorism. This is what the PPWBPC model is able to do.

As a final note for the reader unacquainted with statistics, it is important not to confuse correlation with causation. Correlation does not take the nature of the relationship between two variables into account. When the relationship is statistically significant, correlation can however be suggestive.

Covariate	Representing root cause	Expected relation	Afghanistan	Iraq	Nigeria	Pakistan	Somalia	Yemen
Civil Liberties index	Civil Liberties	See [1]	-0.18	-0.69**	0.046	–	–	0.5
Control of Corruption	Corruption	-	-0.55*	0.4	0.28	-0.29	-0.26	-0.61*
GDP	Income inequality [2]	+	0.96***	0.82***	0.88***	0.94***	0.96***	0.63*
KOF Globalisation	Globalisation	+	0.94***	-0.62*	0.62*	0.52 *	0.94***	0.61*
Political Rights	Political Rights	See [1]	-0.31	-0.39	0.019	-0.87***	0.87***	0.32
Political Stability & Abs..	–[3]	-	-0.54*	-0.32	-0.83***	-0.83 ***	-0.45	-0.93***
Population	Rapid Population Growth	+	0.97***	0.86***	0.88 ***	0.94***	0.96***	0.94***
Precipitation	Draught	-	0.55*	-0.39	-0.15	0.19	0.14	-0.4
Temperature	Climate Change	+	-0.2	-0.15	0.37	-0.13	0.17	0.23
Unemployment	Unemployment	+	0.41	-0.55*	0.41	-0.64 *	0.94 ***	0.61 *
Urbanisation	Urbanisation	+	0.97***	0.86 ***	0.88 ***	0.94 ***	0.96 ***	0.94 ***

[1] Empirical research by Abadie found that countries with intermediate political freedom experience higher terrorist activity than both high authoritarian regimes and countries with high political freedom such as legitimate democracies [30]. Similar reasoning holds for civil liberties. Therefore, countries that improve from severe oppression are expected to be positively correlated with terrorism (i.e. more terrorist attacks) and countries that see their rights detioriate from mediocre to no freedom are expected to show negative correlation (less terrorist attacks)

[2] GDP growth could possibly capture poverty instead, for which a negative relation with terrorism is expected. [3] Section 2.3 did not cover the impact of political stability on terrorism, but a negative relationship is expected (i.e. more stability decreases terrorist activity).

Table 5.2 Comparison of theory on causes of terrorism with correlations found between country statistics and the number of terrorist attacks. Green (red) cells reflect significant correlation of which the direction (does not) corresponds to theory from social science.

Civil Liberties & Political Rights

Recall empirical research by Abadie found a nonlinear relationship between political rights and terrorist activity; countries with intermediate political freedom experience higher terrorist activity than both highly authoritarian regimes and countries with high political freedom such as legitimate democracies [30]. Therefore, countries that improve from severe oppression are expected to be positively correlated with terrorism (i.e. more terrorist attacks) and countries that see their rights detioriate from mediocre to no freedom are expected to show negative correlation (less terrorist attacks). Similar reasoning holds for Civil Liberties.

Considering Abadie's findings and comparing figures 4.8 and figure 5.2, the first upsurge of terrorism in Iraq, starting in 2004 and lasting only until 2005, could be interpreted as a consequence of the improvement of civil liberties from extreme expression. Freedom House explains however, that as the United States mobilised military resources for a possible invasion of Iraq, "...Saddam Hussein took some steps to improve human rights conditions in what many observers regard as the most oppressive state in the world" [80]. The positive shift in civil liberties is therefore rather a consequence of the U.S. American invasion. The theorem that the regime change enforced by this invasion created the preconditions for terrorism to arise, remains convincing. The corresponding scatterplot in figure C.2 does not show a clear relationship between civil liberties and the number of attacks.

Correlation with civil liberties is absent for Somalia and Pakistan because the index is constant over time. The scatterplots of the remaining countries show no sign of a meaningful relationship between civil liberties and terrorism activity.

The scatterplot of Pakistan's terrorist attacks versus political rights rating is more convincing, see figure C.4. Clearly, the change from severe oppression to intermediate politcal freedom is associated with higher terrorist activity. The end of military rule and parliamentary elections in 2008 is the reason of improved political rights, which is the exact year of real breakthrough of terrorism, see figure 4.8. The lack of precision of the index (integer value between one and seven), does unfortunately not allow to deduce a more comprehensive conclusion and moreover substantially limits the predictive potential as covariate. Yet, Abadie's findings seems to be supported for Pakistan.

However, in Somalia terrorism breaks out after its Political Rights index deteriorates to the worst level possible, contradicting Adabie's nonlinear relationship. In addition, extremer forms of political oppression in Yemen are met with higher terrorist activity. For the remaining countries, no meaningful relationship was found.

Concluding, there seems to be a connection between political rights and terrorist activity in Pakistan, in line with Adabie's finding discussed in section 2.3.3.5. On the contrary, higher forms of political oppression in Somalia and Yemen seem to be associated with higher terrorist activity. For the remaining countries, there is no sign of a meaningful relationship. Civil Liberties does not seem to affect terrorism directly. The lack of precision of the index (integer value between one and seven), hinders more meaningful conclusions and substantially limits the predictive potential as covariate.

Control of Corruption

The scatterplot found in figure C.6, shows that as Yemen loses grip on corruption, terrorist activity increases. A closer look at figures 5.4 and 4.8, shows that 2009 is the start of a strong decreasing trend of control of corruption, after which terrorism in Yemen seems to emerge. As both measures generally deteriorate, 2013 sees a decrease in terrorist activity and a slight increase in control of corruption, while 2014 sees its worst terrorism and corruption yet. This is in line with the hypothesis that increasing corruption leads to more terrorism. The fact that the Spearman correlation is not convincing is caused by the early years, where Yemen gains more control of corruption, but terrorism does not decay.

The scatterplot of the number of attacks in Afghanistan versus control of corruption, see figure C.1, does not suggest that corruption is able to explain terrorism. However, control of corruption will turn out to capture deviations from the baseline trend very well.

Gross Domestic Product

The significant positive correlations concerning GDP suggest the covariate rather captures income inequality, than (reduction of) poverty. The countries, with exception of Iraq, show however no sign of significant change in trends over more recent years when terrorism truly emerges. In fact, Afghanistan's growth in GDP stagnates in the more recent years and Yemen experiences a decrease in 2011, causing the questionable correlation as displayed in the scatterplot found in figure C.6. Besides, the smoothness of the data suggests the World Bank estimates do not provide the accuracy required to make meaningful statements.

On the contrary, Iraq's data exhibits fluctuations that might provide more insight. First, there is the drastic decline in 2003 due to the aforementioned U.S. invasion and subsequent Iraq war. One year later, as the GDP recovers, terrorism breaks out. Despite the uprising trend, there are unfortunately no striking correspondences that could hint to causality. Concluding, while GDP is (strongly) positively correlated with terrorist activity for all six countries, indications of the nature of these relationships cannot be deduced.

Globalisation

Figure 5.1 shows five countries facing strong increasing globalisation in distant years, while stagnating/decreasing in the more recent years. Comparing this to the exponential increases of terrorist attacks, this makes sense from a socio-economic perspective. As globalisation rises in the countries of interest, phenomena such as income inequality and urbanisation increase. In addition, there is the effect that globalisation can cause a clash between cultures. As a consequence, hypothetically, terrorism emerges/intensifies. At the same time, increases in terrorism have a negative influence on characteristics from globalisation such as foreign investments, trade and tourism, explaining the stagnation over more recent years. The scatterplots follows this line of reasoning and give reason to believe that there is a mutual causality between terrorism and globalisation. The negative correlation exhibited by Iraq is remarkable and conflicting with this statement, foreign intervetion and war might again be the explanation.

Political Stability and Absence of Violence/Terrorism

Although the emphasis of the Political Stability and Absence of Violence/Terrorism indicator lies on political stability, terrorism is included and the negative correlations the countries exhibit are therefore not surprising. What the scatterplot of Yemen fails to reveal, is that when the political stability truly starts to decline in 2008, Yemen experience its upsurge in terrorism, see figures 4.8 and 5.3. In addition, the slight improvement and subsequent severe deterioration of the political situation from 2013 to 2014 is reflected by the behavior of terrorist activity.

Pakistan and Nigeria show similar resemblances, although the fluctuations of terrorist activity and political stability in more recent years do not seem to be related. For Afghanistan and Somalia there does seem to be a negative influence of political stability on terrorism, while once again for Iraq conclusion cannot be drawn.

Population

Due to fact that population estimates for the countries of interest are rough approximations, all six countries experience smooth monotonic growth, causing high positive correlations. It is not possible to draw any conclusions other than that for our data, the population rises when terrorism increases and vice versa.

Precipitation

Against expectation, Nigeria, Somalia and Yemen do not show a significant, negative correlation with precipitation. Comparing the data shows no correspondences between the number of attacks and rainfall. The significant positive correlation for Iraq is not meaningful either. Literature stated environmental factors such as drought, act rather as a threat multiplier than as a direct cause, which seems to be confirmed.

Climate Change
Trying to capture climate change with temperature is of course naive, considering the fact that climate is (often) defined as the weather averaged over 30 years. Indeed, no significant relationship is found. Over short periods of time, temperature does therefore not seem to affect terrorism directly.

Unemployment
Afghanistan sees its terrorism emerge in 2003, the year unemployment strongly deteriorates. Whether there is causality, or both are rather a consequence of the invasion in Afghanistan, is unclear. As stated, increased security concerns of companies after the invasion in Afghanistan in 2001 did lead to a major shift in unemployment. Yemen's true breakthrough in terrorism in 2010 sees a rise in unemployment that remains at a high level. For both Afghanistan and Yemen, higher levels of unemployment do correspond with higher levels of terrorism.

Policies instituted by the U.S. in 2003 pushed Iraq's unemployment levels near 30%, see figure 5.5. Literature argued this was an important precondition for ISI to emerge. Indeed, terrorism sees an upsurge in 2004 and 2005, see figure 4.8. Nevertheless, when unemployment seems to stabilise to levels lower than before the invasion (though still disturbingly high), Iraq experiences a resurgence of terrorism reaching a dramatic climax in 2014. This explains the negative correlation and seems to contradict the hypothesis of unemployment as primary cause of terrorism. Similarly, Pakistan's negative correlation conflicts with this hypothesis. Unemployment rather seems to be the spark that ignites the fire.

Unemployment in Somalia stays near constant caused by a lack of data, which explains the strong positive, but meaningless, correlation.

Urbanisation
Urbanisation makes use of population estimates and therefore approximations will be at least as rough. The result is the same, all six countries experience smooth monotonic growth, causing high positive correlations, to which no further meaning can be assigned.

5.4 Selecting covariates

For the five countries that are consistent with the model assumptions, country statistics need to be selected to implement as covariates. The goal is to select those factors that correlate with terrorism, but do not introduce multicollinearity. This is done by evaluating scatterplots and correlations. However, bivariate correlations are not sufficient to guarantee absence of multicollinearity. The analysis was done assuming a proportional intensity model, for which standards for more formal quantification of multicollinearity (such as variance inflation factors or eigenvalue analysis), are not available. Nevertheless, it will turn out that the amount of country statistics that have the potential to improve the model's results is limited and therefore evaluation of bivariate correlations is deemed sufficient.

There are three reasons why the number of potential covariates is limited. First of all, as mentioned in section 5.1, since countries are modelled individually, factors can only improve the model if they consist of annual data. Secondly, piecewise constant covariates whose values constitute to a linear or exponential curve cannot improve the model fit. The model's baseline estimation can thus only be improved by covariates that show dynamics corresponding to the number of attacks over time. Therefore, the population and urbanisation statistics, and to slightly less extend GDP per capita, are useless for our model. Finally, there is the general fact that annual country data is scarce and their accuracy uncertain.

Applying thresholds, removing attacks that do not generate a country-specific number of attacks, does not significantly change the correlations of covariates with terrorism, with exception of Iraq, see figure 5.6. Part of the findings from section 5.2 can thus be used to exclude statistics that show no correspondences with dynamics in terrorist activity.

In this light, both environmental factors (precipication and temperature) can be omitted from further analysis. Futhermore, the civil liberties index did not reflect dynamics in terrorist activity and is not included. Political rights might be able to capture change points for Pakistan, Somalia and Yemen. The potential of the political rights index for our modelling purposes was nevertheless argued to be questionable, because the rating of the index severely lacks precision.

This leaves five potential covariates: control of corruption, globalisation, political rights, political stability and unemployment (analysing the data for Iraq with a threshold of x=8, results in the same five covariates). This number is further reduced by evaluating the statistics on country level. Finally, to prevent multicollinearity, correlation among covariates is compared to the rule of thumb that allows a maximum correlation in absolute value of 0.6. This results in country-specific covariates sets that will be implemented in the model. For sake of simplicity, the covariates are occasionally abbreviated by C (control of corruption), G (globalisation), PS (Political Stability &Absense of Violence/Terrorism), PR (political rights) and U (unemployment).

Afghanistan Although the rise in unemployment could in 2003 could have sparked the emergence of terrorism in the same year and unemployment remains high while terrorism thrives, see figures 5.5 and 4.8. The unemployment rate will therefore not be able to improve the model. Similarly, Afghanistan's shifts in political rights do not correspond to dynamics in the number of attacks and will thus not capture deviations from the baseline model. This leaves control of corruption, globalisation and political stability, with the intercorrelations C-G: -0.6*, C-PS: 0.61*, G-PS: -0.59*. Comparing this to the rule of thumb, suggests not to implement these covariates in the model simultaneously.

Iraq

For Iraq, none of the five factors can easily be excluded on basis of scatterplots and comparison between timeseries of the country statistics and terrorist activity. The five covariates exhibit the intercorrelations of table 5.3. This results in three covariate sets that always include corruption and political stability: C-PS-G, C-PS-PR and C-PS-U.

	C	G	PR	PS	U
C	1	-0.26	-0.53	0.24	-0.39
G	-0.26	1	0.79	0.11	0.61
PR	-0.53	0.79	1	0.076	0.69
PS	0.24	0.11	0.076	1	-0.36
U	-0.39	0.61	0.69	-0.36	1

Table 5.3 Intercorrelation covariates Iraq

Pakistan

The dynamics of unemployment in Pakistan seen in figure 5.5, do not correspond with changes in terrorist activity, see figure 4.8. Although the negative correlation between unemployment and terrorist activity in Pakistan is counter-intuitive, perhaps even nonsensical, the covariate is not omitted. This leaves corruption (C), globalisation (G), political rights (PR) and political stability (PS), with intercorrelations as in table 5.4. It is clear that political rights and political stability, as well as globalisation and political stability should not be implemented in the model simultaneously. A correlation of 0.52 between control of corruption and politcal stability is high, but acceptable. This results in the covariates sets C-G-PR and C-PS.

	C	G	PR	PS
C	1	-0.16	0.46	0.52
G	-0.16	1	-0.49	-0.72
PR	0.46	-0.49	1	0.84
PS	0.52	-0.72	0.84	1

Table 5.4 Intercorrelation covariates Pakistan

Somalia Unemployment rates for Somalia consist of rough estimates that approximate linearity, making them useless as covariate. The country experiences one shift in political rights in 2006, which is not reflected in the number of attacks, see figures 5.2 and 4.8. In the absense of data, the KOF globalisation index was substituted by the average of Djibouti and Ethiopia. Although questionable, the covariate is not omitted. This leaves the following covariates with intercorrelations C-G:-0.37, C-PS: 0.8 and G-PS: -0.54. This results in two covariate sets C-G and G-PS.

Yemen Although high terrorist activity in Yemen seems to be related to high political oppression, the three years the political rights situation in Yemen changes (2001, 2002 and 2009) do not correspond to changes in terrorist activity, making the index useless as covariate. This leaves four covariates, for which the correlations are found in table 5.5. The several rather high correlations result in three covariate sets: C-G, G-U and PS.

	C	G	PS	U
C	1	-0.029	0.67	-0.63
G	-0.029	1	-0.65	0.32
PS	0.67	-0.65	1	-0.71
U	-0.63	0.32	-0.71	1

Table 5.5 Intercorrelation covariates Yemen

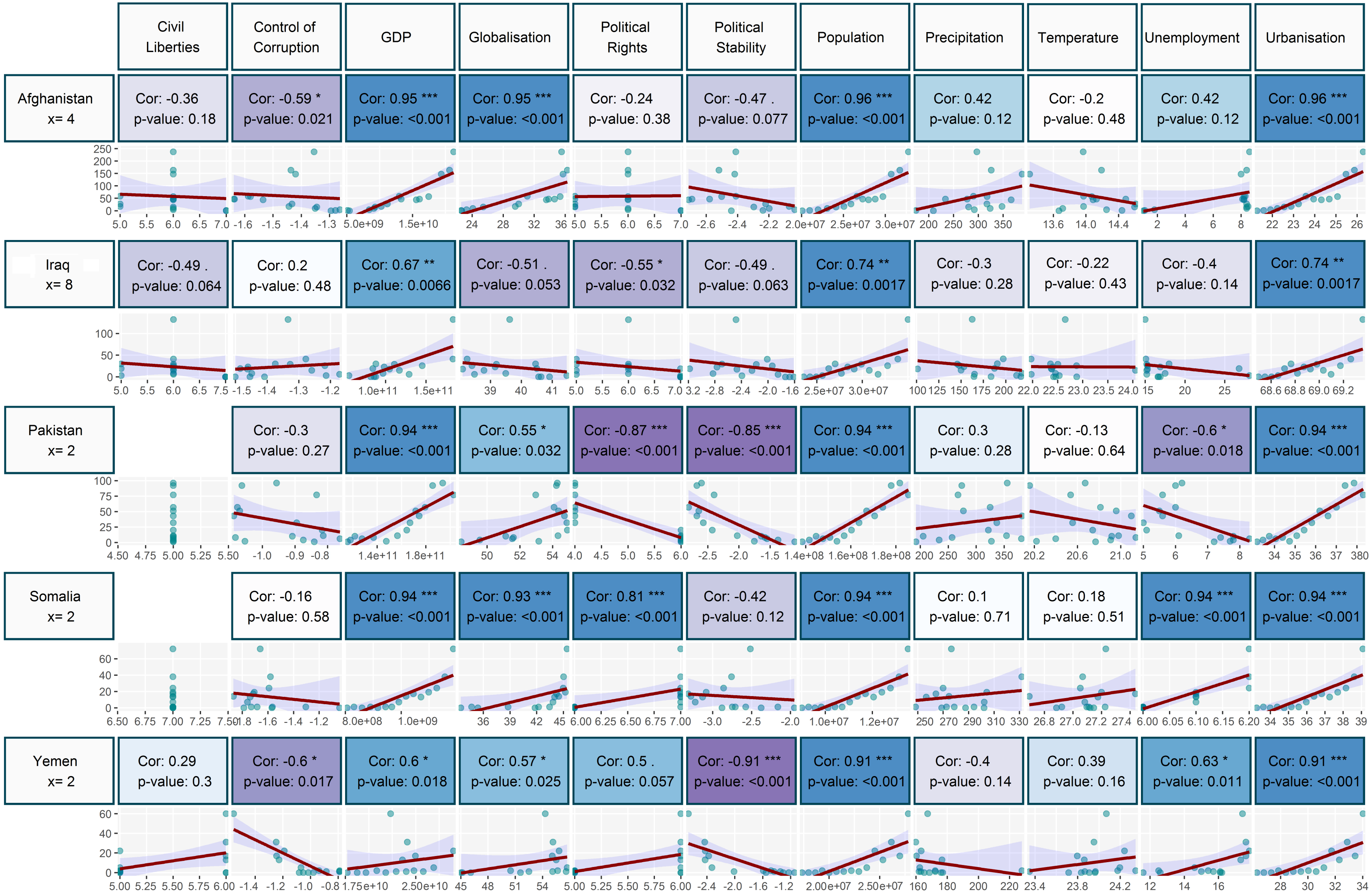

Fig. 5.6 Correlation between covariates and the number of attacks, after applying country-specific thresholds

Chapter 6

Results

6.1 General procedure

This chapter discusses the results of the PPWBPC model. The results of the countries are presented in the order of the number of covariates in the parsimonious model. Without significant covariates, Somalia is discussed first. The annual data is adjusted for outliers and fitted by a Weibull trend, which produces surprisingly good results. Secondly, with one covariate, Afghanistan is evaluated. Control of corruption captures the dynamics in terrorist activity remarkably well, both in-sample and out-of-sample. For Iraq, the two covariates globalisation and control of corruption reasonable capture the strong deviations from the baseline trend. However, the out-of-sample model performance will turn out disappointing. The results for Pakistan and Yemen are less interesting and therefore included in appendix D.

It is important to note that in the analysis, only terrorist attacks are included for which the perpetrator is known, all three criteria of definition 2.1 are met and at least a country specific threshold of x deaths are generated. Details on this data selection can be found in sections 3.4 and 4.3.3.

The bivariate correlations of section 5.4 resulted in country specific sets of covariates. The next step is to perform model selection. The Akaike information criterion (AIC) is an estimator of the predictive potential of statistical models that can be used for model selection, where the preferred model is the one with minimum AIC value. This leads to the following approach applied for each country

1. Fit the model with all covariates
2. Select the covariates on the basis of significance
3. Use the AIC to filter out covariates to find the parsimonious model

After finding the parsimonious model, the in-sample fit is assessed first, by comparing the fitted values to both the naive 95% confidence interval (CI) and the bias-corrected and accelerated (BCa) 95% bootstrap CI. The BCa bootstrap CI corrects the standard bootstrap CI for possible bias and skewness in the distribution of the bootstrap estimates. A residual analysis is performed to check for conflicts with the model's assumptions.

At time of writing, the *new* 06/17 version of the GTD was released, which extends the 06/15 version to include attacks for 2015 and 2016. The new version faced corrections, where incidents assessed as terrorist attacks before, are no longer included in the database and vice versa. As table 6.1 shows, differences in annual attacks are insignificant. The GTD 06/17 version is therefore used in subsequent analysis, since it provides two more years to either train or validate the data.

	Afghanistan x=4		Iraq x=8		Pakistan x=2		Somalia x=2		Yemen x=2	
	06/15	06/17	06/15	06/17	06/15	06/17	06/15	06/17	06/15	06/17
2000	0	0	0	0	4	3	0	0	0	0
2001	1	1	0	0	7	7	1	1	0	0
2002	4	4	0	0	11	11	0	0	0	0
2003	10	10	3	3	8	7	1	0	2	0
2004	17	16	23	23	9	9	1	1	0	0
2005	23	23	34	33	4	4	1	1	1	1
2006	41	41	8	7	14	11	3	2	1	1
2007	78	76	25	26	23	23	10	9	2	2
2008	60	60	22	22	48	46	14	14	2	2
2009	62	62	30	30	61	61	20	20	5	5
2010	60	59	17	17	62	62	19	19	32	32
2011	71	71	7	7	64	64	31	31	17	17
2012	206	204	30	32	129	129	37	38	41	41
2013	224	223	49	49	126	125	50	50	23	23
2014	307	307	148	149	115	117	121	121	76	78
2015	-	349	-	161	-	90	-	73	-	144
2016	-	309	-	216	-	65	-	102	-	90

Table 6.1 Comparison number of attacks of GTD version 06/15 and 06/17. Cells marked red indicate differences between the two versions.

Including attacks until 2016 requires extended covariate data. Table 6.2 shows that the KOF globalisation index is not available for 2016, which should be taken into account when training and/or validating the data.

Covariate	Available until
Control of corruption	2016
Globalisation	2015
Political Stability	2016
Political Rights	2017
Unemployment	2017

Table 6.2 Availability covariates

After assessing the in-sample fit, out-of-sample validation provides insight into the model's forecast performance. A rolling window approach is pursued to obtain a multiperiod indication of forecasting accuracy. Considering that data is available from 2000-2016, (when globalisation as covariate is included this reduces one year), the minimum training period is set at 10 years, circa 60% of the timeseries. The end of the training period, i.e. year 2009, is set as forecast origin T, from which H forecasts are made. The forecasting origin is then updated by $(T+1)$, adding one year to the training data. The model is refit and again H predictions are made. Note that $H \leq 2016 - T$, since attacks after 2016 are unknown and cannot be validated. The forecasts are then compared to the 95% prediction intervals and metrics of predictions errors are presented as measure of out-of-sample performance. The prediction error metrics used in the analysis are the mean absolute error (MAE) and the mean absolute percentage error (MAPE), both widely used and in no need of further explanation.

With an indication of prediction performance, forecasts can be made. Of course, the model suffers from historical bias, since the past, showing an increasing trend, is used as a predictor of the future. The model will only function when the trend continues. All forecasts are made under the assumption that the trend indeed does so.

The forecasts exceed the covariate availibility shown in table 6.2 and therefore three scenarios are created. The first is referred to as the *Constant scenario*, where the covariate path is simply kept constant to its last known value. The second scenario uses Holt-Winters double exponential smoothing (DES) to forecast covariate values, making use of the formulas

$$\hat{y}_{t+1|t} = l_t + hb_t \tag{6.1a}$$

$$l_t = ay_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{6.1b}$$

$$b_t = \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1} \tag{6.1c}$$

where l_t and b_t estimate the level and the trend of the timeseries at time t, respectively, $0 < \alpha < 1$ is a smoothing constant for the level and $0 < \gamma < 1$ for the slope. The smoothing constants determine how fast the weights of the series decay; values close to one reflect large weight on more recent observations and small smoothing constants assign bigger importance to observations in the further past [63].

As Girosi and King write, forecasting covariates can work well when the relationship between the covariate and the dependent variable remains constant in the future and if high-quality forecasts are available [70]. However, the uncertainty of the DES covariate forecast greatly increases the forecast variance. They propose a procedure that forecasts t periods ahead by using covariates that are lagged by t periods. From a sociological perspective, if terrorist organisations respond to changing circumstances, this could indeed have delay, although more than a year is not deemed realistic. The third *lagged* scenario therefore uses covariates of lag one, which allows a one year forecast. The covariates are fortunately available before 2000 and backcasting is not required, though some indices are not available annually and require interpolation.

Because of absence of covariates in the parsimonious model of Somalia and disappointing out-of-sample results for Iraq, forecasting the three scenarios is only pursued for Afghanistan. Finally, the general conclusion in chapter 7 discusses differences and similarities between the countries' performances and covariates.

6.2 Somalia

6.2.1 In-sample performance

Of the five countries considered, Somalia is by far the country with the most unreliable and scarce data. This might be the reason that there are no significant covariates in the parsimonious model of Somalia, resulting in a highly significant trend parameter δ and intercept as in table 6.3. The maximum log-likelihood is 1474.519 and the corresponding AIC equals -2943.04.

Parameter	Estimate	Standard error	z-value	p-value
δ	4.432	0.202	21.954	$< 10^{-16}$
Intercept	-6.378	0.574	-11.117	$< 10^{-16}$

Table 6.3 Parameter estimates Somalia, training data 2000-2016

The in-sample fit is shown in figure 6.2, along with the naive and 10000 sample BCa bootstrap 95% confidence intervals. The number of attacks follow the Weibull model remarkably well, with the exception of the large spike in 2014. From december 2013, there is a clear upsurge of terrorist activity in Somalia, see figure 6.1. This could be due to the devastating cyclone in November that year or to specific events in the ongoing Somalia war. There was no data found that can help the model to explain the outlier.

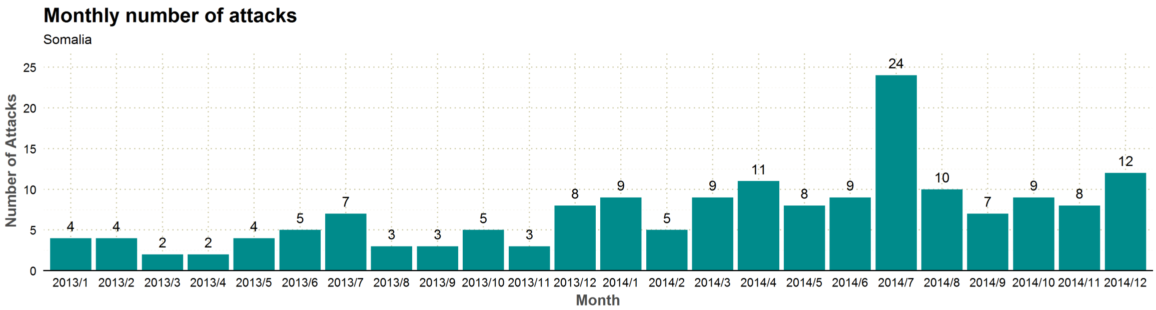

Fig. 6.1 Monthly number of attacks Somalia in 2013 and 2014.

As Chen and Liu write, presence of outliers forecast accuracy can reduce forecast accuracy due to "a carry-over effect of the outlier on the point forecast and a bias in the estimates of model parameters" [53]. In addition, Hillmer [133] and Ledolter [88] found not only that forecast intervals are sensitive to additive outliers, but note that point forecasts are affected when the outlier is positioned near the forecast origin. Chen and Liu propose a procedure to detect outliers and to adjust for their effect, based on the approach by Chang et al. [82], that examines the maximum value of the standardised statistics of the outlier effects.

Based on simulation results, they recommend a critical value between 2.5 and 2.9 for series of a length shorter than hundred observations. The procedure is used to detect additive outliers, which capture isolated spikes as found in Somalia's timeseries, and the required adjustment. For a critical value of 2.75, additive outliers are detected in 2014 and 2016, see table 6.4 and figure 6.3.

Year	Coefhat	t-statistic	GTD 06/17	Adjusted
2014	58.920	27.244	121	63
2016	20.560	4.065	102	82

Table 6.4 Additive outliers Somalia

The number of attacks is adjusted from 121 to 63 in 2014 and from 102 to 82 in 2016. With these adjustments, the carry-over effect of the outliers on the point forecast and the bias in the estimates of the model parameters is reduced [53]. The adjusted fit yields a maximum

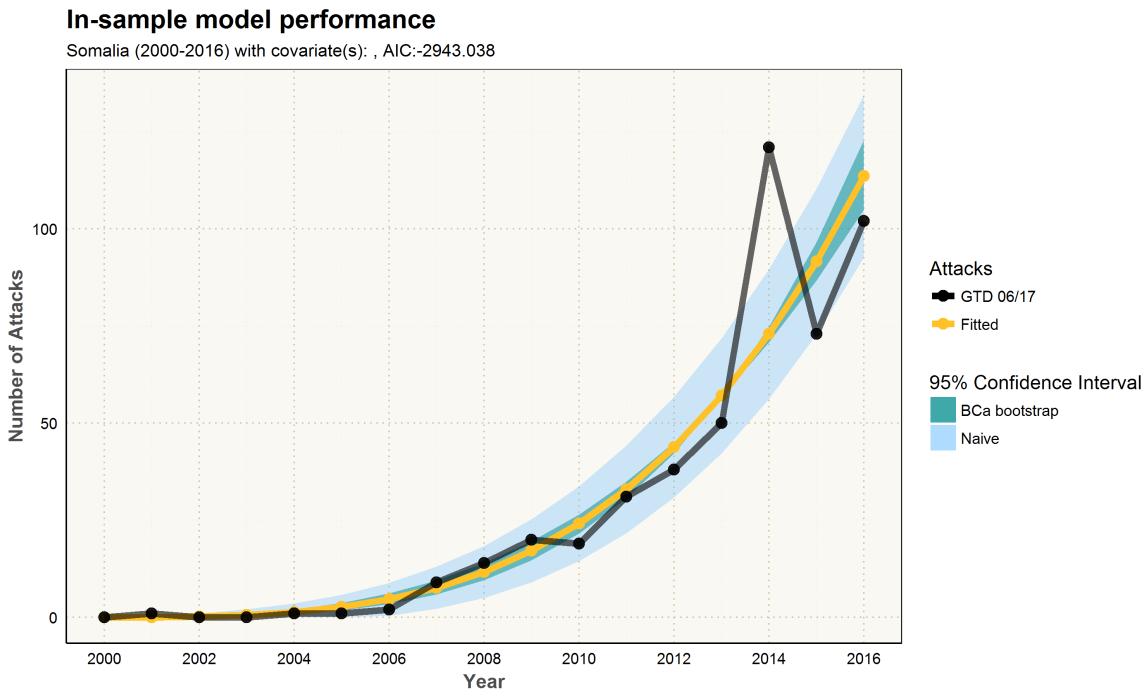

Fig. 6.2 In-sample fit Somalia (2000-2016), including the naive and 10000 sample bias-corrected accelerated (BCa) bootstrap 95% confidence intervals.

log-likelihood of 1139.73 with a corresponding AIC of -2273.45. The parameter estimates are given in table 6.5 and a graphical representation of the fit is found in figure 6.4.

Parameter	Estimate	Standard error	z-value	p-value
δ	4.089	0.203	20.0998	$< 10^{-16}$
Intercept	-5.584	0.579	-9.652	$< 10^{-16}$

Table 6.5 Parameter estimates Somalia outlier adjusted, training data 2000-2016

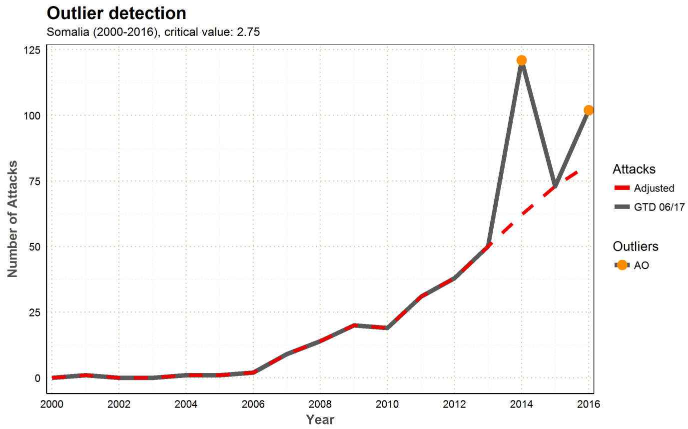

Fig. 6.3 Outlier detection Somalia

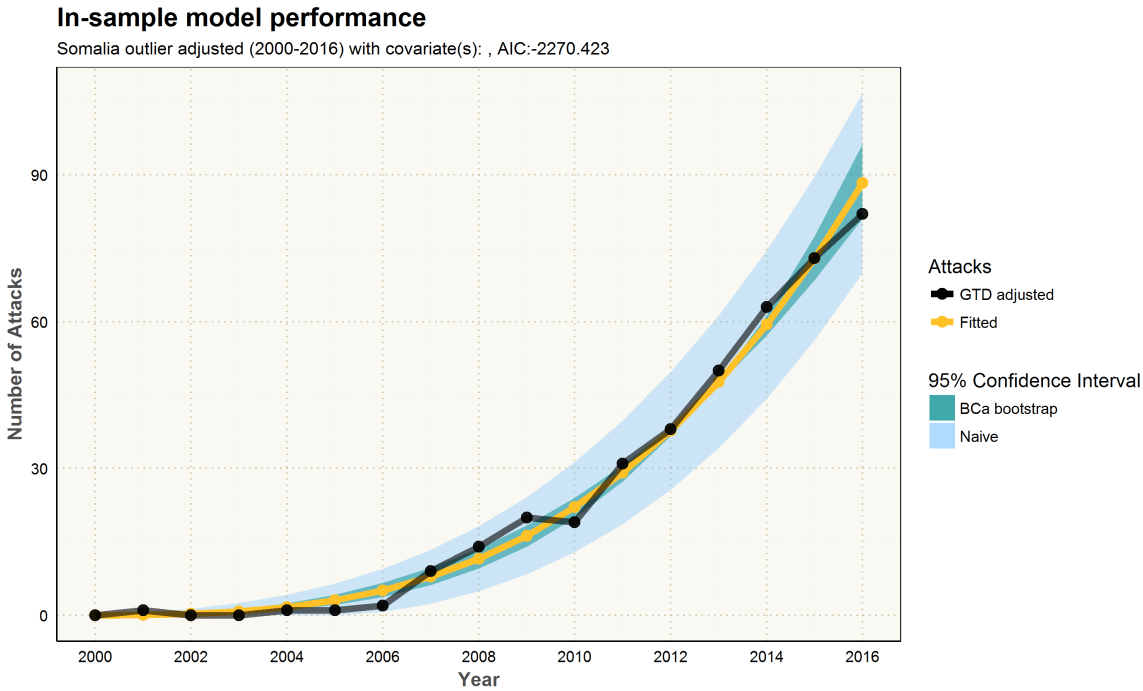

Fig. 6.4 In-sample outlier adjusted model fit Somalia (2000-2016), including the naive and 10000 sample bias-corrected accelerated (BCa) bootstrap 95% confidence intervals.

Although table 6.6 shows the outlier adjusted GTD values are not within the bootstrap BCa confidence interval 9 out 17 times (52.94%), the interval is exceeded by at most two attacks and therefore it can be said the model fits the data very well. In addition, the GTD intensities fall within all of the naive confidence intervals.

Year	GTD adjusted	Fitted	Naive Lower	BCa Lower	GTD adjusted	BCa Upper	Naive Upper
2000	0	0	0	0	0	0	0
2001	1	1	0	1	1	1	1
2002	1	1	0	1	0	1	2
2003	0	1	0	1	0	2	3
2004	1	2	0	2	1	3	5
2005	1	4	0	3	1	5	7
2006	2	6	1	4	2	7	10
2007	9	8	3	7	9	10	14
2008	14	12	5	10	14	14	19
2009	20	17	9	15	20	19	25
2010	19	23	13	20	19	25	32
2011	31	30	19	28	31	31	40
2012	38	38	26	37	38	39	50
2013	50	48	35	48	50	48	62
2014	63	60	45	58	63	62	75
2015	73	73	57	69	73	78	90
2016	82	89	70	82	82	97	107

Table 6.6 In-sample model performance Somalia. The left two columns represent the outlier adjusted GTD 06/17 and fitted values. The columns on the right show the naive and BCa 95% confidence intervals and compare them to the GTD, cells marked red represent GTD values outside of the confidence interval.

The residual plots show no sign of autocorrelation, see figures 6.5 and 6.6, illustrated by the smoothed local polynomial fit. This is confirmed by the ACF/PACF plot of figure 6.7. The model's assumptions seem to be satisfied and out-of-sample validation is performed next.

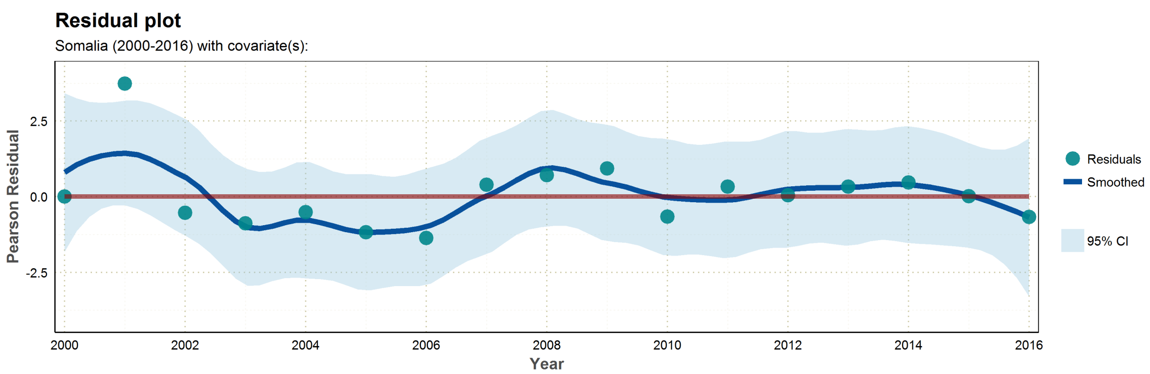

Fig. 6.5 Pearson residuals versus time for Somalia

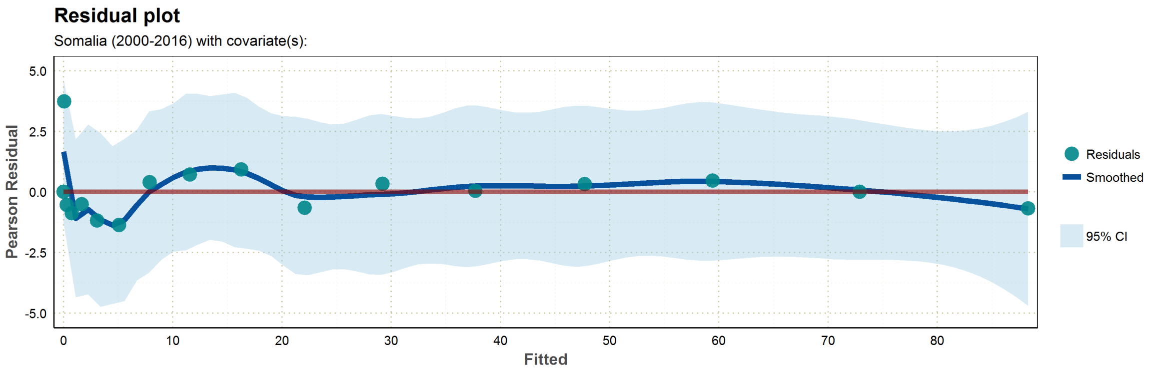

Fig. 6.6 Pearson residuals versus fitted values for Somalia

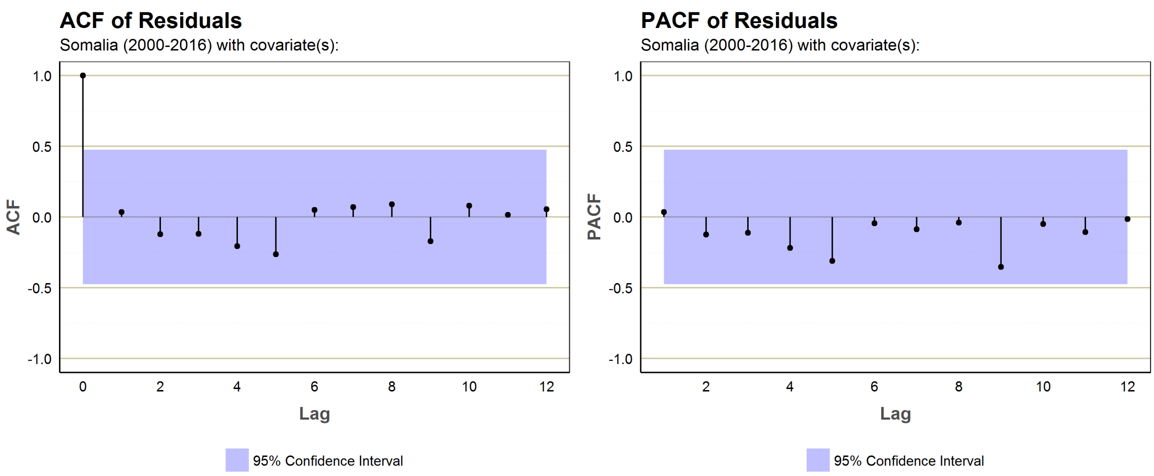

Fig. 6.7 ACF and PACF of residuals for Somalia

6.2.2 Out-of-sample performance

For a minimum out-of-sample period of 10 years (circa 60%), the rolling window approach is pursued with forecast origin T updated from 2009 through 2015. Note that the model is refit for each update. A maximum of three year ahead prediction $H = \min\{3, 2016 - T\}$ for each run, yields the results of table 6.7.

Poor prediction accuracy is observed for a forecast origin of $T = 2009$, with a mean absolute error of 10 attacks and a correponsing mean absolute percentage error of 35.60%. This is caused by a stagnation in 2010 of the strong increasing trend of previous years, see figure 6.8.

Year	2010	2011	2012	2013	2014	2015	2016	MAE	MAPE
GTD	19	31	38	50	63	73	82		(%)
L	17	26	38					10	35.60
U	38	51	66						
L		20	28	38				2	3.09
U		42	53	66					
L			29	40	53			5	9.35
U			54	68	85				
L				37	49	62		3	4.39
U				65	80	97			
L					48	62	78	8	9.37
U					80	97	117		
L						62	78	11	12.26
U						97	116		
L							75	11	13.41
U							112		
							Total	7	12.52%

Table 6.7 Three step ahead predictions for the outlier adjusted attack data of Somalia. L and U represent the lower and upper bound of the prediction intervals, respectively. The most left value of each row represents the one step ahead prediction in year $T+1$ from forecast origin T. A prediction interval is colored red when the GTD value falls outside of the interval. The mean absolute error (MAE) and the mean absolute percentage error (MAPE) are given on the right for each run. The GTD values fall within all prediction intervals.

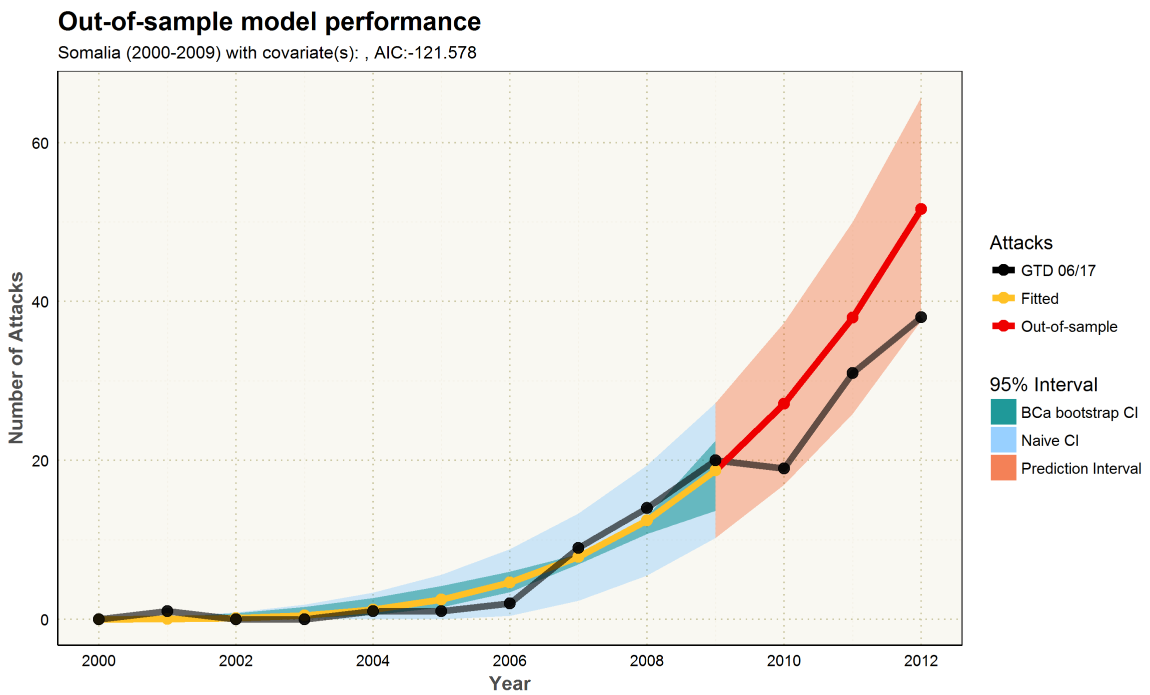

Fig. 6.8 Out-of-sample validation Somalia for training data 2000-2009

Nevertheless, the overall out-of-sample performance is good, with a MAE of 7 attacks and a corresponding MAPE of 12.52% for the total rolling window. In addition, all prediction intervals contain the adjusted GTD values. Forecasts of Somalia's terrorist activity are therefore expected to be a reasonable representation of the future.

6.2.3 Forecast

Since the parsimonious model does not contain covariates, Somalia especially suffers from historical bias; the forecasts that follow are fully based on the trend. The model is trained over the period 2000-2016, with model results as given in section 6.4. A three year ahead forecast leads to the results as given in table 6.8.

Year	**Forecast**	**Prediction Interval**	
		Lower bound	Upper bound
2017	107	87	127
2018	127	105	149
2019	149	125	173

Table 6.8 Forecast Somalia, training data 2000-2016

Based on the general trend, the number of attacks from known perpetrator groups, generating at least 2 deaths, is expected to rise to ca. 150 in 2019. The graphical representation in figure 6.9 might suggest an overestimation of attacks because of the stagnation in 2016. However, this year was outlier adjusted from 102 to 82 attacks.

Considering the devastating impact that terrorism has on Somalia's ongoing humanitarian crisis, the situation in Somalia is expected to deteriorate even further when the Somali government does not manage to get a grip on terrorism.

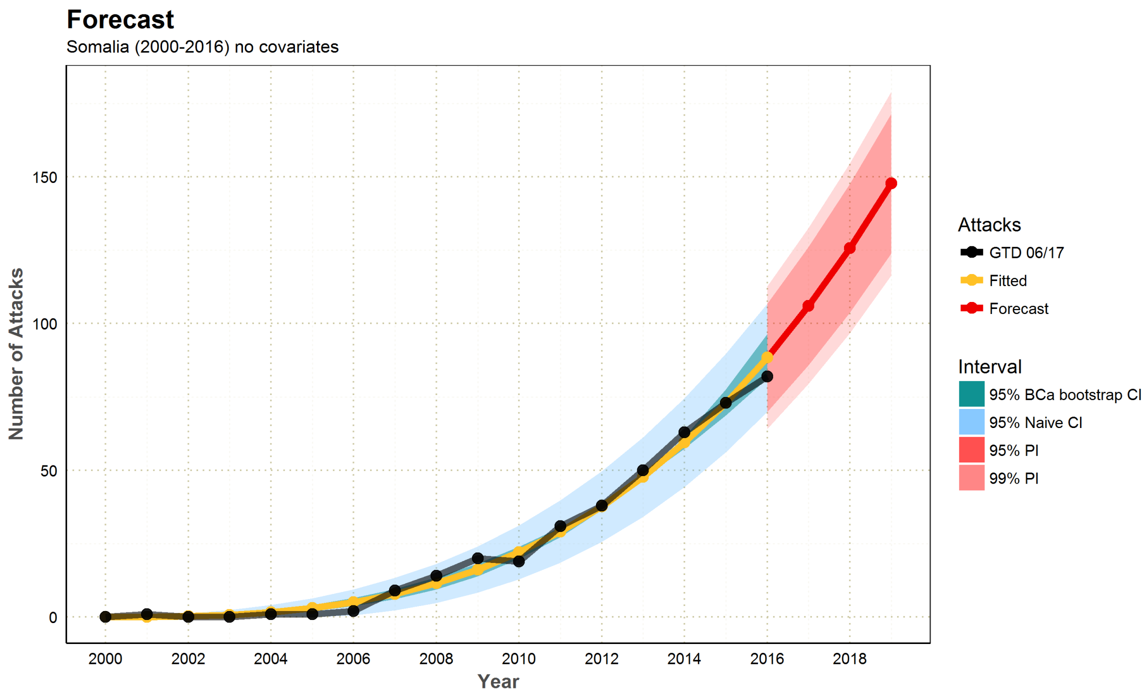

Fig. 6.9 Three year ahead forecast for outlier adjusted Somalia without covariates. Naive and bias-corrected and accelerated 5000 sample bootstrap confidence intervals are shown, as well as the 95% and 99% prediction intervals. The model clearly predicts that the strong increasing trend continues.

6.3 Afghanistan

6.3.1 In-sample performance

Afghanistan's potential covariates are control of corruption, globalisation and political stability. Model selection results in a parsimonious model with control of corruption as only covariate and no intercept included, with a log likelihood of -15373.484, AIC of 7689.742 and parameter estimates shown as in table 6.9. Both control of corruption and the trend parameter δ are extremely significant.

Parameter	Estimate	Standard error	z-value	p-value
δ	3.480	0.0610	57.0320	$< 10^{-16}$
Control of corruption	1.612	0.120	13.470	$< 10^{-16}$

Table 6.9 Parameter estimates Afghanistan, training data 2000-2016

The positive relationship between control of corruption and the number of attacks is remarkable. As section 2.3.3.5 noted, corruption highly weakens Afghanistan's security forces and arms terrorist organisations. Perhaps more important, is the fact that corruption facilitates illicit trade. Afghanistan is by far the world's leading opium producer. The UNODC states that in 2017, between 26-85% of the area under opium poppy cultivation in Afghanistan was located in territory under influence of the Taliban [27]. Furthermore, the UN Security Council Committee estimated in 2011 that half of the overall annual income of the Taliban derives from narcotics [28]. In addition, 95.6% of the selected attacks in Afghanistan come from the Taliban.

Combining the fact that illicit drug trade is of vital importance to the Taliban, with the positive relationship found between control of corruption and the number of attacks, suggests that anti-corruption policies are met with retaliatory measures, because they greatly threaten the organisation's power.

The results of the in-sample fit are shown in figure 6.10, along with the naive and 5000 sample BCa bootstrap 95% confidence intervals. Control of corruption is generally able to capture deviations from the trend. However, between the period 2007 and 2011 the performance is quite poor, with a percentage error of 90.00% in 2007 and -45.76 % in 2010.

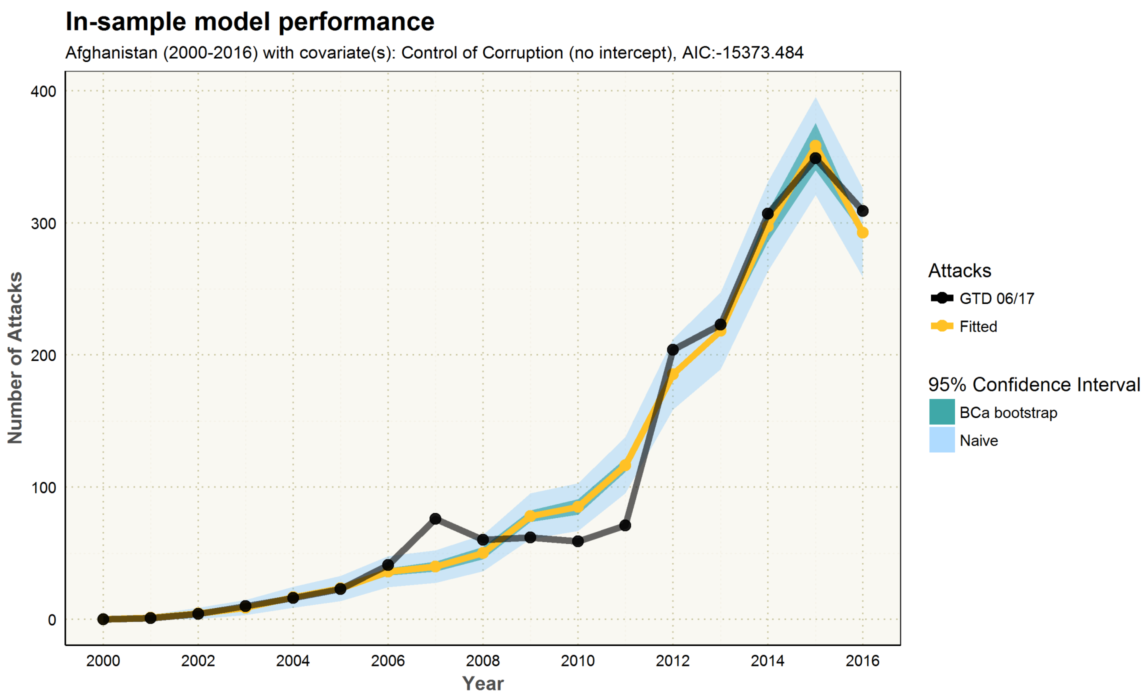

Fig. 6.10 In-sample fit Afghanistan (2000-2016), including the naive and bias-corrected accelerated (BCa) bootstrap 95% confidence intervals.

There are several possible reasons for the mediocre model performance. The model's assumptions could be violated. Plotting the Pearson residuals, does not give indication of a strong signal remaining in the data, see figures 6.11 and 6.12. Still, the wave starting in 2007, illustrated by the smoothed line representing a local polynomial regression fit, could indicate a dependency structure of the residuals. Considering that the ACF and PACF are not significant for any lag, see figure 6.13, there is no reason to believe the model assumptions are violated.

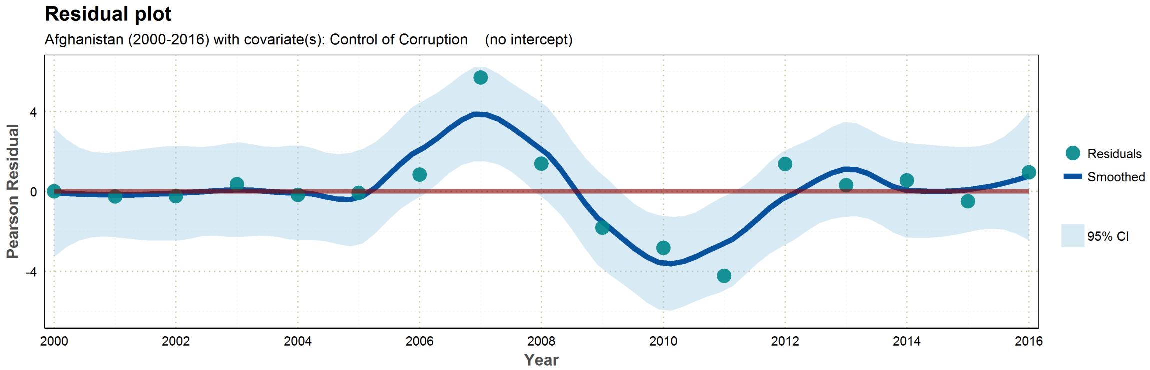

Fig. 6.11 Pearson residuals versus time for Afghanistan (2000-2016)

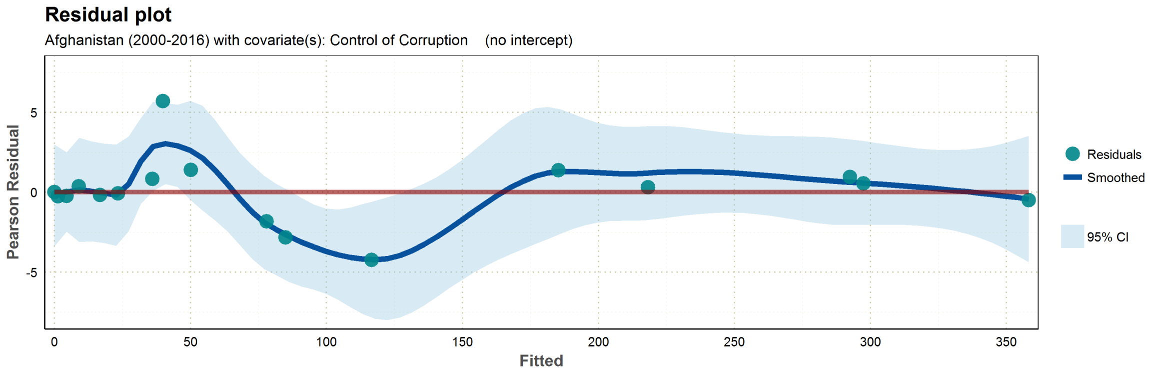

Fig. 6.12 Pearson residuals versus fitted values for Afghanistan (2000-2016)

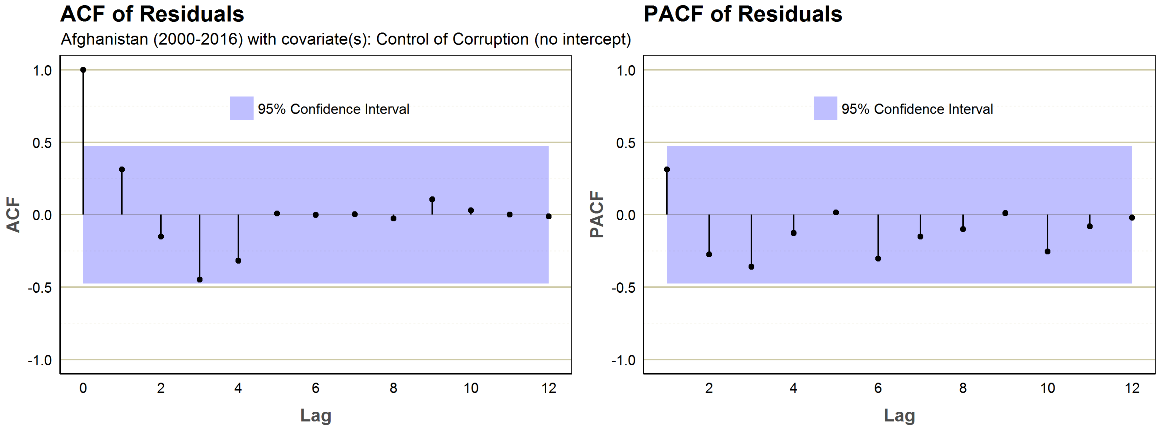

Fig. 6.13 ACF and PACF of residuals for Afghanistan (2000-2016)

As section 3.2 discussed, the GTD switched organisations in April 2008, as well as in November 2011, recognised as a major cause of the vast increase of attacks in 2012. Therefore, the GTD might underestimate the true number of attacks between 2008-2011, a phenomenon which is not reflected by the covariate. Comparing with the other four countries of interest, Pakistan and to less extend Iraq show an abrupt increase of attacks in 2012, while the other countries do not exhibit a pattern that can clearly be linked to changes in data collection methodology.

Nevertheless, a dichotomous dummy is introduced that activates between 2008-2011. If the model fit improves, the database did indeed suffer from the switches of organisations. Model selection results in the parameter estimates of table 6.10.

Parameter	Estimate	Standard error	z-value	p-value
δ	3.235	0.0692	46.765	$< 10^{-16}$
Control of corruption	1.076	0.141	7.631	$2.320 * 10^{-14}$
Organisation dummy	-0.480	0.0800	-5.991	$2.089 * 10^{-9}$

Table 6.10 Parameter estimates Afghanistan including organisation dummy, training data 2000-2016

The *organisation* dummy is highly significant and the AIC of -15408.409 substantially improved, which is illustrated in figure 6.14. Not only did the model fit improve between 2007 and 2011, the dynamics after 2011 are captured even better.

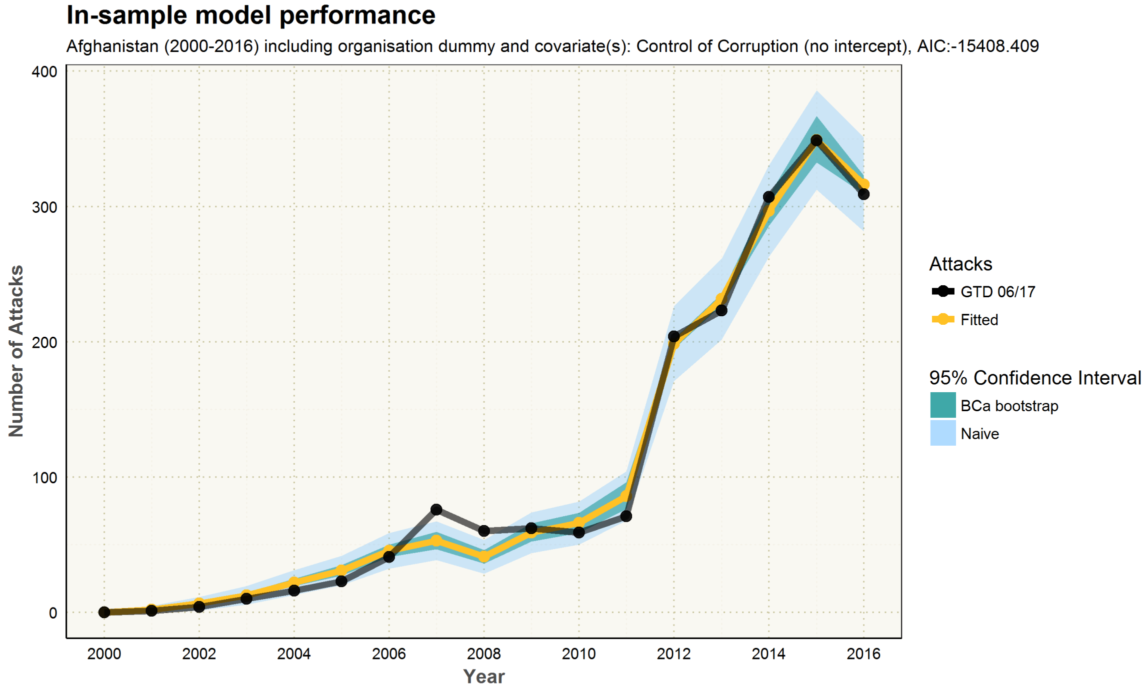

Fig. 6.14 In-sample fit Afghanistan (2000-2016) including organisation dummy, with the naive and bias-corrected accelerated (BCa) bootstrap 95% confidence intervals.

Table 6.11 shows the intensities for both confidence intervals and compares them to the recorded GTD 06/17 attacks. Only 7 out of 17 (41.1%) of the 95% BCa confidence intervals contain the GTD intensities. This gives however a distorted image because in 2006, 2011, 2013 and 2016, the true value lies just outside the interval. The naive confidence interval contains the true number of attacks 15 out of 17 years (88.2%).

The residuals look decent and there is no sign of violation of the model assumption, see figures 6.15 and 6.16. This is confirmed by the ACF/PACF plot of figure 6.17. The out-of-sample validation is therefore performed for the model that includes the organisation dummy.

Year	GTD 06/17	Fitted	Naive Lower	BCa Lower	GTD 06/17	BCa Upper	Naive Upper
2000	0	0	0	0	0	0	0
2001	1	3	0	2	1	3	5
2002	4	7	2	6	4	8	12
2003	10	13	6	10	10	15	20
2004	16	23	13	20	16	25	32
2005	23	32	21	28	23	36	42
2006	41	46	33	42	41	51	59
2007	76	54	39	47	76	60	68
2008	60	42	29	37	60	47	54
2009	62	59	44	53	62	66	74
2010	59	67	51	59	59	74	82
2011	71	86	68	77	71	97	105
2012	204	199	171	195	204	203	227
2013	223	232	202	228	223	236	262
2014	307	297	263	286	307	309	331
2015	349	350	313	333	349	367	386
2016	309	317	282	310	309	323	352

Table 6.11 In-sample model performance Afghanistan. The left two columns represent the GTD and fitted values. The columns on the right show the naive and BCa 95% confidence intervals and compare them to the GTD, cells marked red represent GTD values outside of the confidence interval.

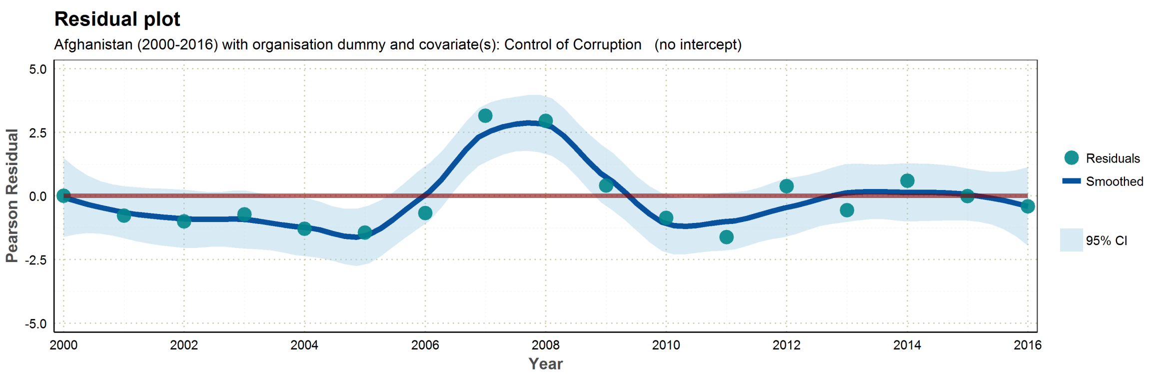

Fig. 6.15 Pearson residuals versus time for Afghanistan (2000-2016), including organisation dummy

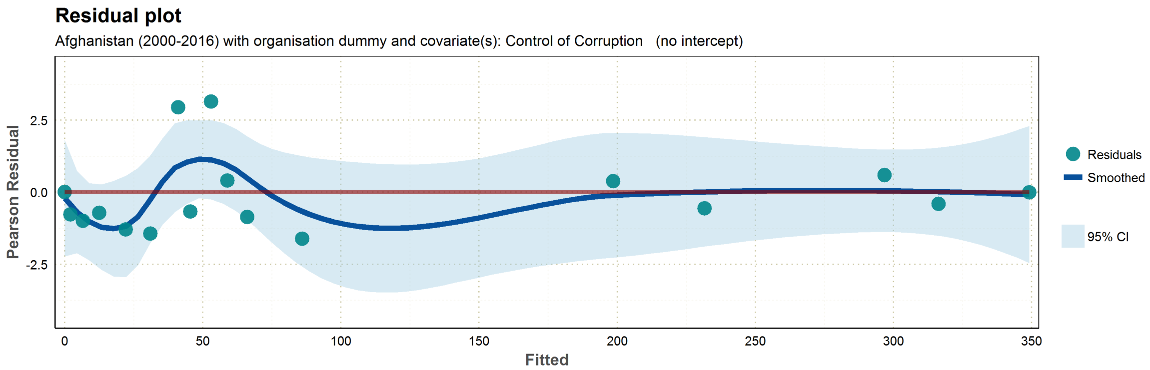

Fig. 6.16 Pearson residuals versus fitted values for Afghanistan (2000-2016), including organisation dummy

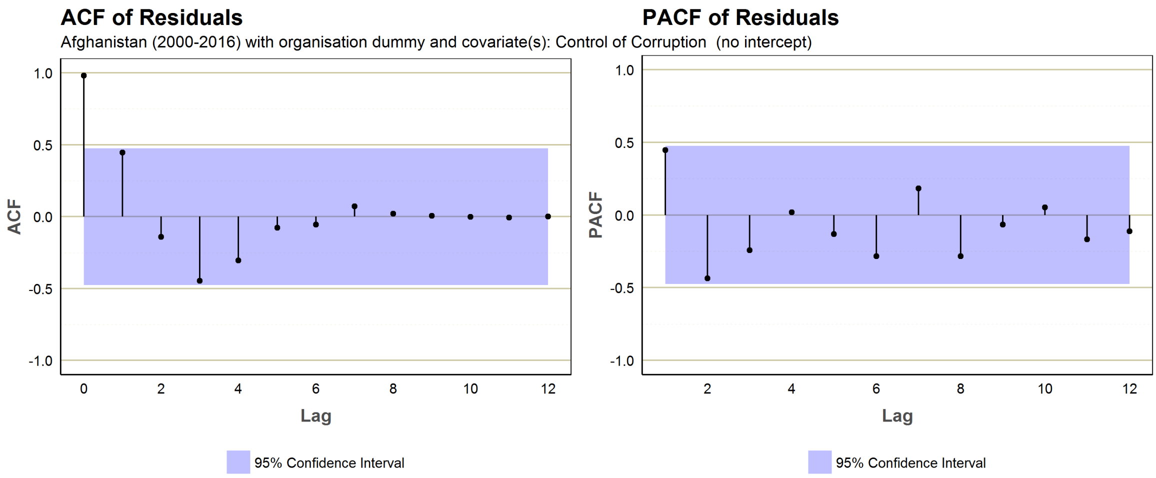

Fig. 6.17 ACF PACF of residuals for Afghanistan (2000-2016), including organisation dummy

6.3.2 Out-of-sample performance

The minimum training period for the out-of-sample validation is set to 10 years (2000-2009), meaning the forecast origin T is updated from 2009 until 2015. Choosing a maximum of three step ahead predictions results in $H = \min\{3, 2016 - T\}$ forecasts for each run. Results are shown in table 6.12.

That the out-of-sample predictions for origins in more distant years are poor, is not surprising. However, instead of the expected underestimation of attacks due to the substantial level shift in 2012, the forecasts overestimate the true number of attacks, see figure 6.18.

Year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**MAE**	**MAPE**
GTD	59	71	204	223	307	349	309		(%)
L	99	139	226					69	86.79
U	142	190	288						
L		131	206	246				58	53.51
U		180	267	311					
L			163	192	247			14	5.52
U			217	250	312				
L				207	269	320		9	3.29
U				267	337	394			
L					259	307	279	9	2.74
U					326	379	348		
L						317	285	8	2.35
U						390	355		
L							284	10	3.24
U							354		
							Total	28	25.75

Table 6.12 Three step ahead predictions Afghanistan. L and U represent the lower and upper bound of the prediction intervals, respectively. The most left value of each row represents the one step ahead prediction in year $T+1$ from forecast origin T. A prediction interval is colored red when the GTD value falls outside of the interval. The mean absolute error (MAE) and the mean absolute percentage error (MAPE) are given on the right for each cycle.

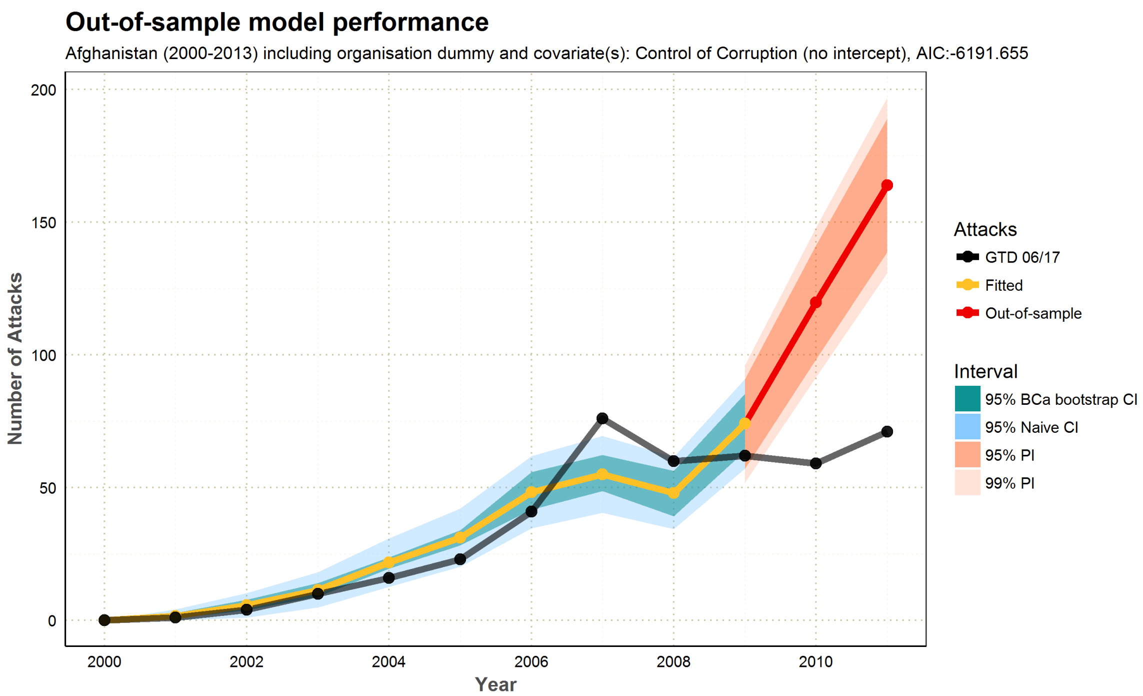

Fig. 6.18 Two year ahead out-of-sample validation Afghanistan, training data 2000-2009, including the naive and BCa 2500 sample bootstrap 95% confidence intervals, plus the 95% and 99% prediction intervals.

This overestimation is caused by the fact that for lower forecast origins T, trend parameter δ is large, see table 6.13. The strong trend is compensated by a larger estimate for both control of corruption -that generally sees a strong decline between 2002 and 2008 that decreases the number of attacks- and the organisation dummy. Note that for larger training data corruption becomes more significant and the organisation becomes extremely signifcant the moment 2012 is added to the training data.

Forecast origin	**δ** par. est.	*p*-value	**Control of corruption** par. est.	*p*-value	**Organisation dummy** par. est.	*p*-value
2009	3.53	$< 10^{-16}$	1.49	$1.51{*}10^{-3}$	-0.38	$1.40{*}10^{-2}$
2010	3.43	$< 10^{-16}$	1.35	$3.26{*}10^{-3}$	-0.46	$1.89{*}10^{-3}$
2011	3.17	$< 10^{-16}$	0.98	$1.29{*}10^{-2}$	-0.46	$1.76{*}10^{-3}$
2012	3.24	$< 10^{-16}$	1.07	$4.56{*}10^{-8}$	-0.50	$5.29{*}10^{-9}$
2013	3.21	$< 10^{-16}$	1.02	$9.37{*}10^{-9}$	-0.48	$3.68{*}10^{-9}$
2014	3.24	$< 10^{-16}$	1.08	$4.81{*}10^{-12}$	-0.49	$4.81{*}10^{-12}$
2015	3.23	$< 10^{-16}$	1.07	$5.13{*}10^{-14}$	-0.49	$2.78{*}10^{-9}$
2016	3.34	$< 10^{-16}$	1.08	$2.32{*}10^{-14}$	-0.48	$2.09{*}10^{-9}$

Table 6.13 Parameter estimates Afghanistan (including organisation dummy) for training data 2000-T, where T is the forecast origin.

Even though the MAE of the total rolling window is 28 attacks, with a corresponding MAPE of 25.75%, the out-of-sample validation can be considered rather decent. The model that does not include the organisation dummy exhibits a large MAE of 51 attacks with a MAPE of 26.59%. Without the dummy, the three year ahead predictions for smaller T are better. However, the large MAE shows that for later forecast origins, the original model clearly has worse results.

Including again the organisation dummy, the MAE is at most 14 attacks with a MAPE smaller than 6% for $T \geq 2011$. That the dynamics of terrorist activity are captured very well for more recent T, is illustrated in figure 6.19. Given that terrorism in Afghanistan is not subjected to structural changes in the upcoming years, our model is therefore expected to give a reasonable representation of the amount of terrorist attacks in the near future.

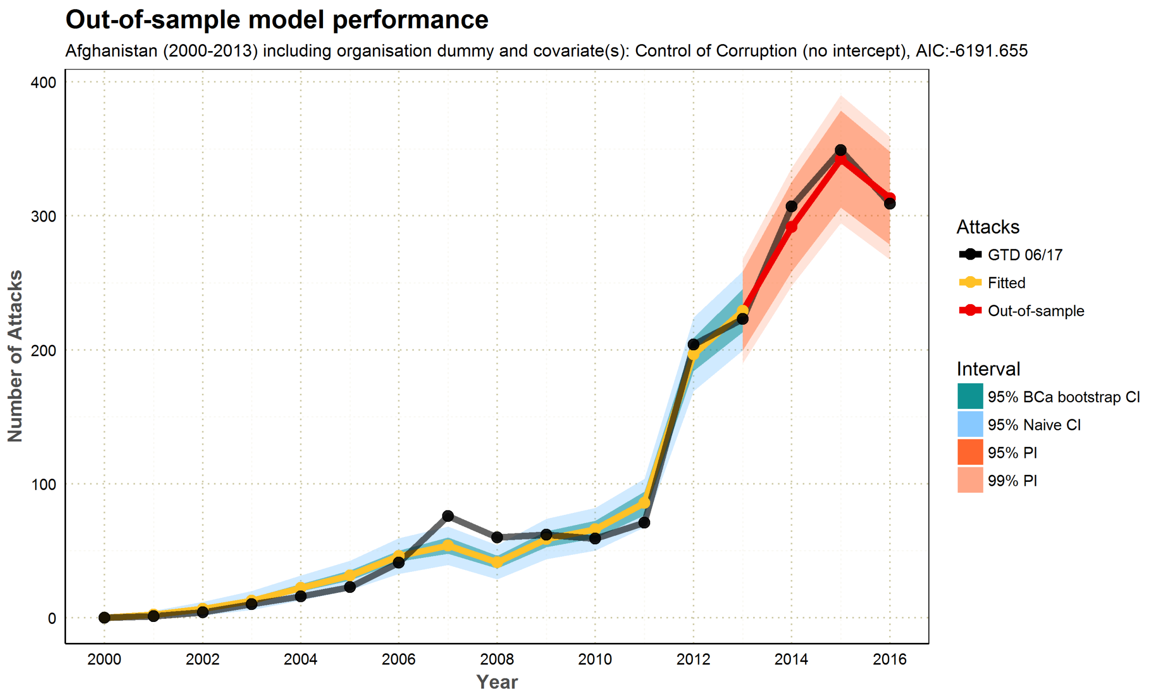

Fig. 6.19 Out-of-sample fit Afghanistan, training data 2000-2013, including the naive and bias-corrected accelerated (BCa) bootstrap 95% confidence intervals, plus the 95% and 99% prediction intervals.

6.3.3 Forecast

As mentioned, the model suffers from historical bias, since the past, showing an increasing trend, is used as a predictor of the future. All forecasts are made assuming this trend continues. For the forecasting, the model is trained over the period from 2000 to 2016, with control of corruption and the organisation dummy included, without an intercept and in-sample results as given in section 6.4.1. Since control of corruption is available until 2016, a three year forecast of terrorist activity requires a three year covariate scenario forecast as introduced in section 6.1. Figure 6.20 shows the covariate trajectories of the scenarios.

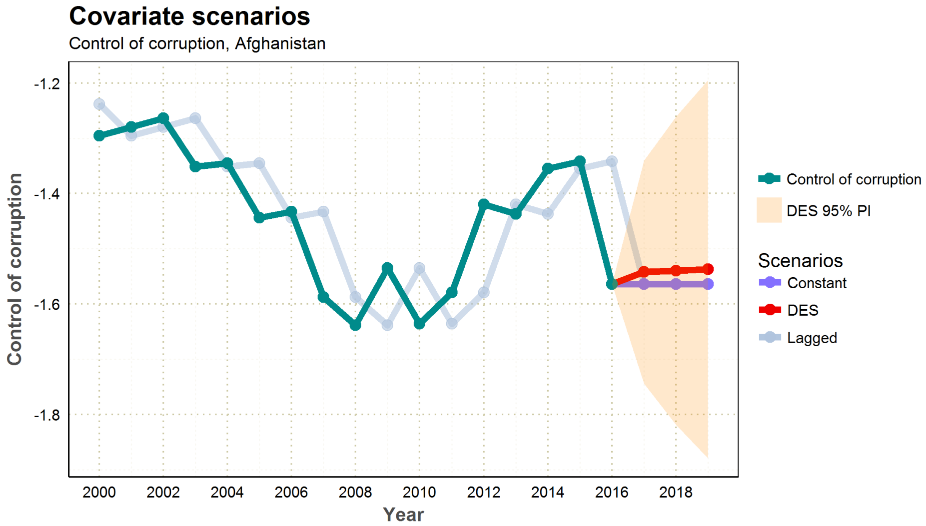

Fig. 6.20 Forecast of the three covariate scenarios of Afghanistan.

The variance of the DES forecast is so large that the forecast itself is not reliable. However, the bounds of the prediction interval do provide an indication of the extreme scenarios. The forecasts for both the constant and DES scenario are shown in figure 6.21. There is little difference between the constant and Holt-Winters DES scenario. The constant scenario naturally follows the baseline trend, predicting 361 terrorist attacks in 2017, 409 in 2018 and 460 attacks in 2019.

Interesting are the extrema of the DES prediction interval. Assuming that terrorist organisations indeed respond to anti-corruption policies, then even in the case of complete deterioration of corruption, which has a decreasing impact on terrorist activity, the annual amount of attacks stays above 300. Conversely, when Afghanistan succesfully fights corruption, terrorism could increase to up to 680 attacks per year in 2019.

Between the period 2000-2016 our data selection comprised only 18.78% of the true amount of attacks in Afghanistan as according to definition 2.2. Therefore, implementing new anti-corruption policies could have severe consequences, if the government and international parties involved do not anticipate accordingly.

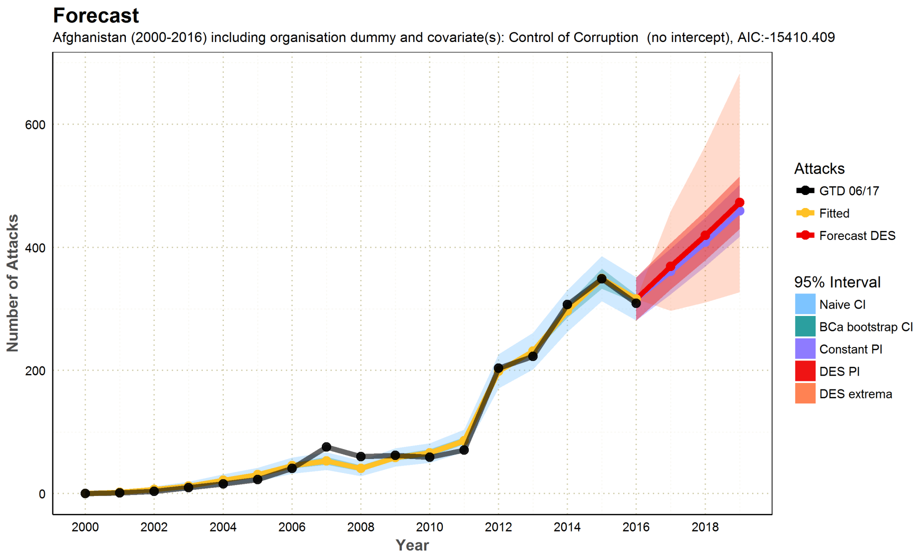

Fig. 6.21 Forecasts for the constant and DES scenario of Afghanistan. The model clearly predicts the strong increasing trend persists. The DES extrema scenario shows that implementing anti-corruption policies could have severe consequences.

For the lag one model, control of corruption is again highly significant, but with a mediocre in-sample fit (AIC -15382.384) as compared to the non lagged model, out-of-sample validation and forecasting are not pursued.

6.4 Iraq

6.4.1 In-sample performance

Model selection results in a parsimonious model with the two highly significant covariates control of corruption and globalisation. Globalisation is scaled beforehand by dividing the index by 25, because the metric is rated from 0 to 100. Recall that the covariate is available only until 2015.

The in-sample model fit for the period 2000-2015 exhibits a log-likelihood of 1804.85 and corresponding AIC of -3599.71, along with the parameter estimates as shown in table 6.14.

The direction of the relationships of the covariates with terrorist activity is as theory from social science claimed. The highly significant, positive relationship between globalisation and terrorist activity seems to confirm the idea of Islamic fundamentalism as a response to changes in social and economic structures imposed by, inter alia, globalisation. This rise

Parameter	Estimate	Standard error	z-value	p-value
δ	4.074	0.130	31.353	$< 10^{-16}$
Intercept	-38.770	3.172	-12.223	$< 10^{-16}$
Globalisation	17.416	1.940	8.976	$< 10^{-16}$
Control of corruption	-4.956	0.676	-7.331	2.292^{-13}

Table 6.14 Parameter estimates Iraq, training data 2000-2015

against modernity, cultural imperialism and inequality was specifically illustrated with the emergence of terrorist organisations in Iraq.

Contrary to Afghanistan, decreasing control of corruption in Iraq sees the expected increase in terrorist activity. According to Iraqi lawmakers, the United States provided weapons to the Iraqi army, of which a part got sold to IS [54]. This not only considerably increased the power of IS, but also made the army less effective in fighting the organisation. Furthermore, the extreme forms of corruption on all levels, while Iraq's oil revenue keeps rising and living conditions deteriorate, could inspire people to resort to radical ideologies.

Note that both the intercept and globalisation take on large absolute values. For globalisation, this can be adjusted by a different scaling of the indicator. However, the intercept is barely effected by scaling the covariates and rather compensates for the large trend parameter δ. This might indicate that the model is overfitting. The out-of-sample analysis includes a grid search to evaluate the trade-off between the in-sample and out-of-sample performance, that shows parameter optimisation bounds of (-40,40) are reasonable.

The model fit is shown in figure 6.22, along with the naive and 5000 sample BCa bootstrap 95% confidence intervals. The country data exhibits rough behavior which explains why the data showed deviations from the straight line in the Weibull plot of figure 4.15. The first true emergence of attacks in 2004 is likely to be the consequence of the 2003 Iraq invasion, as mentioned in section 2.1.1. The vast upsurge of attacks in 2014 is appointed to the seizure of large areas of the country by ISIL [16].

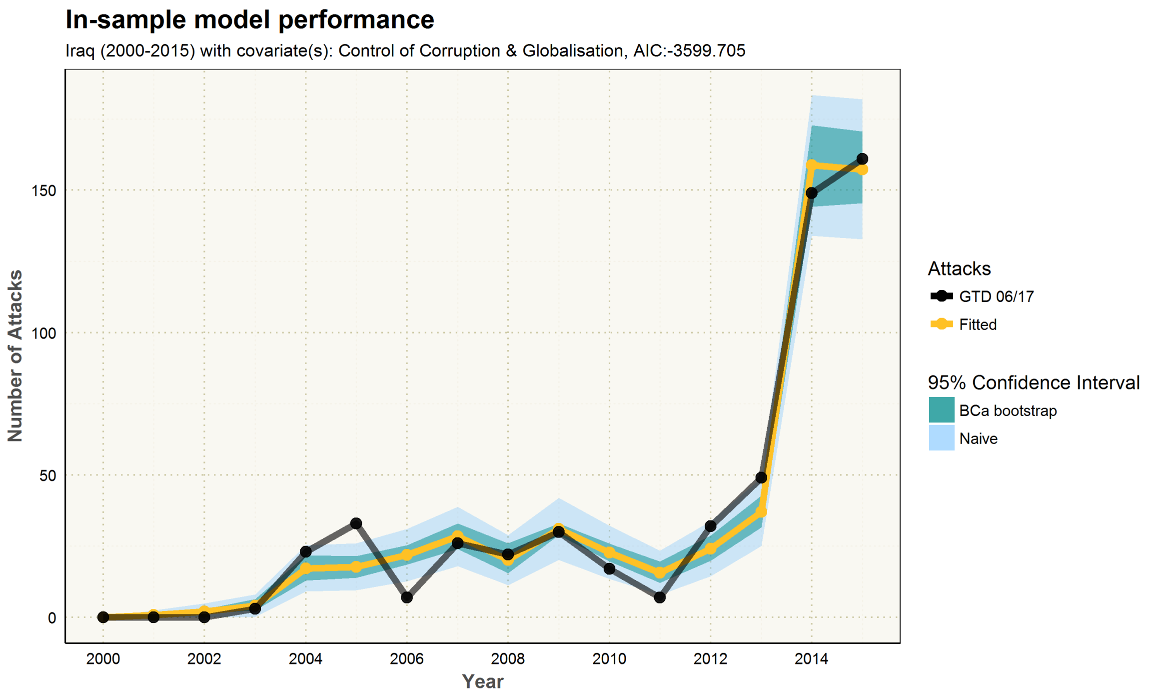

Fig. 6.22 In-sample fit Iraq (2000-2015), including the naive and BCa 5000 sample bootstrap 95% confidence intervals.

Table 6.15 shows the intensities for both confidence intervals and compares them to the recorded GTD 06/17 attacks. Although only 7 out of 16 (43.8%) BCa confidence intervals contain the true value, 13 out of 16 (81.3%) of the naive intervals contains the observed number of attacks. Considering the deviations from the trend, the covariates capture the dynamics of the number of attacks reasonably well.

The model is clearly not accurate in 2005, 2006 and 2011, as figure 6.23 illustrates. However, with an overall mean absolute error of 6 attacks, the results are regarded decent.

Year	GTD 06/17	Fitted	Naive Lower	BCa Lower	GTD 06/17	BCa Upper	Naive Upper
2000	0	0	0	0	0	0	0
2001	0	1	0	1	0	2	3
2002	0	3	0	2	0	3	5
2003	3	5	1	3	3	7	9
2004	23	18	10	13	23	22	26
2005	33	18	10	14	33	22	27
2006	7	22	13	19	7	26	32
2007	26	29	18	25	26	33	39
2008	22	21	12	16	22	27	29
2009	30	32	21	29	30	33	42
2010	17	23	14	20	17	26	33
2011	7	16	8	13	7	20	24
2012	32	25	15	20	32	29	34
2013	49	37	26	32	49	43	49
2014	149	159	135	145	149	173	184
2015	161	158	133	146	161	171	182

Table 6.15 In-sample model performance Iraq. The left two columns represent the GTD and fitted values. The columns on the right show the naive and BCa 95% confidence intervals and compare them to the GTD, cells marked red represent GTD values outside of the confidence interval.

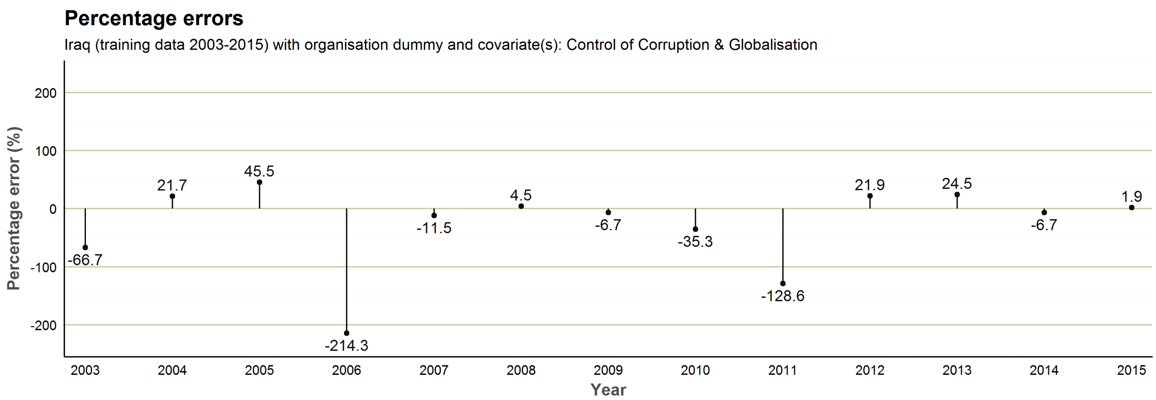

Fig. 6.23 Percentage errors for in-sample fit Iraq. Since the annual number of attacks is zero before 2003, these years are not shown.

Plotting the Pearson residuals over time, shows the residuals are zero on average, see figure 6.24. Furthermore, the residuals versus the fitted values do not show dependency structures, see figure 6.25. Both the autocorrelation and partial autocorrelation do not exhibit significant values, see figure 6.26. There is thus no indication that the model assumptions are violated.

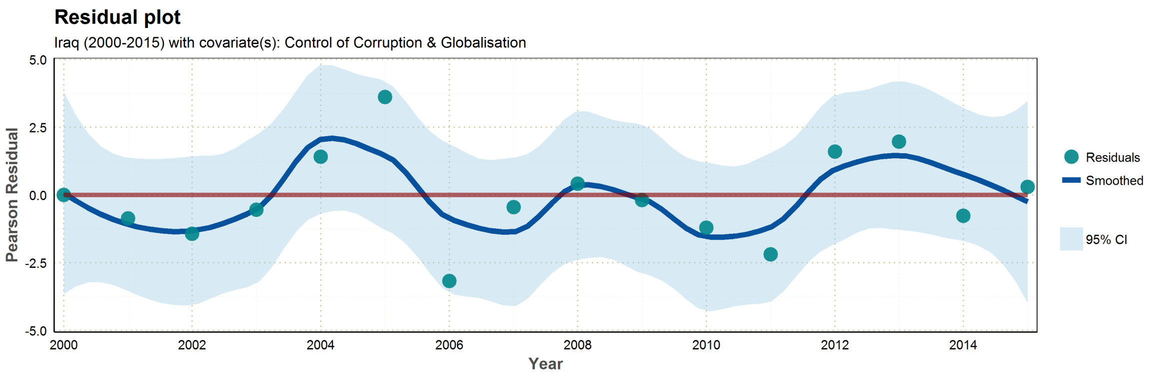

Fig. 6.24 Pearson residuals versus time for Iraq.

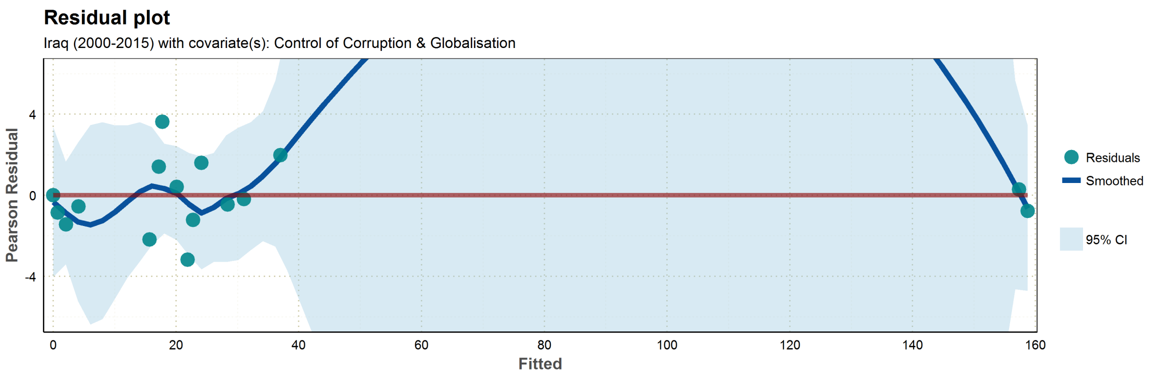

Fig. 6.25 Pearson residuals versus fitted values for Iraq. Smoothing is cutoff for clarity of the figure.

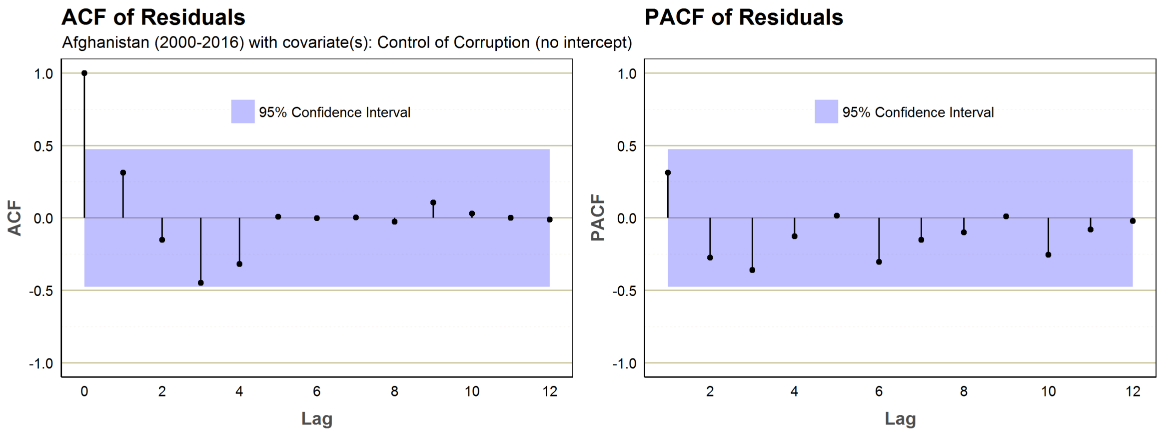

Fig. 6.26 ACF and PACF of residuals for Iraq

6.4.2 Out-of-sample performance

The minimum training period is set to 10 years (2000-2009), such that the forecast origin T is updated from 2009 through 2014. Choosing a maximum of three year ahead predictions results in $H = \min\{3, 2015 - T\}$ forecasts for each cycle. As turns out from performing the rolling window analysis, the out-of-sample performance of Iraq is highly dependent on the bounds imposed on the intercept in the optimisation. A grid search of different bounds reveals the trade-off between in-sample and out-of-sample performance, see table 6.16. It appears that for bounds stricter than circa (-40,40) both the in-sample and out-of-sample performance decrease. For bounds more liberal, the in-sample fit of each rolling window cycle slightly increases at the cost of out-of-sample performance. When comparing results for the bounds inbetween (-35,35) and (-45,45), it turns out that a bound of (-40,40) results in a decent balance between in-sample and out-of-sample performance.

Still, with a MAE of 14 attacks and MAPE of 37.74% for the total rolling window, the model does not perform well. Control of corruption and globalisation do considerably improve the baseline model results (MAE of 51 attacks and MAPE of 106.31% for the total rolling window).

Forecasting contains an extra setback because the globalisation index is not known for 2016. Assigning the 2015 value to 2016 for globalisation, allows to compare the prediction of 2016 to the GTD value, see figure 6.27. With 219 attacks the model is very close to the true 216 attacks (percentage error of 1.39%).

For optimisation bounds (-40,40), the rolling window is extended one year by assuming the globalisation index constant in 2016. Updating forecast origin T from 2009 through 2015, with $H = \max\{3, 2016 - T\}$ predictions per cycle, the window exhibits a MAE of 13 attacks

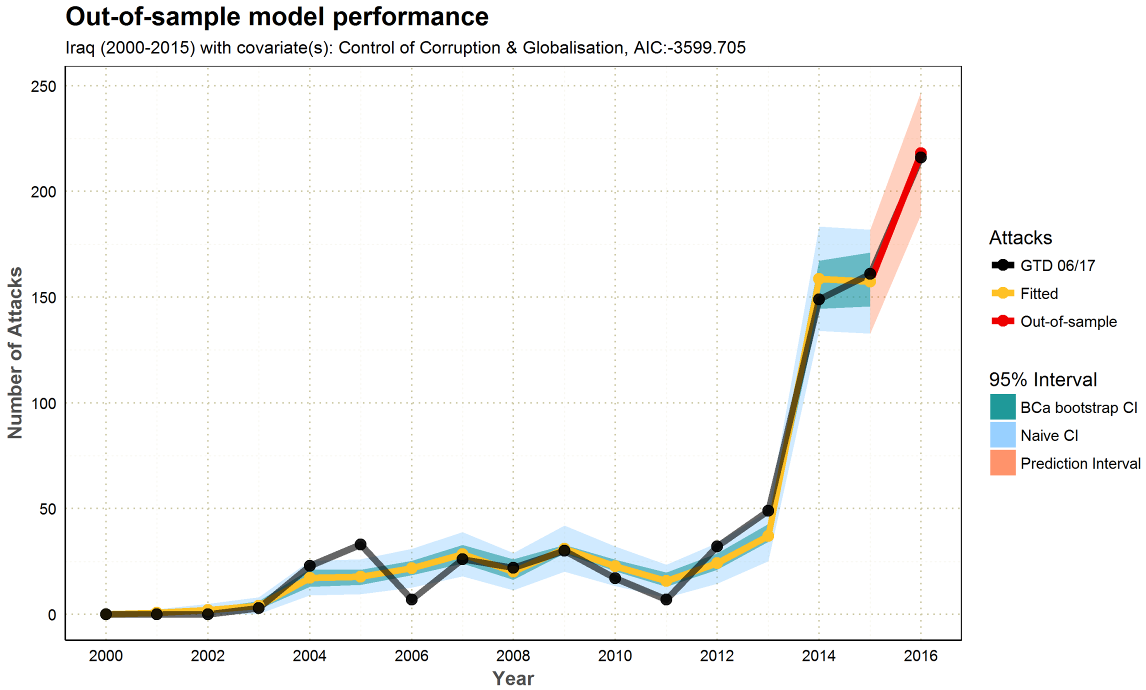

Fig. 6.27 One year out-of-sample validation Iraq, training data 2000-2015, with the globalisation index kept constant in 2016. The naive and 95% BCa 2500 sample bootstrap confidence intervals are shown for the in-sample fit.

with a MAPE of 32.22%. Reducing to two year ahead predictions does not improve results, with a MAE of 11 attacks and 37.17%. Therefore, forecasting can only provide an indication of the general trend in the near future.

6.4.3 Forecast

For the forecast, the model is trained between 2000 and 2016 and the globasliation index is assumed constant in 2016. A two year ahead forecast is shown in figure 6.28. Considering the out-of-sample results, the only conclusion that can be drawn is that the strong increasing trend in expected to persist.

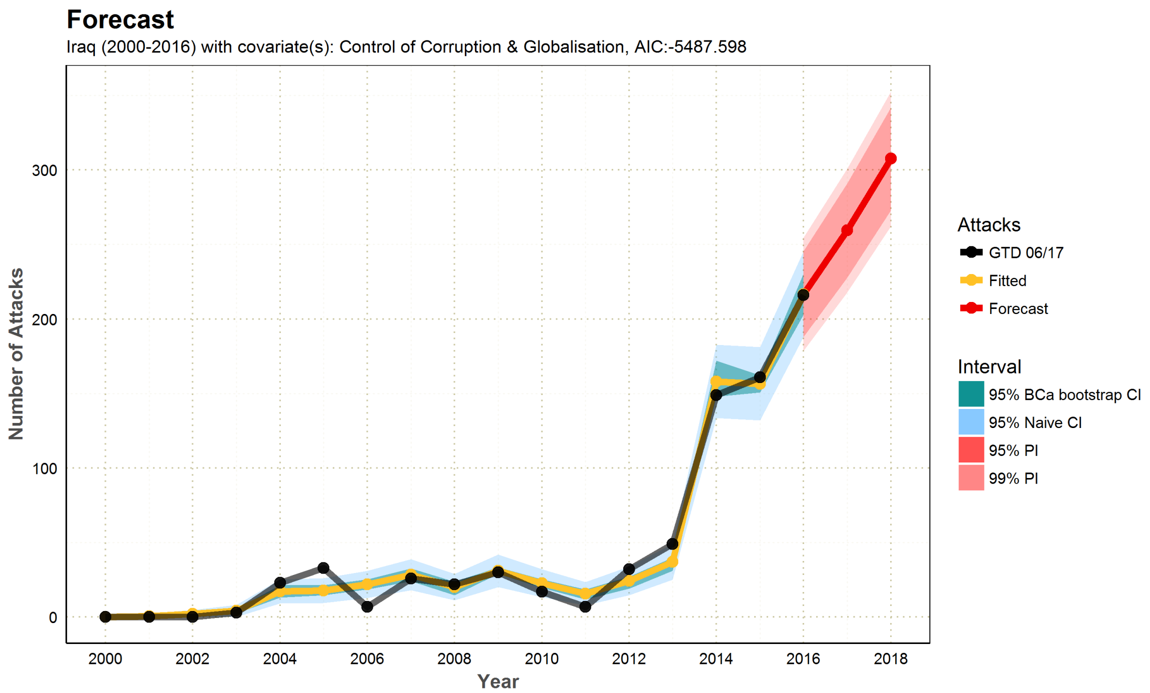

Fig. 6.28 Two year ahead forecast Iraq (2000-2016) for the constant covariate scenario, including the naive and BCa 2500 sample bootstrap 95% confidence intervals, as well as the 95% and 99% prediction intervals. Note that training the model over the period 2000-2016 required an extra value of the globalisation index, the 2015 value was used.

Bounds	**Metric**	2009	2010	2011	2012	2013	2014	**Total**
	AIC	-587.64	-648.89	-661.18	-811.14	-1093.36	-2286.94	-
(-100,100)	MAE	7	13	44	71	191	7	53
	MAPE (%)	47.87	50.28	52.21	52.54	123.78	4.35	57.37
	AIC	-587.52	-648.63	-660.31	-810.83	-1092.43	-2286.94	-
(-50,50)	MAE	8	12	20	40	90	7	29
	MAPE (%)	53.55	49.23	35.51	32.40	58.19	4.35	42.19
	AIC	-587.04	-647.94	-659.13	-810.02	-1091.21	-2286.94	-
(-45,45)	MAE	8	12	15	18	51	7	18
	MAPE (%)	62.96	53.32	31.93	18.59	32.89	4.35	38.03
	AIC	-586.16	-646.79	-657.40	-808.66	-1089.45	-2286.94	-
(-40,40)	MAE	8	12	22	11	18	7	14
	MAPE (%)	72.36	57.04	36.04	13.93	11.83	4.35	37.74
	AIC	-584.86	-645.14	-655.06	-806.72	-1087.11	-2285.65	-
(-35,35)	MAE	9	12	29	27	10	10	17
	MAPE (%)	83.72	60.08	40.51	24.55	6.00	6.21	42.99
	AIC	-583.12	-642.95	-652.08	-804.15	-1084.16	-2279.62	-
(-30,30)	MAE	9	12	34	40	32	16	25
	MAPE (%)	104.48	71.60	45.96	39.90	33.21	13.66	57.73
	AIC	-578.22	-636.83	-643.94	-796.88	-1076.20	-2252.50	-
(-20,20)	MAE	13	12	42	60	67	29	36
	MAPE (%)	130.58	74.64	48.42	46.41	43.24	18.01	66.97

Table 6.16 In-sample and out-of-sample results Iraq, for three year ahead rolling window cycles from forecast origin T for different optimisation bounds. The years represent forecast origin T from which $H = \max\{3, 2015 - T\}$ predictions are made. For each cycle, the AIC of the in-sample fit (training period from 2000 to T) and the MAE and MAPE of the predictions are given. The outer right column represents the MAE and MAPE for the total rolling window for each optimisation bound. Note that for every cycle the parameters stay significant. For bounds stricter than (-40,40) both the in-sample and out-of-sample performance decreases. For bounds more liberal, the in-sample fit of each rolling window cycle slightly increases, at the cost of out-of-sample performance.

Chapter 7

Conclusion

This study pursued two main goals. First, it examined the relationship between various annual country statistics and the number of attacks. Secondly, it assessed the potential of forecasting terrorist activity. This section discusses to what extend these goals have been achieved and summarises the findings obtained over the course of this study. Recall that in the analysis, attacks have been selected that have an identified perpetrator and satisfy all three criteria of definition 2.1.

First of all, various conclusions that concern the data are discussed. Although the GTD is widely recognised as the most comprehensive open-source terrorism database, the changes in data collection methodology impacted the model results of Afghanistan. This confirms the claim by Pape *et al.* [123] that the GTD is a questionable source when interpreting recent trends. Future research should take this limitation into account.

In addition, the database includes a significant amount of incidents which are legal according to international humanitarian law. These events are not recognised as terrorist attacks according to the view of world leading international organisations. Nevertheless, the GTD allows the possibility to filter out these incidents, which solves the problem.

The choice of problematic countries in less developed regions of the world, is interesting from a research perspective, but limits the model results, because the country statistics can be scarce and unreliable. Especially demographic data such as population growth and urbanisation rate, is not accurate enough for improving results when modelling countries individually. The lack of accurate, reliable country statistics is a huge loss to modelling terrorism. Fortunately, modelling countries individually does eliminate the flaw of limited comparability of statistics between countries due to different collection procedures and estimation methods.

Combining the limitations of the data with the fact that a mathematical model, i.e. an simplification of a complex phenomenon by a formula, was able to produce partially decent model results, seems to provide a tentative answer to the general debate whether terrorism can be explained by root causes. Still, for two out of five countries, the results were poor. Although Somalia showed good results, there were no significant covariates. This could be due to the fact that country statistics of the country are very scarce. Still, the attacks occurred according to a natural pattern likely to be rooted in sociology and human psychology, that follows a Weibull trend remarkably well. Afghanistan was the only country for which the covariate captures the dynamics of terrorist activity both in-sample and out-of-sample.

Combining that the countries showed a range of different covariates in their parsimonious model with the fact of mixed model results, seems to suggest that a general statement on explaining terrorism might not be applicable. It appears that, although countries do share a common trend, conflicts are so diverse that explaining terrorism should rather be pursued conflict-, group-, country-, or case specific. This study tried to accomplish this by modelling countries individually, in addition to exclusively selecting attacks that are not legal according to international humanitarian law, coming from identified perpetrators.

Still, the model results gave rise to interesting findings from a sociological perspective. Above all, the control of corruption index was able to explain the dynamics of terrorist terrorist activity in Afghanistan remarkably well, both in-sample and out-of-sample. The direction of the relationship was surprising; as Afghanistan gets more grip on corruption, the number of attacks rise. The combination of the facts that corruption facilitates illicit trade, half of the overall annual income of the Taliban derives from narcotics (mainly opium) and the vast majority of attacks included in the analysis come from the Taliban, led to conclude that anti-corruption policies are met with retaliatory measures, because they greatly threaten the organisation's power. Assumed this is true, the forecasts of extreme corruption scenarios showed that increasing the fight against corruption could have severe consequences, if the government and international parties involved do not anticipate accordingly.

Backed by good out-of-sample model results, forecasts for Somalia clearly indicated a continuation of the strong increasing trend in number of attacks. Considering the devastating impact that terrorism has on Somalia's ongoing humanitarian crisis, the situation is expected to deteriorate even further.

The in-sample fit of Iraq showed that control of corruption and globalisation reasonably explain the rather rough dynamics in terrorist activity. Contrary to Afghanistan, decreasing control of corruption saw the expected increase in the number of attacks. The positive relationship between globalisation and terrorist activity seemed to confirm the idea of Islamic fundamentalism as a response to changes in social and economic structures, imposed by, inter alia, globalisation. The government of Iraq should therefore take measures to disperse the benefits of globalisation more equally within the nation. However, considering the fact that

Iraq is one of the most corrupt countries in the world, with corruption on all levels of society [22], this is unlikely to happen. Since the out-of-sample results were poor, more research on the relationship between these covariates and terrorism is required to support these findings.

Reflecting on Rapoport's wave theorem, the first three waves lasted about fourty years. The current, *religious* wave emerged around 1980, meaning that if history were to repeat itself, it would end around 2020. Based on the model predictions and the general worldwide increasing trend, the religious wave will almost certainly last longer than its predecessors. Rapoport's argument that religious communities are very durable, as compared to those inspired by a secular cause, seems to hold.

Finally, this study yielded various mathematical findings. Attack data of five countries were shown to be consistent with a NHPP, after removing attacks that did not generate more than a country-specific threshold x amount of deaths. These countries were Afghanistan (x=4), Iraq (x=8), Pakistan (x=2), Somalia (x=2) and Yemen (x=2). Attacks from Nigeria are not suited for a Poisson model, because they are highly clustered.

For the period 2000-2014, graphical evaluation showed that a Weibull baseline model is reasonable for all countries. However, extension to the GTD 06/17 version revealed a structural change in trend for Pakistan, starting in 2012 that persists through 2016. Part of the introduced aggregated targets seem to follow the Weibull trend, however there are clear deviations such as Military & Police. By rough graphical evaluation, piecewise proportionality between the intensities of the countries seemed reasonable, except for Nigeria and Iraq.

The distribution of lethal attacks ($x = 1$) of the six countries combined, was shown to follow a Pareto distribution with finite mean and infinite variance.

Considering the limitations of the data (both the GTD and the country statistics) and the fact that two out of five countries showed decent out-of-sample results, the PPWBPC model is deemed reasonable. Even though out-of-sample results for the other countries turned out poor, a relatively simple model was able to produce at least partly satisfactory results for a very complex phenomenon. Therefore, the conclusive answer to the potential of forecasting the number of attacks, is that this study indeed sees potential in forecasting terrorist attacks, but recommends a case-specific, detailed approach, rather than a general cross-national study.

7.1 Future Work

Considering the results of the model, there are opportunities for future research on modelling terrorism. As was concluded, this study recommends the modelling of a detailed selection of case-, conflict-, organisation-, or region specific attacks, rather than a general cross-national study. In particular, attacks from the Taliban should be examined closely for their relationship with anti-corruption policies and opium production. A more detailed modelling approach could provide the accuracy needed for the implementation in policy planning. In this way, the government and international parties involved can construct anti-corruption strategies to fight the Taliban, while taking the appropriate measures to the expected retaliation. Since the GTD provides the geographical location of attacks, terrorism within areas of opium production could be compared to attacks in major cities, to see if there is a difference in the relationship to anti-corruption policies.

Instead of merely modelling the number of attacks, Young argues that counting the number of deaths may be a more valid way of measuring terrorism, since the impacts are more obvious [94]. This can be done by using a marked point process, that, in addition to modelling the attack times, considers the lethality of the attacks. An introduction to applying marked point processes to terrorist attacks is found in chapter five of Lum *et al.* [110].

At the very end of writing this thesis, an interesting paper by Tench *et al.* [132] was found, which models improvised explosive device attacks in North Ireland by a Hawkes process. This process takes into account that attacks are not independent. The idea is that the intensity has a background rate that gets a boost when an attacks happens, temporarily increasing the probability of a subsequent attack. Next to a more accurate prediction of the number of attacks, they find that Hawkes processes can provide insight on how terrorist groups will respond to different events. The latter has great potential for counter-terrorist strategies and requires further research.

Appendix A

Brief historical review terrorism

Rapoport's four wave characterisation of the evolution of modern terrorism was introduced in section 2.1. The three waves preceding the religious wave that concerns this study, are briefly discussed to provide a historical and political context to terrorism.

A.1 Anarchist wave (1880-1920)

The Anarchist wave started in Russia in the late 1880s, which is also seen as the beginning of modern terrorism. Russian revolutionaries propagated that terrorism, a title used proudly, would spark off a revolution[37]. Assassinating goverment officials was seen as the quickest and most effective means to destroy conventions. While terrorism used to be violence by governments that tried to impose their radical new order, it now became associated with non-governmental groups as it still is today.

Within a decade after emergence, this form of terror appeared in Europe, Asia and with it, the first global terrorism episode in history emerged. The spread of revolutionary- and anarchist ideas depended on a number of factors, the main one being the industrialisation and urbanisation [125]. There was a transformation in communication and transportation patterns: telegraph, daily mass newwpapers and railroads prospered the end of the nineteenth century. People moved from countryside to dense cities, living in urban slums where ideas were easily spread and living- and working conditions were inhumane. There, the message of revolution ending classes and the capitalist system caught on. The anarchist wave lasted for about fourty years, ending with the onset of World War I.

A.2 Anticolonial wave (1920-1960)

The second wave was not triggered by a domestic political situation, but rather by the Treaty of Versailles that brought World War I to an end in 1919. After the war, European colonies as well as a number of European ethnicities, felt that political independence was within reach [87]. Furthermore, American president Woodrom Wilson introduced fourteen points in January 1918, presented as a foundation for global peace and prosperity [141]. The points were central during the negotiations for peace. The fifth was the concept of national self-determination, which gave people the right to decide about their sovereignty. Incorporated in the Treaty of Versailles, the victors of the war could use this principle to divide the empires of the defeated states (mainly in Europe). This led to nationalist groups within their own empires, that arose to protest against the colonial powers. The second wave's climax was twenty-five years later than the Treaty of Versailles, which Rapoport explains by stating that World War II reinforced and enlarged the implications of Versailles.

The organisations called themselves freedom fighters, resisting against government terror, because their anticolonial cause seemed more legitimate than those of the anarchists. Vice versa, the governments called the organisations terrorists. As Rapoport remarks, this new form of anticolonial terrorism was crucial in establishing new states such as Ireland, Israel, Cyprus and Algeria.

A.3 New Left wave (1960-1990)

The third wave emerged in the 1960s, stimulated by the Vietnam war (1955-1975) that convinced people that the current capitalistic system was vulnerable. In the West and Global South, groups arose consisting of youth aiming for a revolution to break down the surpressing structures of imperialism, capitalism and facism [41]. Driven by the advancement in communications technology, opposition to the war and to the United States (be it anti-Americanism), became a global phenomenon [87]. One third of this wave's international attacks involved U.S. American targets.

Rapoport notes that after the Vietnam War ended in 1975, the Palestine Liberation Organisation (PLO) replaced the Viet Cong as the face of third wave terrorism. The PLO, dedicated to restoration of the Palestinian homeland, received strong support of the Arab states and the Soviet Union. This combined with the training facilities in Lebanon they made available to the Palestinians and other groups, gave them an even more central role.

Terrorism meant extreme forms of nationalism and radicalism again, as with the Anarchist wave. Examples of emerging groups are the Red Brigades - founded in Italy 1970 to overthrow the government and replace it with a communist system -, the Red Army Fraction (RAF) - an anti-capitalist, anti-imperialist, West German terrorist organization active in the 80s - and the Japanese Red Army - which was a communist militant group active between 1971 and 2001, with the goals to overthrow the Japanese goverment and to start a world revolution. The Soviet Union supported the organisations, facilitating training and weapons.

Another similarity with the Anarchist wave was the reappearance of what Rapoport calls "theatrical targets", this time in the form of plane hijackings, kidnappings and the assassination of prominent figures. The abandoned term "international terrorism" was revived. Groups bonded, which intensified when international training facilities arose. Targets had international character and some groups were more active abroad than in their country of origin.

Inseparable with this wave, was a wave of right-wing violence. In several countries in South America and in Europe, right-wing groups staged attacks as if coming from the left, to legitimise the shift to the right [41].

In the 1980s, the New Left groups were getting defeated and the third wave saw its decline. International cooperation to fight terrorism became increasingly effective and Israel's invasion of Lebanon in 1982 eliminated PLO's training facilities.

Appendix B

Secondary country statistics

Besides the five country statistics implemented as covariates, this study considered various other statistics that are introduced in this appendix.

B.1 Population Growth

The United Nations Population Division calculates annual population estimates that are available from 1960 until present. The estimates are based on censuses, surveys, populations registers and other sources [29]. For countries in less developed regions, demographic information might be limited or unreliable. This is illustrated by the fact that the most recent data available for Somalia was 1975 and for Pakistan 1998. Missing data is interpolated (extrapolated) based on demographic models [19]. Errors might be substantial, because of limitations in resources to perform a full census and because of both conscious and unconscious government shortcomings. In addition, the comparability between countries is limited because of different collection procedures and estimation methods by the agencies that collect the data.

The population estimates are shown in figure B.1. Increasing population is expected to substantially increase terrorist activity.

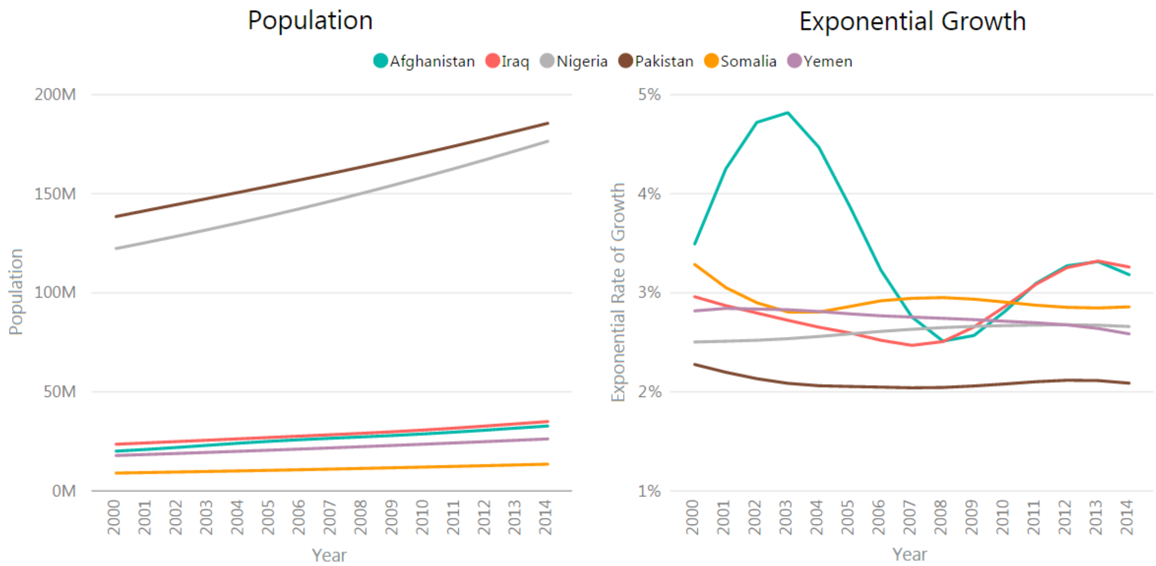

Fig. B.1 Annual population growth in percentage

B.2 Urbanisation

Another demographic statistic estimated by the United Nations Population Division is urbanisation, defined as the percentage of people living in urban areas [26]. The previously discussed population estimates and urban ratios from the United Nations World Urbanization Prospects are used to calculate the urbanisation rate. Therefore, the uncertainties of the population estimates causes the urbanisation data to be at least as unreliable.

Most countries experience significant increases of urbarnisation rates, see figure B.2. Iraq does not see this strong increase, but deals with a general urbanisation rate of almost 70%. Urbanisation was predicted to exacerbate terrorism, but because of unreliable data, results should be interpreted with caution.

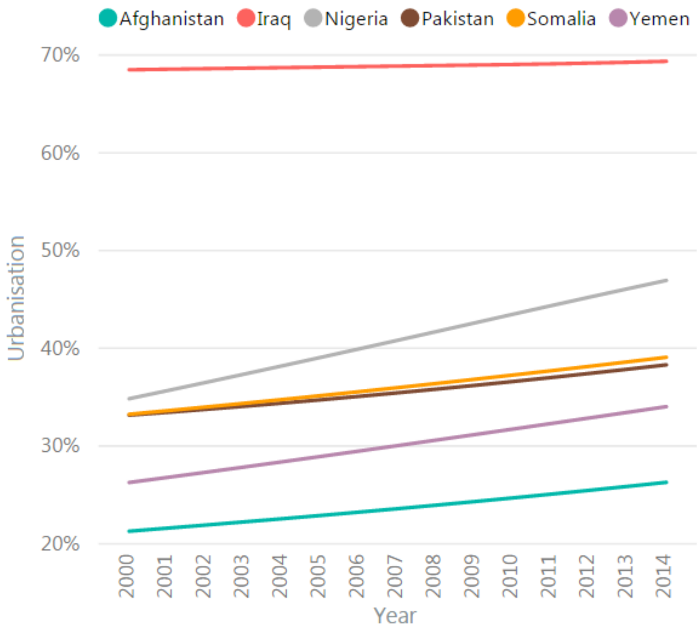

Fig. B.2 Annual urbanisation as percentage of total population

B.3 Precipitation & Temperature

The Climate Research Unit of the University of East Anglia processes climate data collected by thousands of weather stations across the world. Figure B.3 shows the annual precipitation and temperatures. Nigeria shows a decreasing trend that tends towards an annual average of 1000 mm. The other countries have climates that are much dryer by nature. Iraq experienced dramatic low precipitation in 2008. Precipitations is expected to be negatively correlated with terrorist activity.

Section 2.3.3.3 did not specifically link rising temperatures to terrorism, temperature is however included to see if it can capture a more general influence of climate change. Apart from normal fluctuations, Iraq experiences a record high average temperature in 2010. Rising temperature is expected to increase the amount of terrorism attacks.

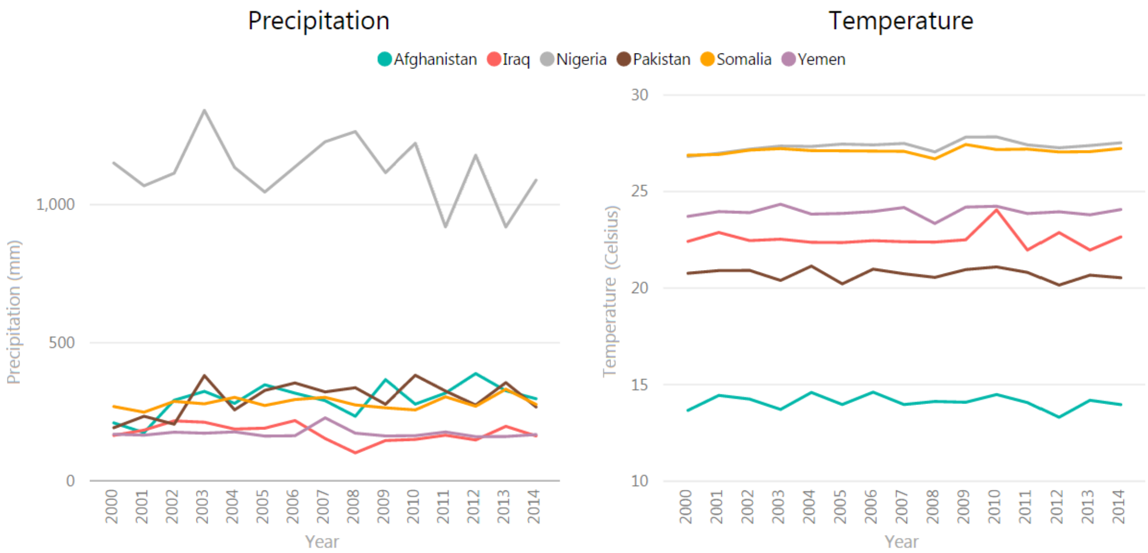

Fig. B.3 Annual precipitation and temperatures

B.4 Gross Domestic Product

The World Bank defines the gross domestic product (GDP) as "...the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output" [19]. Ideally, GDP per capita is used, but this would involve the uncertainty of the population statistics as mentioned in section B.1. Therefore annual GDP at constant 2010 prices in US Dollars is incorporated, see figure B.4.

All countries experience growth, apart from two major decreases, Iraq in 2003 and Yemen in 2011. Considering the correspnding increases in terrorist activity, GDP is more likely to represent rising income inequality than (the reduction of) poverty. In this light, a rise in GDP is expected to increase the number of terrorist attacks. However, GDP per capita makes use of population statistics, therefore the unreliability is inherited and results should be interpreted with caution.

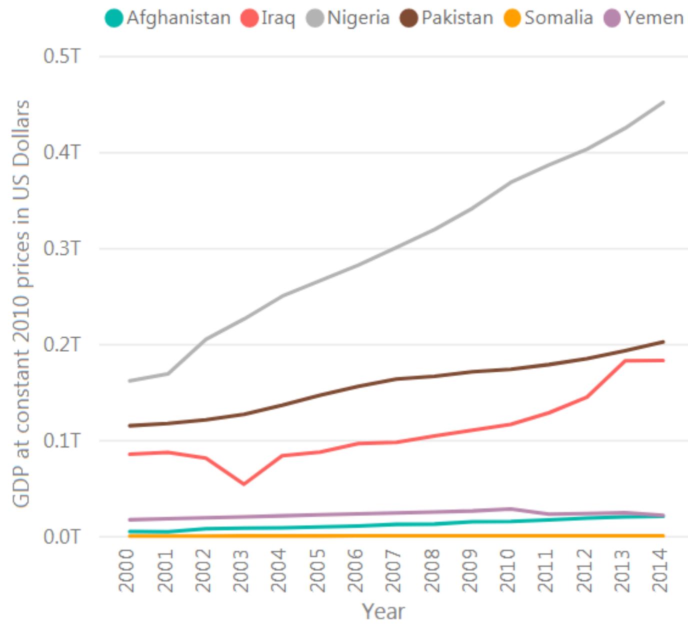

Fig. B.4 Annual gross domestic product at constant 2010 prices in US Dollars

B.5 Human Development Index

The Human Development Index (HDI) was created by the UN as a metric to assess the average achievement in three basic dimensions of human development: health, education and income [3]. Since 1993, the index is annually published in the Human Devlopment Report of the UNDR. However, data agencies continually revise their data and country coverages occasionally change, possibly affecting the HDI ranking, causing year-to-year changes of the HDI values for different editions of the report [1]. Therefore UNDP stresses that analyses should not be based on data from different editions of the report.

Fortunately, each report contains recalculations of the HDI scores of former years. The Human Development Report 2015 provides the index for 1990, 2000 and 2010-2014, as shown in figure B.5. Data for Somalia can again be replaced by the average of Djibouti and Ethiopia. Data between 2000 and 2010 can be linearly interpolated. This leads to a general linear increasing trend from 2000 through 2014, without fluctuations that could be helpful to explain differences with the general increasing trend in terrorist activity. In addition, the scarce data makes it hard to interpret results. The HDI is therefore not further considered in this study.

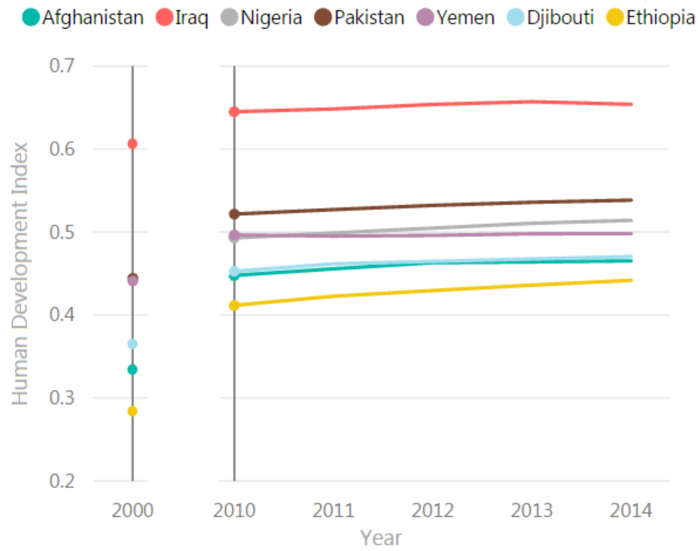

Fig. B.5 Human Development Index

Appendix C

Correlation plots

The six figures that follow are used for a general review of the relationships between terrorist activity and annual country statistics, as well as to select covariate sets to be implemented in the model. In the upper triangular, Spearman's rank correlation coefficient, possibly tie corrected, and the corresponding statistic significance are given. For a general overview, different levels of correlation are represented by different colors (dark blue strong positive relationship, dark purple strong negative relationship) and the significance is indicated by the common asterisks. The lower triangular contains the corresponding scatter plots, including linear model fits and 95% confidence intervals. For Pakistan and Somalia, the correlations regarding civil liberties are absent because the index is constant over time.

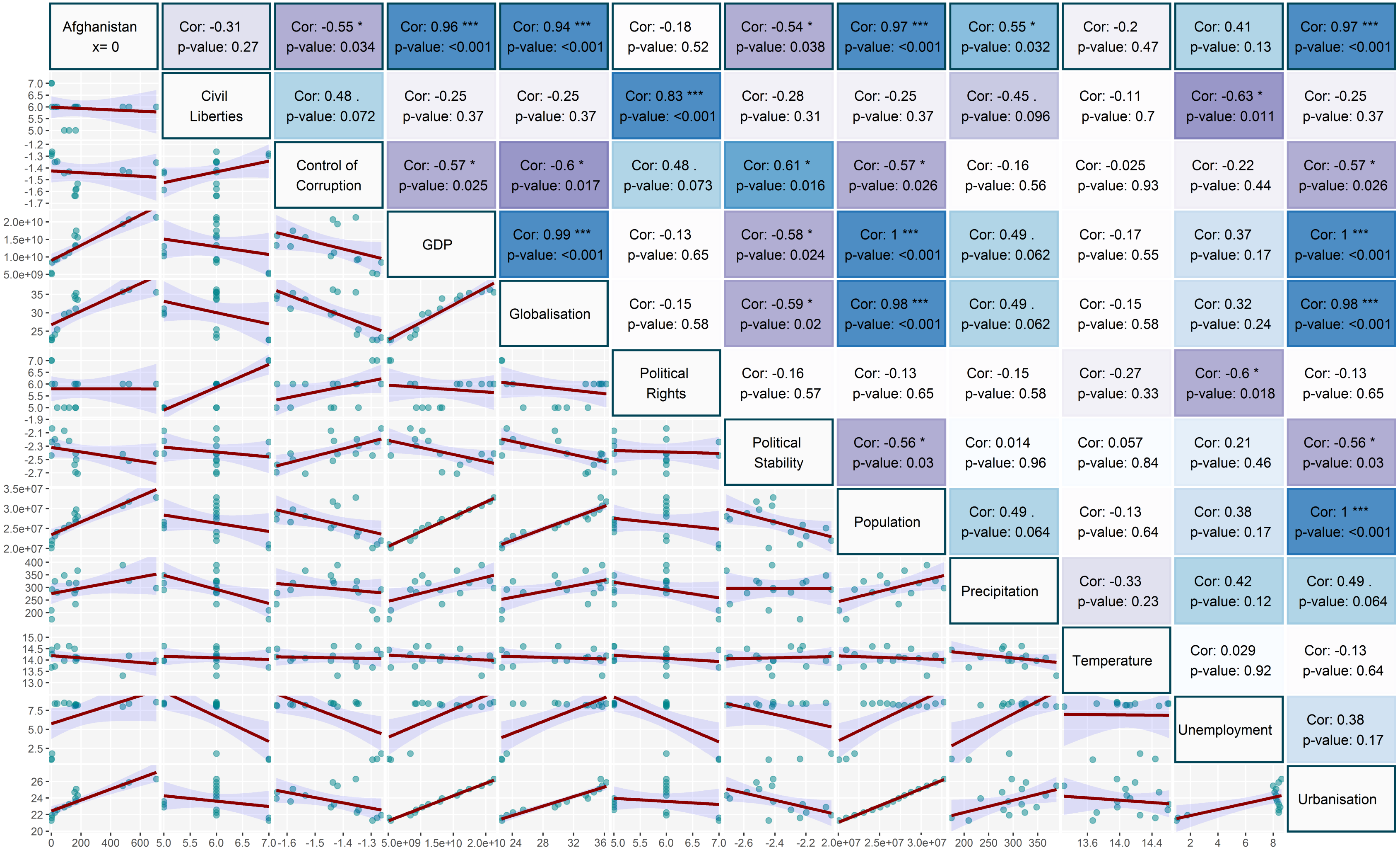

Fig. C.1 Correlation plot Afghanistan

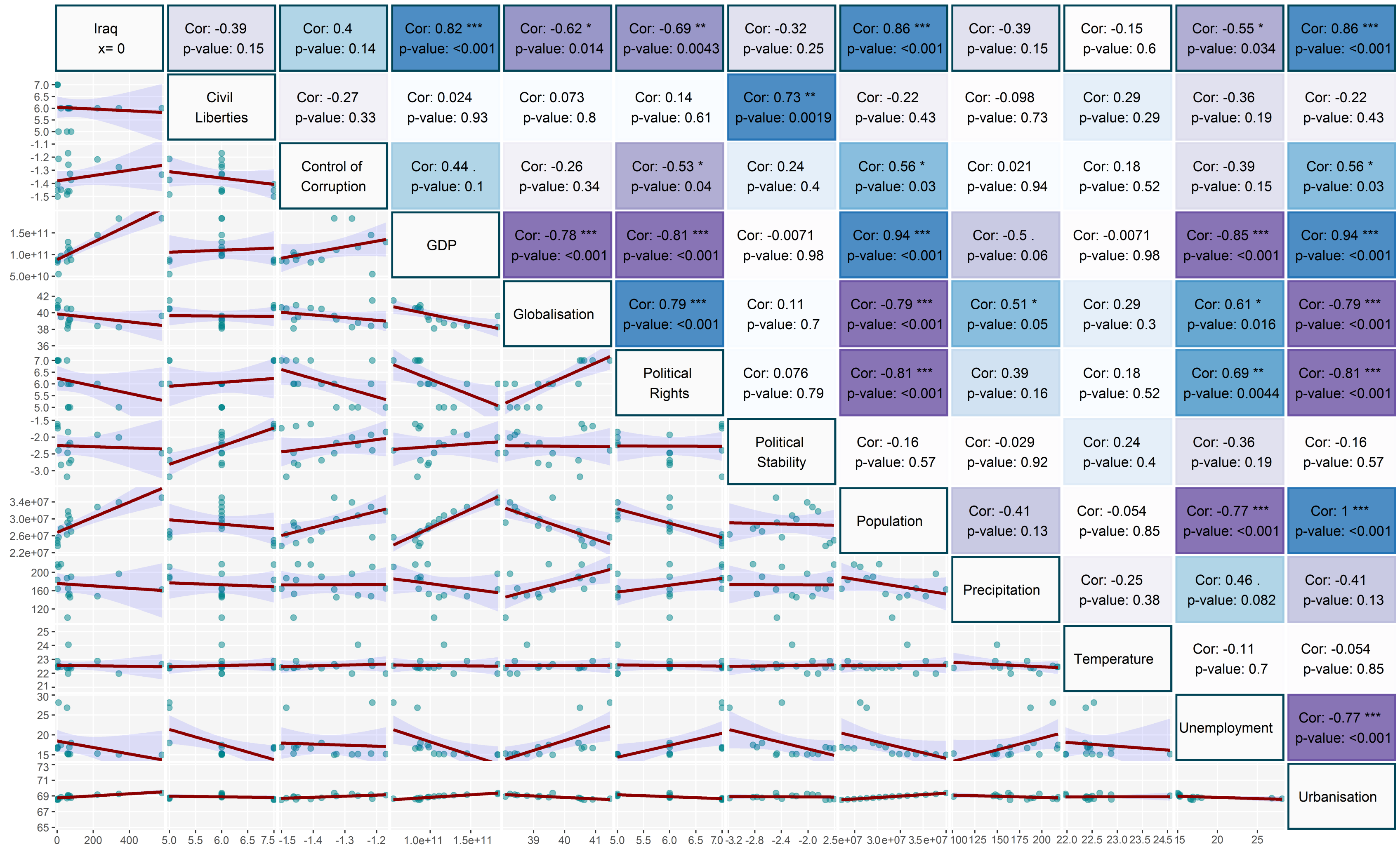

Fig. C.2 Correlation plot Iraq

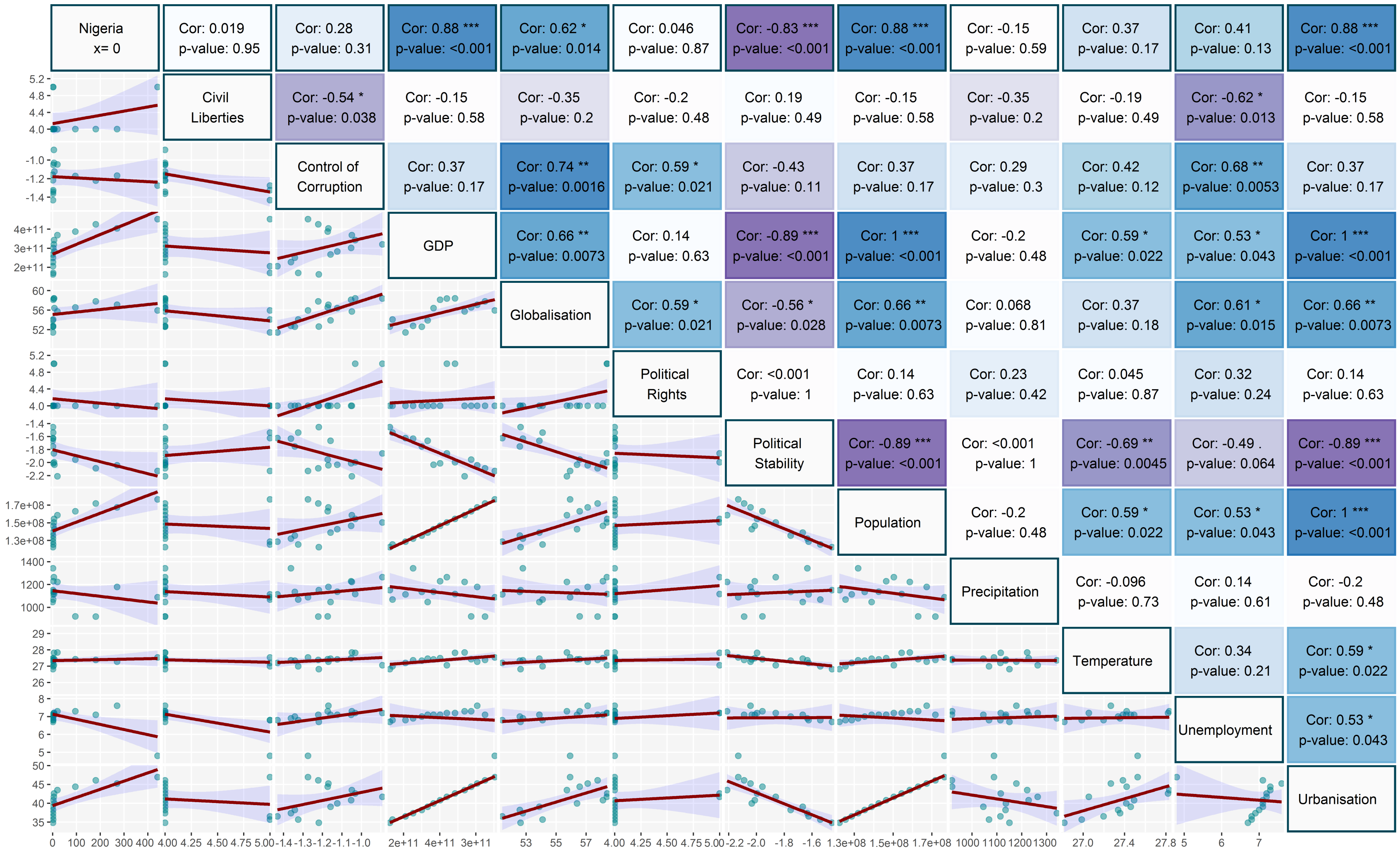

Fig. C.3 Correlation plot Nigeria

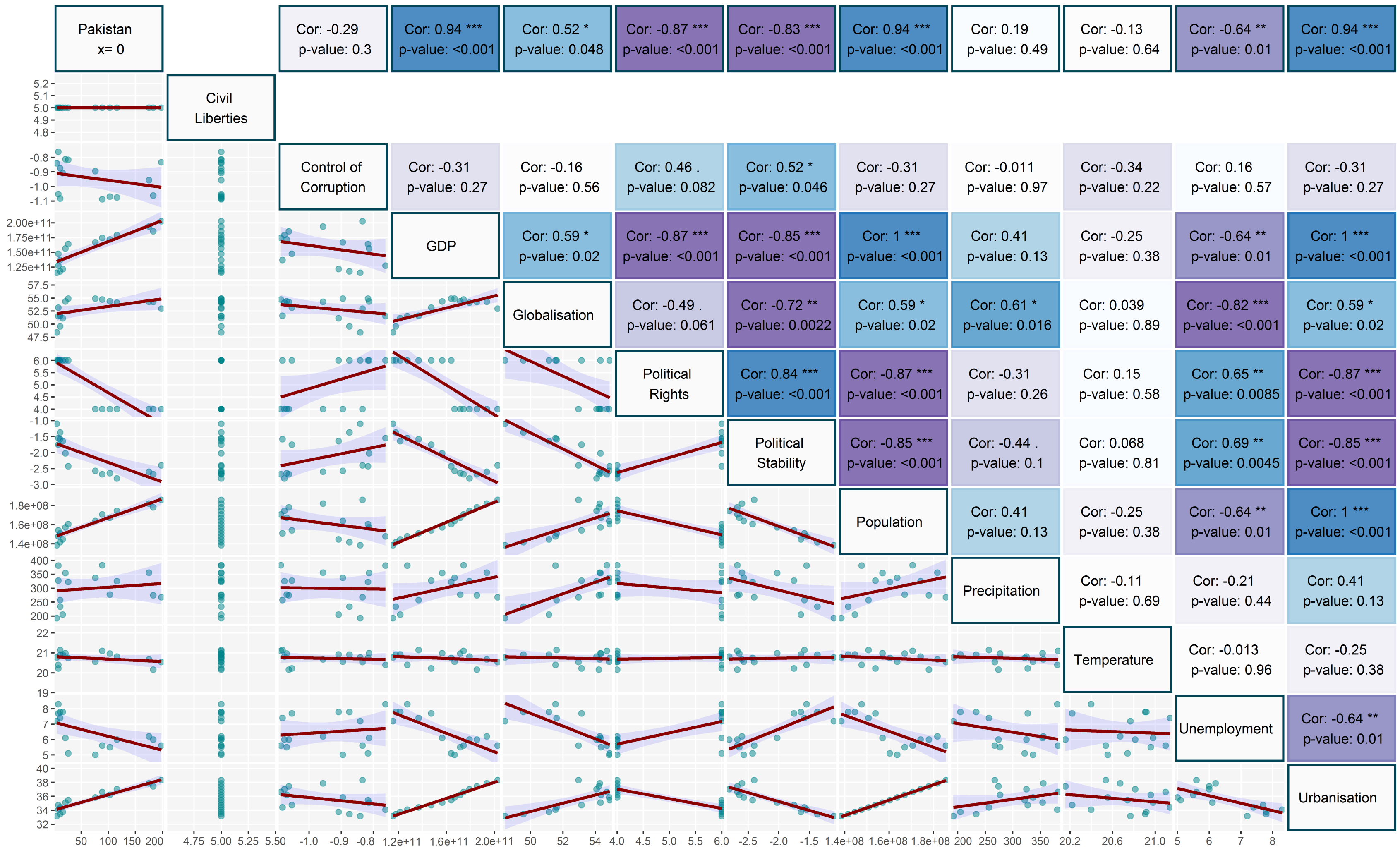

Fig. C.4 Correlation plot Pakistan

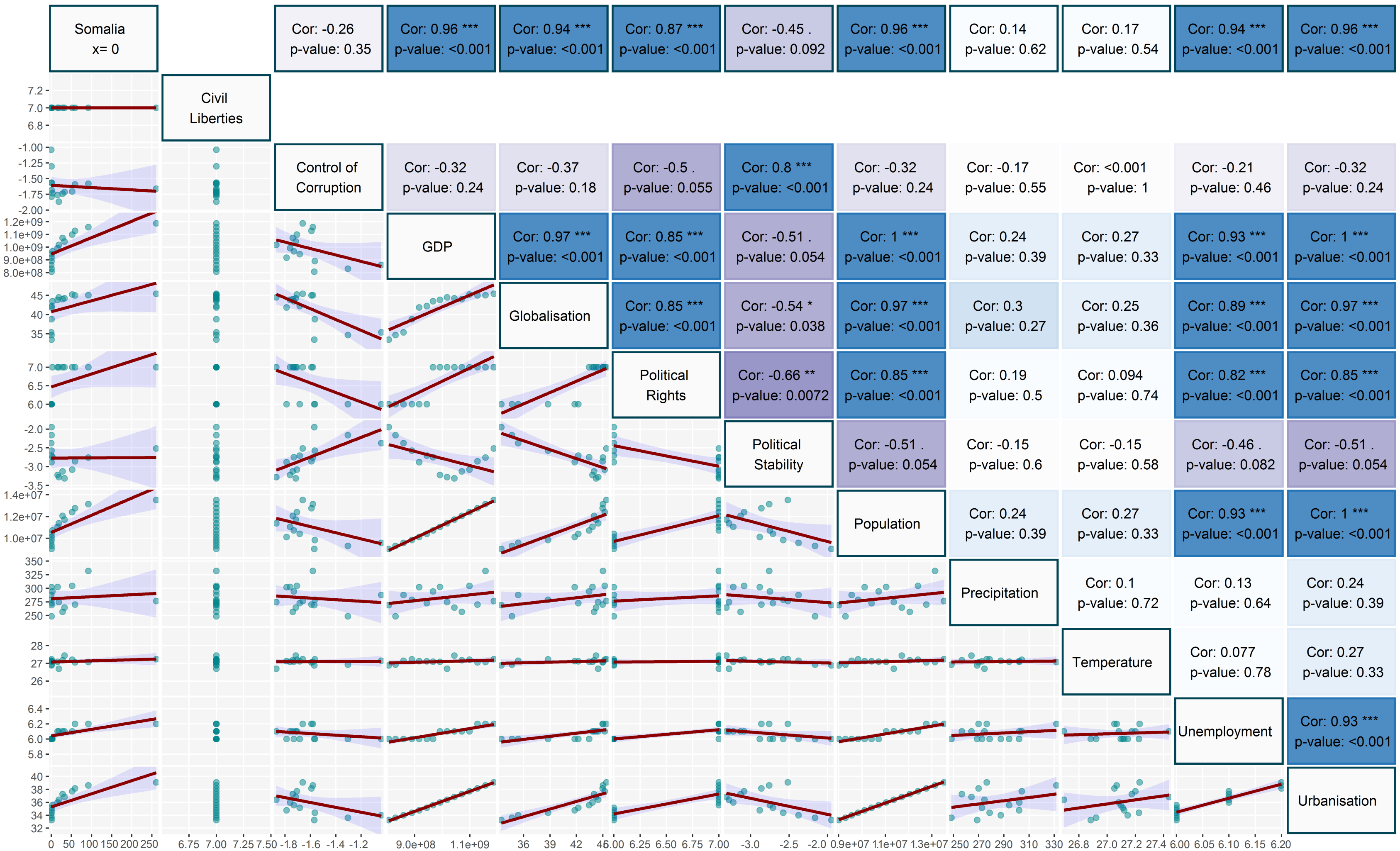

Fig. C.5 Correlation plot Somalia

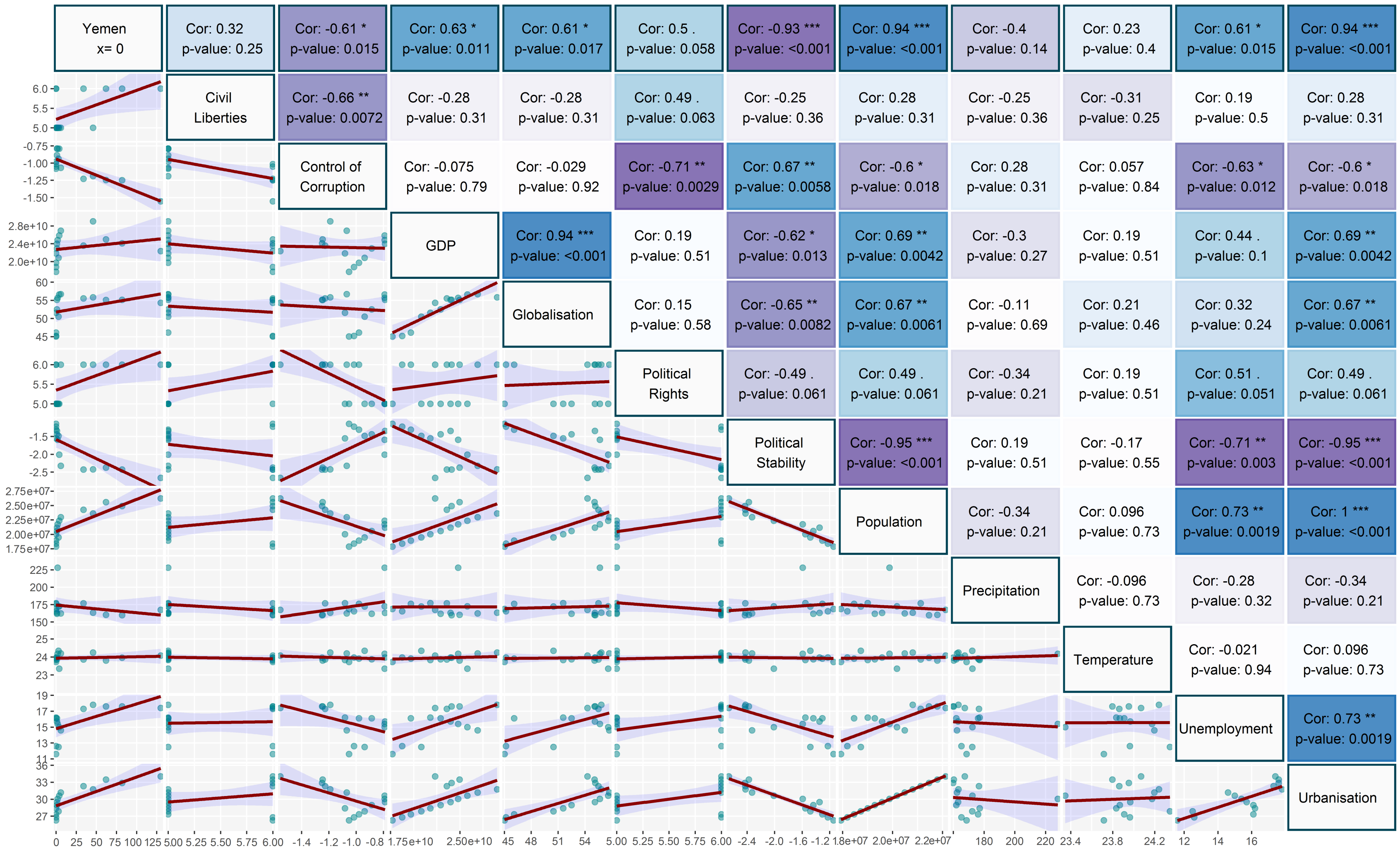

Fig. C.6 Correlation plot Yemen

Appendix D

Results Pakistan and Yemen

D.1 Pakistan

Problematic for Pakistan, is that the extension to the GTD 06/17 version reveals a general change in trend, as illustrated in figure D.1. This hits one of the biggest limitations of our model; it will only work as long as there is a general increasing trend.

The supremum Wald test by Andrews [67] can be used to assess the structural break. The Weibull baseline trend is then used before the break and a distinct, to be determined trend afterwards. Using two different baseline models does come with the problem of overfitting. Furthermore, the data to determine the second baseline trend is limited. Still, the main reason not to implement a structural break model, is the fact that the in-sample fit of the parsimonious model for period 2000-2012 is not promising.

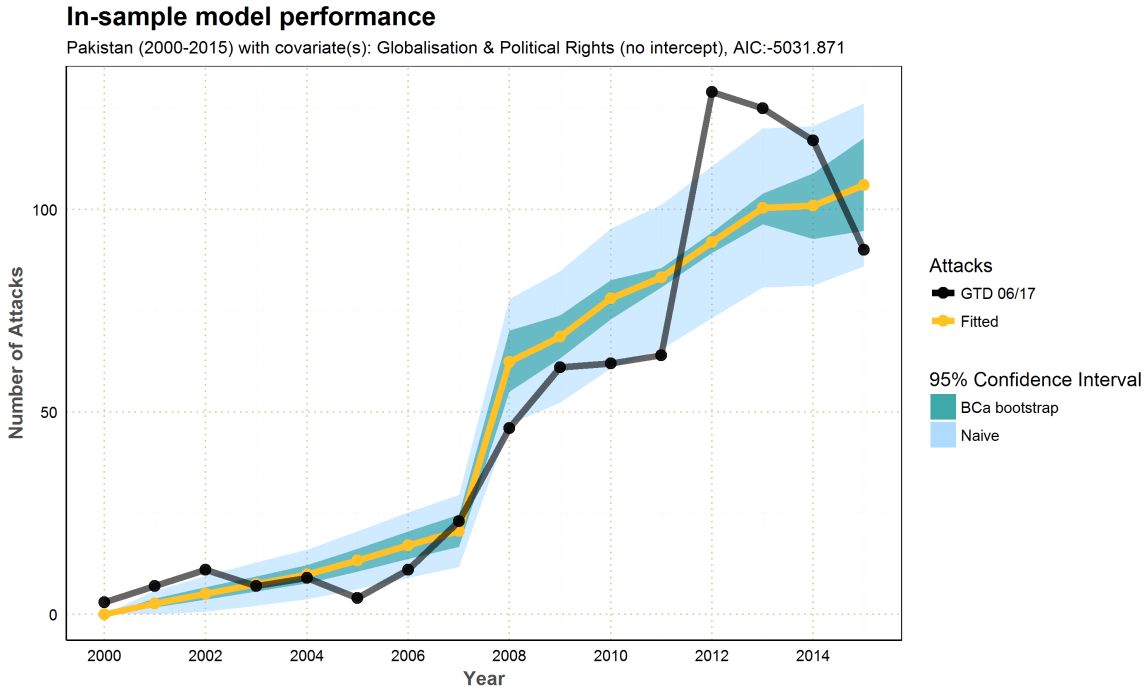

Fig. D.1 In-sample fit Pakistan (2000-2016), including the naive and BCa 2500 sample bootstrap 95% confidence intervals. The extension of the GTD reveals a general change in trend.

Model selection for the period 2000-2012, results in a model with political rights and globalisation as covariates, without an intercept, see table D.1. While the covariates do improve the baseline fit (AIC of -2580.08 opposed to an AIC of -2551.49), the in-sample fit exhibits a mean absolute error of 10 attacks with a corresponding mean absolute percentage error of 49.74%. Figure D.2 shows that the covariates do not capture the characteristic shifts in terrorist activity, such as the decrease of attacks in 2003 and 2005, as well as the strong increase in 2012. This leads to conclude that the covariates are not able to explain the dynamics of terrorism in Pakistan.

Parameter	Estimate	Standard error	z-value	p-value
δ	1.928	0.227	8.459	$< 10^{-16}$
Globalisation	1.608	0.456	3.529	$4.168 * 10^{-04}$
Political rights	-0.5117	0.0962	-5.321	$1.031 * 10^{-07}$

Table D.1 Parameter estimates Pakistan, training data 2000-2012

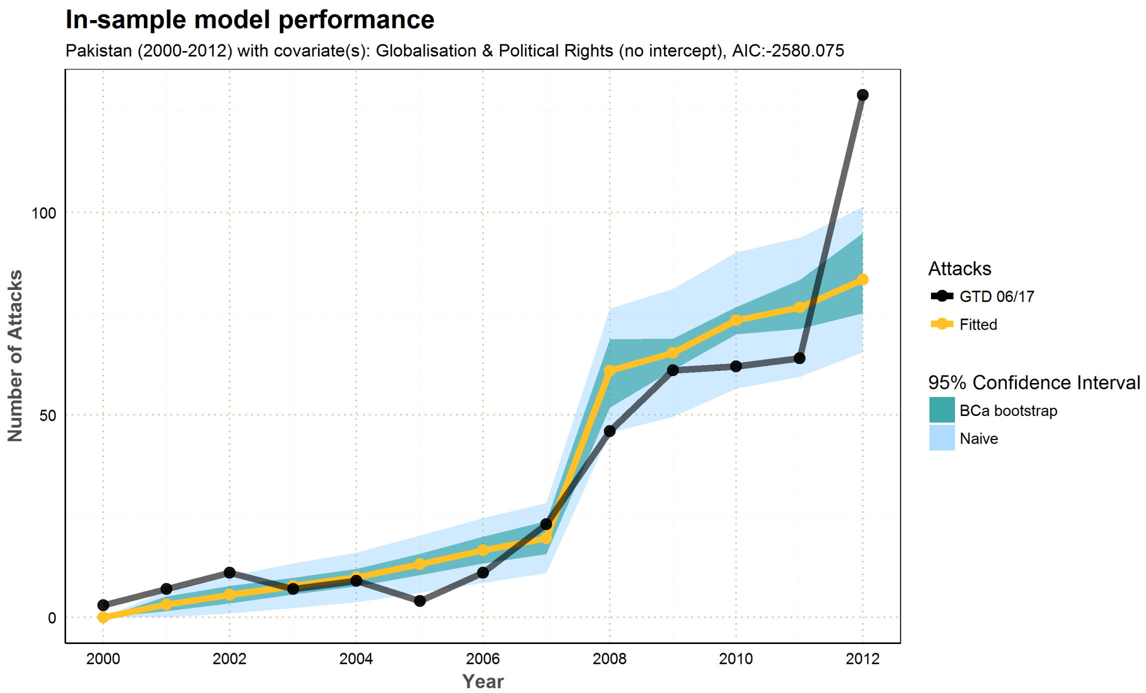

Fig. D.2 In-sample fit Pakistan (2000-2012), including the naive and BCa 2500 sample bootstrap 95% confidence intervals.

D.2 Yemen

The parsimonious model for Yemen does not include an intercept and has globalisation as covariate. The globalisation index is again scaled by dividing the index by 25. Maximum likelihood estimation for the period 2000-2015 results in a log-likelihood of 1085.98 and corresponding AIC of -2165.99, along with the parameter estimates from table D.2.

Parameter	Estimate	Standard error	z-value	p-value
δ	5.703	0.248	23.005	$< 10^{-16}$
Globalisation	-4.595	0.321	-14.304	$< 10^{-16}$

Table D.2 Parameter estimates Yemen, training data 2000-2015

The model fit is shown in figure D.3. The level shifts in 2014 and 2015 are captured very well, but the dynamics between 2010 and 2013 are completely absent. Since the first attack in Yemen occurred in 2004 and terrorism only truly emerged after 2008, there is limited data available to train and test the model. Out-of-sample validation and forecasting are therefore not pursued.

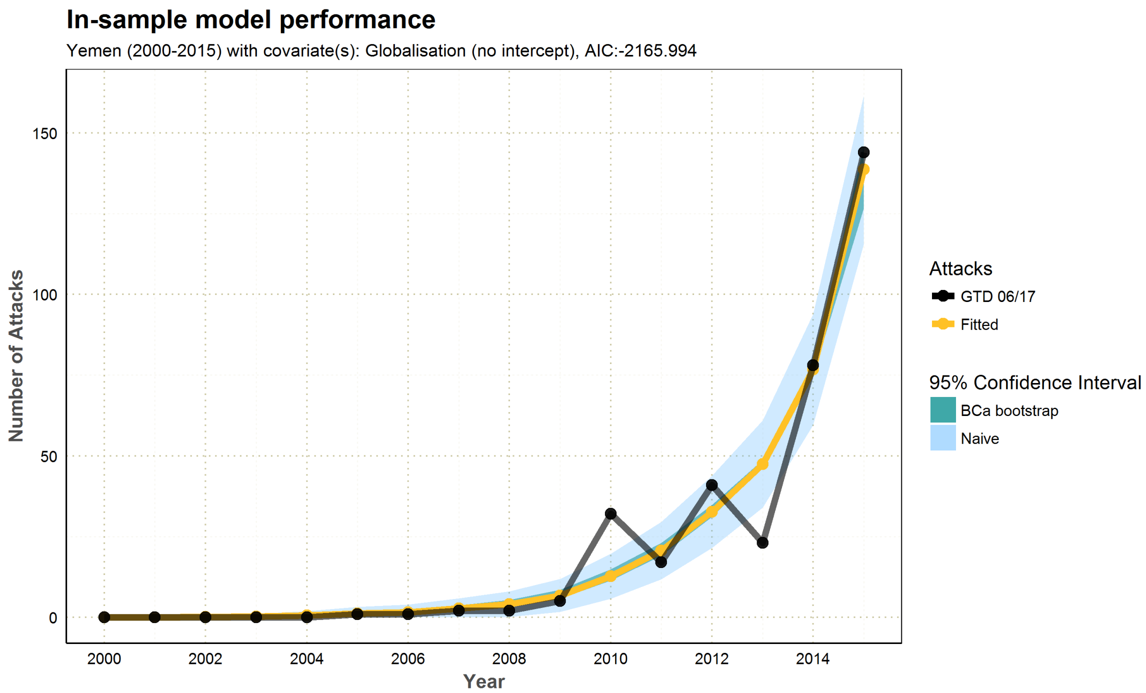

Fig. D.3 In-sample fit Yemen (2000-2015), including the naive and bias-corrected accelerated (BCa) bootstrap 95% confidence intervals.

References

[1] Briefing note for countries on the 2016 Human Development Report . http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/IND.pdf.

[2] Globalisation and the state: Challenges for economic growth and human development. Technical report, United Nations Department of Economic and Social Affairs, New York.

[3] Human Development Index (HDI). http://hdr.undp.org/en/content/human-development-index-hdi.

[4] Resolation A/RES/40/61. http://www.un.org/documents/ga/res/40/a40r061.htm, December 1985.

[5] Human development report 1997. Technical report, United Nations Development Programme, Oxford, 1997.

[6] The Arab Convention for the Suppression of Terrorism. page 2, April 1998.

[7] A fair globalization: Creating opportunities for all. Technical report, The World Commission on the Social Dimension of Globalization, International Labour Office, Geneva, Switzerland, February 2004.

[8] A global alliance against forced labour. Technical report, International Labour Office, Geneva, Switzerland, 2005.

[9] Globalization and inequality. Technical report, World Economic Outlook, International Monetary Fund, Washington D.C., U.S.A., October 2007.

[10] CESifo DICE Report 4/2011 (Winter). page 67, April 2011.

[11] 2014 climate change adaptation roadmap. Technical report, Department of Defense, United States of America, October 2014.

[12] Isis rebels declare 'islamic state' in iraq and syria. BBC News, http://www.bbc.com/news/world-middle-east-28082962, 2014.

[13] World Population Sheet 2014. http://www.prb.org/pdf14/2014-world-population-data-sheet_eng.pdf, 2014.

[14] The world's 10 richest terrorist organisations. *Forbes International*, December 2014.

[15] Codebook, Inclusion Criteria and Variables). pages 3–53, June 2015.

[16] Country reports on terrorism 2014. Technical report, United States Department of State, Bureau of Counterterrorism, July 2015.

[17] Global Employment Trends for Youth 2015. Technical report, International Labour Organization, Geneva, Switzerland, 2015.

[18] Universal Declaration of Human Rights. Technical report, United Nations, 2015.

[19] World Development Indicators. 2015.

[20] Insurgency, terrorism and organised crime in a warming climate. Technical report, Federal Foreign Office, Bundesrepublik Deutschland and Adelphi, Berlin, October 2016.

[21] Freedom in the World 2018. https://freedomhouse.org/report/methodology-freedom-world-2018, 2017.

[22] Iraq's public administrative issues: Corruption. *Review of Public Administration and Management, Vol. 5 Issue 2*, page 3, 2017.

[23] Political Stability and Absence of Violence/Terrorism. http://info.worldbank.org/governance/wgi/pdf/pv.pdf, 2017.

[24] Terrorism, corruption and the criminal exploitation of natural resources. Technical report, OECD, October 2017.

[25] Unemployment, total (% of total labor force) (modeled ILO estimate). https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=AF, 2017.

[26] Urban population (% of total). https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=AF, 2017.

[27] World drug report 2017, market analysis of plant-based drugs. Technical report, United Nations Office on Drugs and Crime, 2017.

[28] World drug report 2017, the drug problem and organisation crime, illicit financial flows, corruption and terrorism. Technical report, United Nations Office on Drugs and Crime, 2017.

[29] World population prospects, the 2017 revision, methodology of the united nations population estimate and projections. page 3, 2017.

[30] Adabie A. Poverty, Political Freedom, and the Roots of Terrorism. page 2, October 2004.

[31] Agresti A. *Foundations of Linear and Generalized Linear Models*. Wiley, 2015.

[32] Azzam A. Defence of the muslim lands, the first obligation after iman. https://www.slideshare.net/baadshahm/defence-of-the-muslim-lands, 1979.

[33] Clauset A. and Woodard R. Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics, Vol. 7, No. 4*, pages 1838–1865, 2013.

[34] MacEwan A. Globalization and Stagnation. *Monthly Review, Vol. 45*, pages 1–16, April 1994.

[35] MacEwan A. Lake Chad, almost gone. *Vital Water Graphics, An Overview of the State of the World's Fresh and Marine Waters*, pages 1–16, 2016.

[36] Rabasa A. *Political Islam in Southeast Asia: Moderates, Radical and Terrorists*. Oxford University Press, London, May 2003.

[37] Roberts A. The Changing Faces of Terrorism. http://www.bbc.co.uk/history/recent/sept_11/changing_faces_02.shtml, August 2002.

[38] Salkida A. Africa's vanishing lake chad action needed to counter an "ecological catastrophe". http://www.un.org/africarenewal/magazine/april-2012/africa%E2%80%99s-vanishing-lake-chad, April 2012.

[39] Krueger A.B. and Maleckova J. Education, Poverty and Terrorism: Is There a Causal Connection? *Journal of Economic Perspectives, Vol. 17, No. 4*, pages 119–144, 2003.

[40] Al Bassam A.M. Urbanisation and Migration in Saudi Arabia: The Case of Buraydah City. page 10, December 2011.

[41] de Graaf B. *Theater van de angst*. Boom, Amsterdam, 2010.

[42] de Graaf B. Terrorisme. DWDD University presenteert, 2016.

[43] Hoffman B. "Holy terror": The implications of terrorism motivated by a religious imperative. *Studies in Conflict & Terrorism*, pages 271–284, 1995.

[44] Hoffman B. *Inside Terrorism*. Colombia University Press, 2006.

[45] Norton B. Saudi arabia funds and exports islamic extremism: The truth behind the toxic u.s. relationship with the theocratic monarchy. http://www.salon.com/2016/01/06/saudi_arabia_funds_and_exports_islamic_extremism_the_truth_behind_the_toxic_u_s_relationship_with_the_theocratic_nation/, 2016.

[46] World Bank. Control of Corruption. https://info.worldbank.org/governance/wgi/pdf/cc.pdf.

[47] Tai B.C. and Machin D. *Poisson Regression*, chapter 5. Regression Methods for Medical Research. Wiley, 2013.

[48] Lindqvist B.H. and Rannestad BB. Monte Carlo Exact Goodness-of-fit Tests for Nonhomogeneous Poisson Processes. June 2011.

[49] Lutz B.J. and Lutz J.M. Globalisation and Terrorism in the Middle East. *Perspectives on Terrorism, Vol.9, No. 5*, 2015.

[50] Jenkins B.M. The Study of Terrorism: Definitional Problems. page 1, 1980.

[51] Barber B.R. *Jihad vs. McWorld, Terrorism's Challenge to Democracy*. Ballentine Books, New York, 1995.

[52] P. Bühlmann and M. Mächler. Computational Statistics. July 2014.

[53] Chen C. and Liu L. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association, Vol. 88, No. 421*, pages 284–297, March 1993.

[54] Friedersdorf C. How u.s. weapons fall into the hands of terrorists. https://www.theatlantic.com/international/archive/2014/11/how-us-weapons-fall-into-the-hands-of-terrorists/383095/, 2014.

[55] Kelley C., Mohtadi S., Cane M., Seager R., and Kushnir Y. Climate change in the fertile crescent and implications of the recent syrian drought. 112:3241–3246, March 2015.

[56] Kwaja C. Nigeria's pernicious drivers of ethno-religious conflict. *Africa Security Brief, No. 14*, page 4, July 2011.

[57] Lehalle C. and Laruelle S. *Market Microstructure in Practice* . World Scientific Publishing, 2014.

[58] Richardson C. Relative Deprivation Theory in Terrorism: A Study of Higher Education and Unemployment as Predictors of Terrorism. April 2011.

[59] United Nations Security Council. Resolution 1566 (2004). page 2, 2004.

[60] Collett D. *Applied Life Data Analysis*. John Wiley & Sons, 1982.

[61] Eggen D. and Wilson S. Suicide bombs potent tools of terrorists. https://www.washingtonpost.com/archive/politics/2005/07/17/suicide-bombs-potent-tools-of-terrorists/e11ed483-9936-45c0-b6c6-2653d4519ff5/?utm_term=.bb781a65079b, 2005.

[62] Ehrenfeld D. Globalisation: Effects on Biodiversity, Environment and Society. *Conservation and Society, Vol. 1, Issue 1*, pages 100–111, 2003.

[63] Koopman D. Internship haensel ams. page 5, 12 2015.

[64] Sterman D. Don't Dismiss Poverty's Role in Terrorism Yet. http://time.com/3694305/poverty-terrorism/, 2015.

[65] Rapoport David C. The four waves of modern terrorism. Published in 'Attacking terrorism: elements of a grand strategy' page 46-73, by Audrey Kurth Cronin, James M. Ludes, Georgetown University, 2004.

[66] Cox D.R. and Lewis P.A.W. *The Statistical Analysis of Series of Events*. Springer, 1966.

[67] Andrews D.W.K. *Tests for parameter instability and structural change with unknown change point*. July 1993.

[68] Cinlar E. *Introduction to Stochastic Processes*. Dover, 2013.

[69] The Economist. A three-headed hydra. https://www.economist.com/blogs/freeexchange/2014/07/measuring-inequality, July 2014.

[70] Girosi F. and King G. *Demographic forecasting* . Princeton University Press, 2008.

[71] Roudi F. Youth Population and Employment in the Middle East and North Africa: Opportunity or Challenge? July 2011.

[72] Zadeh F.E. *Islam versum Terrorism: An Intimate View of Understanding Islam and the Culture of the Middle East*. Twin Lakes Pub, 2003.

[73] Flannery F.L. *Understanding Apocalytic Terrorism*. Routledge, 2016.

[74] National Consortium for the Study of Terrorism and Responses to Terrorism. History of the GTD. https://www.start.umd.edu/gtd/about/History.aspx.

[75] Takahara G. Exponential Distribution and Poisson Processes. page 199, 2017.

[76] GTD. Gtd including palestinan resistance acts. http://www.start.umd.edu/gtd/search/Results.aspx?search=israel&sa.x=0&sa.y=0.

[77] Long H. Inequality caused rise of ISIS, Piketty says . http://money.cnn.com/2015/12/03/news/economy/isis-money-thomas-piketty/index.html, December 2015.

[78] Cooper H.H.A. Terrorism: The Problem of the Problem of Definition. *Chitty's Law Journal, 26*, page 106, 1978.

[79] Challenges High-level Panel on Threats and Change. A more secure world: our shared responsibility. page 52, 2004.

[80] Freedom House. Iraq. https://freedomhouse.org/report/freedom-world/2003/iraq, 2003.

[81] Gregg H.S. Defining and Distinguishing Secular and Religious Terrorism. *Perspectives on Terrorism, Vol. 8*, pages 1–16, 2014.

[82] Chang I., Tiao G.C., and Chen C. Estimation of time series parameters in the presence of outliers. *Technometrics, Vol. 30, No. 2*, pages 193–204, May 1988.

[83] Saha I. and Paul B. *Essentials of Biostatistics*. Academic Publisher, April 2010.

[84] US Foreign Policy in Perspective. Soviet union (target: Afghanistan) 1979-91. http://www.us-foreign-policy-perspective.org/index.php?id=322, 2009.

[85] Amnesty International. The Arab Convention for the Suppression of Terrorism: a serious threat to human rights. page 6, 2002.

[86] Amnesty International. European union directive on counterterrorism is seriously flawed. https://www.amnesty.nl/actueel/european-union-directive-counterterrorism-seriously-flawed, 2016.

[87] Kaplan J. Waves of Political Terrorism. http://politics.oxfordre.com/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-24, November 2016.

[88] Ledolter J. The effect of additive outliers on the forecasts from arima models. *International Journal of Forecasting, Vol. 5, Issue 2*, pages 231–240, 1989.

[89] Baker J.A. and Hamilton L.H. The Iraq Study Group Report. Technical report, Iraq Study Group, 2006.

[90] Piazza J.A. Is Islamist Terrorism More Dangerous?: An Empirical Study of Group Ideology, Organization, and Goal Structure. *Terrorism and Political Violence, 21*, pages 62–88, 2009.

[91] Lawless J.F. Regression Method for Poisson Process Data. *Journal of the American Statistical Association, Vol. 82, No. 399*, pages 808–815, September 1987.

[92] Lee J.H., Han G., Fulp W.J., and Giuliano A.R. Analysis of overdispersed count data: application to the Human Papillomavirus Infection in Men (HIM) Study. *Epidemiology and Infection, Vol. 140*, pages 1087–1094, August 2011.

[93] Consul J.I. Flexible Bayesian Modelling of Covariate Effects on Survival. page 21, November 2016.

[94] Young J.K. Measuring Terrorism. *Terrorism and Political Violence*, page 6, October 2016.

[95] Kvaløy J.T. and Neef L.R. Tests for the Proportional Intensity Assumption Based on the Score Process. *Lifetime Data Analysis, Vol. 10*, pages 139–157, October 2003.

[96] Balakrishnan K. and Basu A.P. *Exponential Distribution: Theory, Methods and Applications*. Gordon and Breach Publishers, 1995.

[97] Dalacours K. *Islamist Terrorism and Democracy in the Middle East*. Cambridge University Press, Cambridge, 2011.

[98] Töpfer K. UN Environment Chief Urges World To Fight Root Causes Of Civil Unrest That Can Lead To Terrorism. http://www.start.umd.edu/gtd/search/Results.aspx?search=israel&sa.x=0&sa.y=0, 2007.

[99] van den Born Karin. Ayatollah khomeiny. Andere Tijden ,http://anderetijden.nl/aflevering/495/Ayatollah-Khomeiny.

[100] Goldstein K.B. Unemployment, Inequality and Terrorism: Another Look at the Relationship between Economics and Terrorism. *Undergraduate Economic Review; Vol 1, Article 6*, 2005.

[101] Brown L., Gans N., Mandelbaum A., Sakov A., Shen H., Zeltyn S., and Zhao L. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. *Journal of the American Statistical Association, Vol. 100, No. 469*, page 39, 2005.

[102] López-Calva L. and Lustig N. Explaining the decline in inequality in latin america: Technological change, educational upgrading, and democracy. page 2, January 2010.

[103] Bain L.J. and Engelhardt M. *Statistical Analysis of Reliability and Life-Testing Models. Theory and Methods. (2nd Ed.)*. 1991.

[104] Axworthy M. Revolutionary Iran: A History of the Islamic Republic, 2013.

[105] Crenshaw M. The Causes of Terrorism . *Comparative Politics, Vol. 13, No. 4.*, page 381, 1981.

[106] Mc Ginnis M., Wilhelm R., and R.E. Ward. *Progress in Material Handling and Logistics*. Springer-Verlag, 1991.

[107] Sandra M. An uncertain future for afghanistan's security sector. *Stability: International Journal of Security and Development, Vol 3., Art. 35*, November 2014.

[108] van Rossem M. Hoe amerika het is-monster baarde. Maarten, https://www.maartenonline.nl/nl/artikel/42051/hoe-amerika-het-is-monster-baarde.html, 2015.

[109] Celebrado M.B. Terrorism: A problem in the philippines. March 2014.

[110] Porter M.D. *Innovative Methods for Terrorism and Counterterrorism Data*, chapter 5, pages 91–112. Evidence-Based Counterterrorism Policy. Auerbach Publications, 2012.

[111] Hutchinson M.K. and Holtman M.C. Analysis of Count Data Using Poisson Regression. *Research in Nursing & Health*, pages 408–418, 2005.

[112] Butt M.S., Islam F., and Shahbaz M. The Income Inequality and Domestic Terrorism Nexus: Fresh Evidence from Pakistan. *Sukkur IBA, Vol. 1*, pages 102–111, October 2014.

[113] Hautsch N. *Modelling Irregularly Spaced Financial Data*. Springer, April 2004.

[114] United Nations. 2011 Report on Terrorism. 2012.

[115] E. Newman. Exploring the "Root Causes" of Terrorism . *Studies in Conflict and Terrorism*, pages 749–772, 2006.

[116] Campos N.F. and Gassebner M. International Terrorism, Political Instability and the Escalation Effect. page 19, March 2009.

[117] Perry N.J. The Numerous Federal Legal Definitions of Terrorism: The Problem of too Many Grails . *Journal of Legislation, 30*, pages 249–274, 2004.

[118] Kingsley O. Religion and Terrorism: A Socio-Historical Re-Consideration. *Journal of Alternative Perspectives in the Social Sciences Vol. 2, No. 2*, pages 550–576, 2010.

[119] Winckler O. Rapid Population Growth and the Fertility Policies of the Arab Countries of the Middle East and North Africa. *Middle Eastern Natural Environments, Bulletin 103*, pages 444–466, November 2017.

[120] Azen R. and Walker C.M. *Categorical Data Analysis for the Behavioral and Social Sciences*. Routledge, 2011.

[121] Bakker R., Hill D.W., and Moore W.H. Modelling Terror Attacks: A Cross-National, Out-of-Sample Study. March 2013.

[122] Ezcurra R. and Palacios D. Terrorism and spatial disparities: Does interregional inequality matter? *European Journal of Political Economy Vol. 42*, pages 60–74, January 2016.

[123] Pape R., Ruby K., and Bauer V. Government data exaggerate the increase in terrorist attacks. https://www.washingtonpost.com/news/monkey-cage/wp/2014/07/21/government-data-exaggerate-the-increase-in-terrorist-attacks/?utm_term=.d48a36e749d6, July 2014.

[124] Reuveny R. and Thompson W.R. *Coping with Terrorism: Origins, Escalation, Counterstrategies, and Responses*. State University of New York Press, New York, U.S.A, 2010.

[125] Jensen R.B. *The Battle against Anarchist Terrorism: An International History, 1878–1934*. Cambridge, 2013.

[126] O'Hara R.B. and Kotze D.J. Do not log-transform count data. *Methods in Ecology and Evolution, Vol. 1, Issue 2*, pages 118–122, June 2010.

[127] Wade R.H. Is Globalization Reducing Poverty and Inequality? *World Development, Vol 32, Issue 4*, pages 567–589, April 2004.

[128] Atran S., Axelrod R., Davis R., and Firschhoff B. Challenges in researching terrorism from the field . *SCIENCE Vol. 355 Issue 6323*, pages 352–354, 2017.

[129] Gygli S., Haelg F., and Sturm J.E. The KOF Globalisation Index – Revisited. February 2018.

[130] Kim S. and Whitt W. Using Simulation to Study Statistical Tests for Arrival Process and Service Time Model for Service Systems. *Winter Simulation Conference*, pages 1223–1232, 2013.

[131] Kim S. and Whitt W. Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes? . *Manufacturing & Service Operations Management, Vol. 16, No. 3*, pages 464–480, 2014.

[132] Tench S., Fry H., and Gill P. Spatio-temporal patterns of IED usage by the Provisional Irish Republican Army. *European Journal of Applied Mathematics, Vol. 27, Issue 3*, January 2016.

[133] Hillmer S.C. Monitoring and adjusting forecasts in the presence of additive outliers. *Journal of Forecasting, Vol. 3*, pages 205–215, April 1984.

[134] 4th Meeting Sixth Committee. Agreed definition of term 'terrorism' said to be needed for consensus on completing comprehensive convention against it. https://www.un.org/press/en/2005/gal3276.doc.htm, 2005.

[135] Ross S.M. *Stochastic Processes*. John Wiley & Sons, Inc., 1996.

[136] Ross S.M. *Introduction to Probability Models - 10th edition*. Elsevier, 2010.

[137] Bjø rgo T. *Root Causes of Terrorism: Myths, Reality and Ways Forward*. Routledge, London, 2005.

[138] Homer-Dixon T. Why Root Causes Are Important. *Toronto Globe and Mail*, September 2001.

[139] Krieger T. and Meierrieks D. What causes terrorism? *Public Choice*, page 9, April 2011.

[140] Krieger T. and Meierrieks D. Does Income Inequality Lead to Terrorism? *Public Choice*, page 22, March 2016.

[141] Paterson T., Clifford J.G., Brigham R., Donoghue M., and Hagan K. *American Foreign Relations: Volume 2: Since 1895*. 2014.

[142] Wagle U. *Three Related Concepts of Poverty*, pages 15–53. Springer US, New York, NY, 2008.

[143] Scientific United Nations Educational and Cultural Organization. Globalisation and culture . http://www.unesco.org/new/en/culture/themes/culture-and-development/the-future-we-want-the-role-of-culture/globalization-and-culture/.

[144] Scientific United Nations Educational and Cultural Organization. Poverty. http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/poverty/.

[145] Nelson W. *Modelling Survival Data in Medical Research*. CRC Press, Taylor & Francis Group, 2015.

[146] Feng Y. *Democracy, Governance, and Economic Performance: Theory and Evidence*. MIT Press, January 2003.

[147] Goryakin Y., Lobstein T., James W.P.T., and Suhrcke M. The impact of economic, political and social globalization on overweight and obesity in the 56 low and middle income countries. *Social Science and Medicine, Vol. 133*, pages 67–76, May 2015.