**Modelling South African social unrest between 1997 and 2016**

by

Sally-Anne Smart

Supervisors:

Dr FJC Beyers

Dr PJ van Staden

Ms M Venter

Submitted in partial fulfilment of the requirements for the degree

MSc Actuarial Mathematics

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

(August 2019)

*"When the fabric of society is so rigid that it cannot change quickly enough,*

*adjustments are achieved by social unrest and revolutions."*

*-John Boyd Orr-*

# Abstract

*Social unrest, terrorism and other forms of political violence events are highly unpredictable. These events are driven by human intent and intelligence, both of which are extremely difficult to model accurately. This has resulted in a scarcity of insurance products that cover these types of perils. Links have been found between the incidence of political violence and various economic and socioeconomic variables, but to date no relationships have been identified in South Africa. The aim of this study was to address this. Firstly, by identifying relationships between the incidence of social unrest events and economic and socio-economic variables in South Africa and secondly by using these interactions to model social unrest. Spearman's rank correlation and trendline analysis were used to compare the direction and strength of the relationships that exist between protests and the economic and socio-economic variables. To gain additional insight with regards to South African protests, daily, monthly, quarterly and annual protest models were created. This was done using four different modelling techniques, namely univariate time series, linear regression, lagged regression and the VAR (1) model. The forecasting abilities of the models were analysed using both a one-step and n-step forecasting procedure. Variations in relationships for different types of protests were also considered for five different subcategories.*

*Spearman's rank correlation and trendline analysis showed that the relationships between protests and economic and socio-economic variables were sensitive to changes in data frequency and the use of either national or provincial data. The daily, monthly, quarterly and annual models all had power in explaining the variation that was observed in the protest data. The annual univariate model had the highest explanatory power ($R^2$ = 0.8721) this was followed by the quarterly VAR (1) model ($R^2$ = 0.8659), while the monthly lagged regression model had a $R^2$ of 0.8138. The one-step forecasting procedure found that the monthly lagged regression model outperformed the monthly VAR (1) model in the short term. The converse was seen for the short-term performance of the quarterly models. In the long term, the VAR (1) model outperformed the other models. Limitations were identified within the lagged regression model's forecasting abilities. As a model's long-term forecasting ability is important in the insurance world, the VAR (1) model was deemed as the best modelling technique for South African social unrest. Further model limitations were identified when the subcategories of protests were considered. This study demonstrates that with the use of the applicable economic and socio-economic variables, social unrest events in South Africa can be modelled.*


KEYWORDS: Social unrest, protests, South Africa, insurance, modelling techniques, VAR (1) model.

# Declaration

I, Sally-Anne Smart, declare that the thesis/dissertation, which I hereby submit for the degree MSc Actuarial Mathematics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: Sally-Anne Smart

DATE: 31 July 2019.

# Preface

It is 09:00 on the 22nd of March 2016. I stayed up till the early hours of the morning to study for a board exam that I would write in two weeks' time. Grasping for the last moment of sleep before the next day's studying commenced, my phone started to explode. I grab my phone to stop my alarm, only to realise that it is not the only cause for the noise. The Brussels airport has just been bombed! A bombing that claimed the lives of 32 people and those of three perpetrators. This was not the first terrorist attack to occur, but this time it is different. My sister was on her way to that exact airport. Thankfully, a quantum physics class delayed her departure.

This was the day that the effects of these types of special risks became tangible. My head started racing: "is it ethical to exclude this from insurance contracts?", "how long will it be before these exclusions cause reputational problems?", "is there a way to predict these types of events?" … These were not new questions, but the answers that were always given were no longer adequate. I knew the answers to all my questions were rooted in creating models for special risks. In South Africa, the special risk that was equivalent to Europe's terrorist risk was social unrest, as very little was known about the interactions that were present. With this, the project was born.

I would like to thank everyone who was involved in this project! A special thanks to my three supervisors, Conrad Beyers, Marli Venter and Paul van Staden, SASRIA and Ane Neetling. Without your help none of this would have been possible. I would also like to thank the nine University of Pretoria Actuarial undergraduate students who helped with the construction of the social unrest database. Your help was invaluable.

None of this would have been possible without the financial assistance of the Absa Chair in Actuarial Science, the South African Department of Science and Technology (DST) Risk Research Platform, under coordination of the North-West University (NWU). Lastly, thank you to my family and friends. You are the wind beneath my wings!

The opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to Absa, the DST, the NWU, University of Pretoria or SASRIA.

*"May your choices reflect your hopes and not your fears"*

*-Nelson Mandela-*

# Table of Contents

# List of acronyms

| | |
|---|---|
| ABM | Agent based models |
| ACLED | Armed Conflict Location & Event Data Project |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Autoregressive Moving Average |
| CPI | Consumer Price Index |
| Crime SA | Crime South Africa |
| FAO | Food and Agriculture Organisation of the United Nations |
| FPI | Food Price Index |
| GDP | Gross Domestic Product |
| IFTRIP | International Forum for Terrorism Risk (Re)Insurance Pools |
| ISS | Institute of Safety and Security |
| MAE | Mean Absolute Error |
| OECD | Organisation for Economic Co-operation and Development |
| RMAE | Root Mean Absolute Error |
| SARB | South African Reserve Bank |
| SASRIA | South African Special Risk Insurance Association |
| Stats SA | Statistics South Africa |
| UK | United Kingdom |
| UNDP | United Nations Development Program |
| USA | United States of America |
| VAR | Vector Autoregressive model |

# List of definitions

**Assassination** is a murder of an important person for a political or religious reason (Collins English Dictionary, 2018).

**Civil war** is a war fought between diverse groups of people who live in the same country (Collins English Dictionary, 2018).

**Communal conflicts** are local groups that clash with each other as part of territorial disputes, resources, historical disagreements or local power inequalities (Fjelde and Østby, 2014).

**Gang violence** is an illegal act that is committed by a group of individuals, who identify themselves with a common name (Seymour et al., 2002).

**Political militia** is an organisation that operates like an army but whose members are not professional soldiers (Collins English Dictionary, 2018).

**Property damage**, in this study, includes any products that can currently be insured in the South African general insurance industry.

**Revolution** is a forcible overthrow of a government or social order, in favour of a new system (Oxford Dictionary, 2018).

**Riot** is a violent disturbance of peace by a crowd (Oxford Dictionary, 2018).

**Social unrest** a way in which a group of individuals show their dissatisfaction with a given situation and may range from signing petitions, sit-ins, boycotts, traffic blockades to crowd demonstrations (Renn, Jovanovic and Schröter, 2011). Protests and Social unrest are used interchangeably.

**Terrorism** is the unlawful use of violence and intimidation in the pursuit of political aims (Oxford Dictionary, 2018).

**Vigilantism** is when law enforcement is undertaken without legal authority by a self-appointed group of people (Oxford Dictionary, 2018).

**Xenophobia** is a strong and unreasonable dislike or fear of people from other countries (Asagba, 2008; Collins English Dictionary, 2018).

# List of tables

# List of figures

*"When the power of love overcomes the love of power, the world will know peace"*

*-Jimi Hendrix-*

# Chapter 1: Introduction

In a world where reading about political violence has become a norm, it is important for insight to be gained to allow a better understanding of factors that influence its occurrence. Political violence has numerous definitions. One of these is described as “the force used by a group with a political purpose or motivation, that is often designed to secure resources and access or alter the path to power” (Fjelde and Østby, 2014, p. 92). This term is not limited to one type of event and may include civil wars, social unrest, political assassinations and terrorism among others. Many of these types of events, like social unrest and terrorism, are driven by human intention and human intelligence, making them highly unpredictable (Major, 2002). There are very few techniques that are capable of modelling human behaviour. Thus, models for various forms of political violence are almost non-existent. This has led to many insurers being reluctant to cover these types of perils (Major, 2002).

The different subcategories of political violence (civil wars, social unrest, political assassinations and terrorism among others) have different characteristics and the data availability also varies. It is therefore ill-advised to model political violence collectively as essential interactions may be overlooked. A bottom-up technique that combines Agent Based Modelling (ABM), Dynamic Bayesian Networks and game theory could be more applicable for the subcategories with limited data. While time series analysis, a top-down approach, is applicable for the subcategories that have an abundance of chronological data.

South Africa has been very fortunate when it comes to terrorism and civil wars. There have been no recorded terrorist attacks or civil wars between 1997 and 2016 (Raleigh et al., 2010). Other forms of political violence have occurred during this time frame. The largest sub-category of political violence in South Africa between 1997 and 2016 is social unrest. During this 20-year period, 7172 social unrest events in South Africa were recorded by the Armed Conflict Location & Event Data Project (ACLED). This makes social unrest the most relevant form of political violence in the South African insurance context.

The main objective of this study was to deconstruct social unrest events in South Africa to gain valuable insight in the matter. The aims of the study were firstly to identify relationships between the incidence of social unrest events and economic and socio-economic variables in South Africa and secondly to use these interactions to model social unrest. The final aim was to identify variations in these interactions when considering five subcategories of protests, namely the incidence of violence, property damage and the three main reasons for protests, which are education, labour and municipal service-related protests.

This report is comprised of seven main sections and a set of appendices at the end. The background on political violence, by describing the definition and landscape in South Africa is looked at in Chapter 2. This chapter also takes a deeper look into the broader subcategories of political violence, namely social unrest and terrorism. The conditions that may potentially affect the likelihood of political violence events are described in Chapter 3. Chapter 4 gives the background on modelling techniques that may be used. The research methodology is explained in Chapter 5 and the results and discussion are in Chapter 6. The conclusion and recommendations can be found in Chapter 7. Supporting information and additional details pertaining to the study are included in the 13 appendices at the end of the report.

# Chapter 2: Political violence

Political violence is a broad term that includes several different aspects relating to violent behaviours by different groups that have some political motivation. This may include civil wars, social unrest, political assassinations and terrorism among others. For the impact of political violence on society to be understood better, it needs to be defined and looked at more closely.

This chapter is broken down into five sections. In the first section, political violence is defined, followed by an overview of political violence within South Africa. The third section describes two subcategories of political violence, namely social unrest and terrorism. Conditions that may affect the likelihood of unrest events are considered in the fourth section. The study's focus is on social unrest, for completeness we discuss terrorism and we do also provide a modelling technique which could be applicable to subcategories of political violence where there is very little data available.

## 2.1 Definition of Political violence

Multiple definitions for political violence exist. It is generally defined as the combination of all potentially deadly conflicts that relate to either elections, territorial disputes, riots or any other form of physical violence (Linke, Schutte and Buhaug, 2015). However, political violence it is also described as "the force used by a group with a political purpose or motivation, that is often designed to secure resources and access or alter the path to power" (Fjelde & Østby, 2014, p. 92). When all the definitions for political violence are combined, it serves as an umbrella term that include civil wars, communal conflicts, political militias, social unrest, riots, vigilantism, gang violence, assassinations, revolutions, xenophobia and terrorism; illustrated in Figure 1 (Asagba, 2008; Renn, Jovanovic, and Schröter, 2011; Raleigh, 2014; Linke et al., 2015).

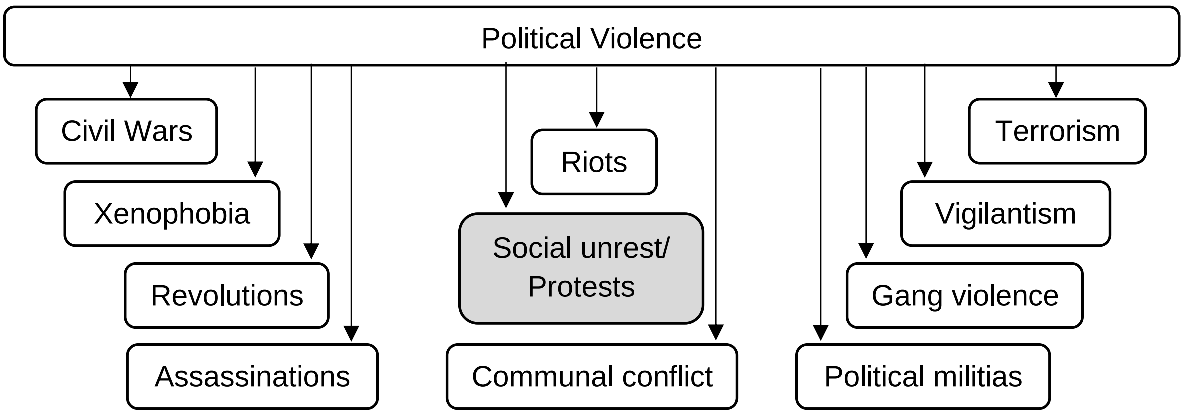



*Figure 1: Factors that form part of the umbrella term for political violence. Adapted from: (Asagba, 2008; Renn, Jovanovic, and Schröter, 2011; Raleigh, 2014; Linke et al., 2015).*

Within the geopolitical field the occurrence of political violence events is seen as political risks (Levinsohn, 2002; Mabasa, 2010). The general shared characteristics of these risks are events that are driven by human intent and human intelligence (Major, 2002). This makes them unpredictable, infrequent and a single event may result in substantial losses to life and property (Levinsohn, 2002; Mabasa, 2010). These risks may also result in clusters of unrest simultaneously breaking out in various locations within a country or following each other longitudinally (Mabasa, 2010).

Some of the most common damage inflicted by political violence events include the loss of income due to business interruption, physical damage to property and infrastructure, as well as the loss of life and health. These damages lead to insurance pay-outs as well as legal liability claims (Airmic Technical, 2013; IFTRIP, 2017). The knock-on effects are also evident in the stock market, level of foreign investments and economic growth rate (Mabasa, 2010).

As a result of the unpredictability of and risk associated with political violence events, especially the effect of terrorism acts, a number of countries have created risk pools that insure different aspects of political violence (Mabasa, 2010; Airmic Technical, 2013). Most of these pools are situated in Europe and North America. In 2013 there were only two countries in Africa, namely South Africa and Namibia, with risk pools that insure these types of risks (Airmic Technical, 2013).

## 2.2 Political violence in South Africa

South Africa has an abundant history of political violence, with the dynamics of the different political violence categories all being distinct from one another (Alexander, 2010). Since 1997 there have not been any terrorist attacks, civil wars or revolutions in South Africa, but all other forms of political violence are common (Raleigh et al., 2010).

Due to the commonality of these events, South Africa created an insurance pool so that these risks can be mitigated. The South African Special Risk Insurance Association (SASRIA) was created after the 1976 Soweto uprising which led to an increase in the number of politically motivated protests within the country (IFTRIP, 2017).

Initially SASRIA only insured political risks, but in 1998 SASRIA expanded to include non-politically motivated risks (Airmic Technical, 2013; IFTRIP, 2017). By law SASRIA has a monopoly in insuring these special risks (Mabasa, 2010). Presently SASRIA insures terrorism, riots, civil unrest, strikes, lockouts and labour disturbances that occur within the country (Mabasa, 2010; Airmic Technical, 2013). Their current portfolio includes commercial and private property, motor vehicles, construction losses, business interruption and money and goods in transit. Life and personal injury cover remain unavailable at the moment (Airmic Technical, 2013; IFTRIP, 2017).

Protests are a common occurrence in South Africa and are not isolated to either urban or rural areas. The number of protests in townships are, however, higher than any other region (Du Toit, 1993; Alexander and Pfaffe, 2014). This may be a consequence of poverty, the lack of a political authority or a result of deep historic or cultural roots (Du Toit, 1993). The bulk of social unrest events in South Africa are due to slow responses to citizens' complaints about a lack of service delivery, employment, housing, electricity and education (Alexander & Pfaffe, 2014; Lodge and Mottiar, 2016). Less significant contributors to unrest include crime and corruption (Alexander & Pfaffe, 2014). In recent years there has been an increase in the proportion of protests with violence. This trend may be a result of police being more likely to use violence as a way of controlling protests (Alexander & Pfaffe, 2014; Lodge & Mottiar, 2016).

There is little consensus about the official number of political violence events within South Africa. The number of public gatherings/protests (Table 1) and political violence incidence (Table 2) within the country indicates this. The number of recorded protests, in Table 1, is considerably higher than the number of political violence events that are depicted in Table 2. The definition of political violence used by CrimeSA, in Table 2, includes protests. This makes the values recorded in Table 1 and 2 counterintuitive. Differences within the definitions of protests, public gathering and public violence events, used by various organisations, could partly be responsible for this. It is recognised that some of the values in Table 1 are unreliable, making Table 2's figures more dependable (Alexander, 2010).

*Table 1: The number of gatherings that were recorded in South Africa's provinces during a financial year which starts on the 1st of April and ends on the 31st of March (Vally, 2009; Alexander, 2010).*

| | 2004/2005 | 2005/2006 | 2006/2007 | 2007/2008 | Total |
|---|---|---|---|---|---|
| Gauteng | 1932 | 2205 | 1888 | 1451 | 7476 |
| Western Cape | 557 | 511 | 577 | 642 | 2287 |
| KwaZulu-Natal | 1891 | 2529 | 1774 | 1146 | 7340 |
| Limpopo | 660 | 915 | 665 | 642 | 2882 |
| Eastern Cape | 754 | 1383 | 1626 | 733 | 4496 |
| North West | 1108 | 1341 | 1159 | 1502 | 5110 |
| Free State | 506 | 728 | 713 | 483 | 2430 |
| Mpumalanga | 295 | 336 | 337 | 4 | 972 |
| Northern Cape | 301 | 489 | 427 | 400 | 1617 |
| Total | 8004 | 10437 | 9166 | 7003 | 34610 |

*Table 2: The number of political violence events that were recorded in South Africa's provinces during a financial year which starts on the 1st of April and ends on the 31st of March (ISS Crime Hub, 2018a; ISS Crime Hub, 2018b).*

| | 2004/2005 | 2005/2006 | 2006/2007 | 2007/2008 | Total |
|---|---|---|---|---|---|
| Gauteng | 167 | 225 | 186 | 197 | 775 |
| Western Cape | 285 | 308 | 406 | 257 | 1256 |
| KwaZulu-Natal | 95 | 131 | 111 | 76 | 413 |
| Limpopo | 28 | 39 | 49 | 31 | 147 |
| Eastern Cape | 191 | 130 | 102 | 88 | 511 |
| North West | 57 | 52 | 39 | 94 | 242 |
| Free State | 68 | 74 | 58 | 59 | 259 |
| Mpumalanga | 48 | 63 | 50 | 61 | 222 |
| Northern Cape | 35 | 22 | 22 | 32 | 111 |
| Total | 974 | 1044 | 1023 | 895 | 3936 |

The number of political violence events that occur in South Africa is much higher than what is seen in many other countries (Lodge & Mottiar, 2016). Ponticelli and Voth (2011) looked at the annual number of chaos events that were reported in 26 European countries from 1919 to 2009. The number of chaos events was defined as the total number of attempted revolutions, demonstrations, political assassinations, riots and strikes.

In Figure 2 the average number of chaos events in these 26 countries can be seen in black, and the highest incidence of chaos events in any of the individual countries can be seen in grey. The maximum number of chaos events that were recorded in any of these countries was always less than 40 per year, with an average of less than five per year. The average number of political violence events reported in these countries between 1919 and 2009 was 1.5 events per year (Ponticelli and Voth, 2011). Unfortunately, no equivalent graph was found for developing countries.

The number of political violence events that were recorded in South Africa during the financial year starting 1 April and ending 31 March (Table 2) shows a stark contrast to the numbers seen in 26 European countries. In Europe, no country recorded more than 40 protests in a 365-day time frame. For the same time frame in South Africa, the fewest number of protests recorded was 895, with an average over the four years of 984 events per year. The only South African province that was able to consistently record values lower than 40 protests in 365 days was the Northern Cape.

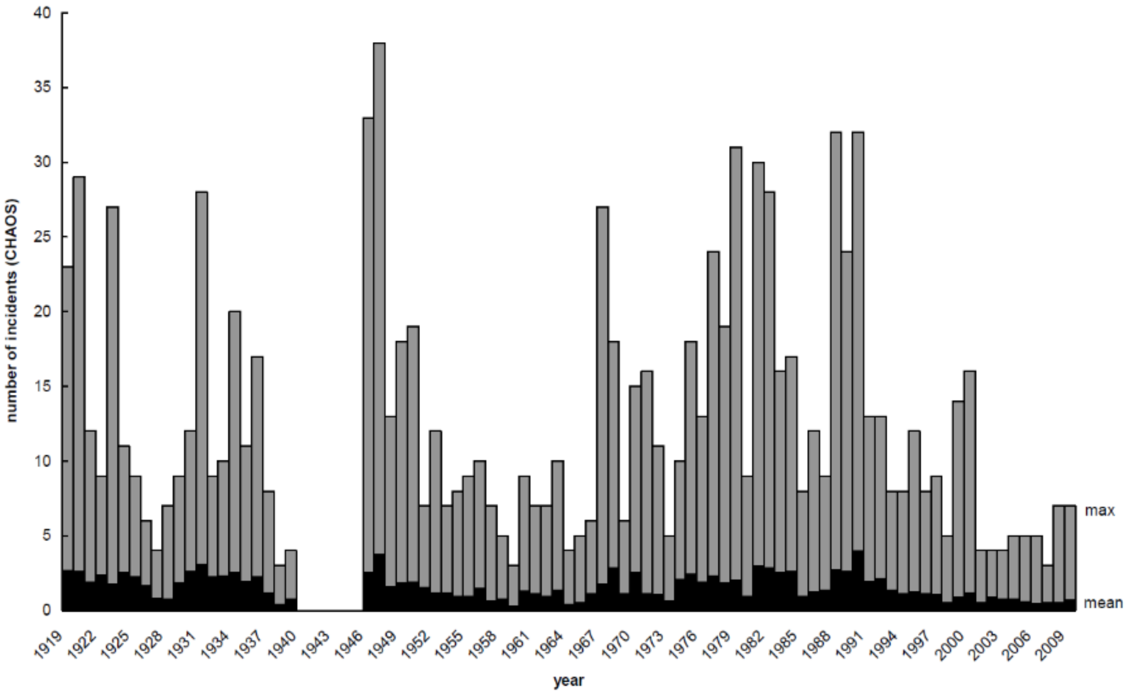



*Figure 2: The number of chaos incidents in 26 European countries from 1919 till 2009. The black indicates the average for the year, and the grey the maximum number of Chaos events in any of the 26 countries. (Ponticelli & Voth, 2011)*

## 2.3 Types of political violence

The definition of political violence includes civil wars, communal conflicts, political militias, social unrest, riots, vigilantism, gang violence, assassinations, revolutions, xenophobia and terrorism (Asagba, 2008; Renn et al., 2011; Raleigh, 2014; Linke et al., 2015). When considering the types of political risks that are insured in South Africa by SASRIA, the scope mainly focuses on social unrest (including riots) and terrorism.

### 2.3.1 Social unrest

Social unrest, otherwise known as protests, are used as a way for individuals to show their dissatisfaction with a given situation and may range from signing petitions, sit-ins, boycotts, crowd gatherings and traffic blockades to crowd demonstrations (Renn et al., 2011). Most riots start out as demonstrations, which makes it difficult to differentiate between these two events (Weinberg and Bakker, 2014). As a result, the term social unrest often includes riots. Socially acceptable methods of political expression also vary from one country to the next and changes over time (Renn et al., 2011).

Protests have the ability to bring about positive advances in a region but more often it can come with devastating consequences. Loss of life and destruction of infrastructure can have negative effects on a region (Renn et al., 2011). Social unrest is both a cause and a consequence, making it difficult to identify and quantify the risk that is involved (Renn et al., 2011). As a result, this topic is surrounded with a lot of uncertainty and ambiguity (Weinberg & Bakker, 2014).

Current literature presents different theories that attempt to explain the underlying causes of social unrest (Fox and Bell, 2016). These theories investigate the opportunities, means and motives that influence the intensity and regularity of unrest events.

The first theory is based on a grievance-based concept which attributes the occurrence of protests to some injustice, lack of service delivery or socio-economic inequalities, which were committed against individuals (Fox & Bell, 2016). The second is the mobilisation theory, that considers how easy it is to organise a protest. The easier it is, the higher the likelihood of protests becomes. Factors such as larger populations size make it easier to organise unrest events, as there is a higher probability of other individuals who are willing and able to join the protest (Fox & Bell, 2016). Other theories include the political opportunity argument and the resource mobilisation theory, where the latter is based on a collective action by a workforce to achieve a specified outcome (Fox & Bell, 2016).

A slightly different theory makes use of an economic approach to protests. This entails an opportunity cost argument on the financial gain available to protesters. Based on this theory, protests are a more common occurrence when the cost to protest is low but the gain from the protest itself is high (DiPasquale and Glaeser, 1998). This described “cost” is not only limited to the lost wages for a day of protesting, but it may go as far as considering the punishment such as time spent in jail. The size of the protest also plays a role in mitigating some of the personal cost mentioned already, as the likelihood of persecution decreases when the size of the social unrest event increases (DiPasquale & Glaeser, 1998).

These underlying causes may be triggered by an array of conditions, resulting in the occurrence of a social unrest event. In South Africa, economic, technological, environmental, political, policy, demographic and cultural related conditions have previously resulted in unrest events (Renn et al., 2011; Linke et al., 2015).

Studies have found discrepancies between the relationships that are present in different regions of the world. This indicates that historical conditions, cultural background and social circumstances of each region play a role in the motivation and incidence rate of protests (Bellemare, 2014; Weinberg & Bakker, 2014; Albertus, Brambor and Ceneviva, 2016).

This strongly suggests that the region itself is an important factor behind the relationships, drivers and motivators of the protests that are observed. However, studying individual regions leads to limited data availability. It also shows gaps within the current understanding of social unrest, as protests in one region may spill over to another region (Linke et al., 2015). This phenomenon makes it extremely difficult to predict social unrest because data related to the individual views and beliefs of regions have to be incorporated into a predictive method, and this data are not readily available.

Social unrest is a multidimensional and multicultural problem, requiring an interprofessional and inter-disciplinary approach to understand it better. The true drivers of unrest lie in the perception of injustice, instigated fear, blame-culture and lack of trust in the environment. These drivers can result from any one or a combination of potential factors and may manifest as frustration or dissatisfaction by one or more individuals (Renn et al., 2011). In this situation, a tipping point needs to be reached before any unrest is observed. Unfortunately, the time taken to reach this tipping point is not constant. It may be spontaneous or occur after a long delay of several years, making modelling complicated (Renn et al., 2011).

The degree of social unrest displayed may be categorised in four distinct levels: communication of dissatisfaction, the organisation of unrest event, mobilisation and actions of violence. Escalation or de-escalation of the level of social unrest results in the movement from one category to the next, derived by the Organisation of Economic Co-operation and Development (OECD), is shown in Figure 3 (Renn et al., 2011).

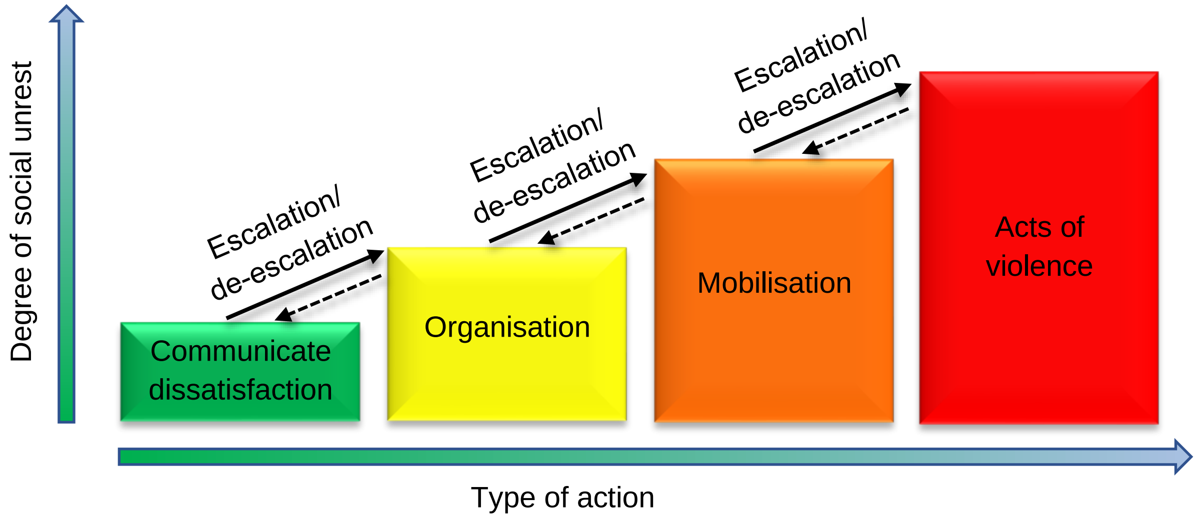



*Figure 3: Ladder of social unrest from OECD. (Renn et al., 2011)*

The level of social unrest can escalate or de-escalate from one category level to the next as a result of the political landscape, government policies, the timing of legislative acts and corruption (Renn et al., 2011; Khmelko and Pereguda, 2014). Failed infrastructure and services, as well as low levels of access to resources and slow response to grievances, may also fuel escalation between the different categories (Renn et al., 2011).

Changes between these categories are not only attributed to factors within the control of regional, or national government. Sizes of the group, the level of motivation of individuals, their expectation of the response, their leaders' attitude to using violence and their trust or distrust of police are all factors that the government have no control over (Renn et al., 2011). Unrest can escalate to acts of violence based on the interest of the media in the matter, messages on social media platforms and the depth of how the cause resonates with the rest of the population (Renn et al., 2011).

Violence occurs whenever a considerable number of individuals have grown sufficiently dissatisfied or frustrated by the political, social and economic conditions that they are experiencing (Parvin, 1973). It is possible that this is driven or influenced by a type of peer pressure or other social expectations (DiPasquale & Glaeser, 1998; Alexander & Pfaffe, 2014). There are ethnic and racial elements that are involved in rioting (DiPasquale & Glaeser, 1998).

The behaviour of police on the scene plays a key role. Their actions and official response after violence has erupted has a large impact on the outcome of the unrest event (Renn et al., 2011).

Another factor that may play a role in the incidence of acts of violence is the prior level of violence in the region (Linke et al., 2015). People assume that there is an increased likelihood of future violence in a country if there were recent episodes of violence in that specific country (Jensen and Young, 2008). In Sub-Saharan Africa it has been found that this premise is correct (Linke et al., 2015). The society's approval of using violence is powerful in predicting violent political unrest events on both a national and a regional level (Linke et al., 2015). However, in Sub Saharan Africa political conflict typically spreads from a region with a high concentration of violence to a region with low levels of violence and is not limited by the borders surrounding the country (Linke et al., 2015). Thus, the use of violence in one country not only increases the likelihood of future use of violence in that specific country but increases the likelihood of violence in neighbouring countries (Linke et al., 2015). This is known as the diffusion effect and it not only describes the spread of violence between different countries, but also between different regions (Linke et al., 2015).

The proximity in time and space is a powerful predictor of violent political unrest events. The combined effect of the proximity and approval of violence has more predictive power than when these are considered alone (Linke et al., 2015).

### 2.3.2 Terrorism

Definitions of terrorism is region specific due to the cultures, beliefs and ideologies that serve as its distinctive features. The South African definition for terrorism, described in Act 33 of 2004, is very comprehensive (Protection of Constitutional Democracy Against Terrorist and Related Activities Act 33 of 2004 (RSA); Kokott, 2005). It is defined as: "any act committed in or outside the Republic which endangers the life, or violates the physical integrity or physical freedom, or causes serious bodily injury to or the death of, any person, or any number of persons, causes the destruction of or substantial damage to any property, natural resource, or the environmental or cultural heritage, whether public or private; causes any major economic loss or extensive destabilisation of an economic system or substantial devastation of the national economy of a country; or creates a serious public emergency situation" (Protection of Constitutional Democracy Against Terrorist and Related Activities Act 33 of 2004 (RSA)). The definition continues by including cyberattacks and biological warfare. Additionally, it describes the purposes under which an attack is seen as a terrorist attack as "an act committed, directly or indirectly, in whole or in part, for the purpose of the advancement of an individual or collective political, religious, ideological or philosophical motive, objective, cause or undertaking" in Protection of Constitutional Democracy Against Terrorist and Related Activities Act 33 of 2004 (RSA).

Different countries have different definitions. Australia and the United Kingdom have similar definitions based on using intimidation to advance a cause (Lord Carlile of Berriew, 2007; Wallace, Pennell, and Robertson, 2014). The United States of America has a different approach to defining terrorism. Acts need to be certified in concurrence with the Secretary of State and the Attorney General of the United States to be defined as an act of terrorism and needs to exceed damages to the value of $5,000,000 (Terrorism Risk Insurance Act of 2002 (USA)).

When the definitions of terrorism are compared, there are seven characteristics that are most often present (Armborst, 2010). In using these seven characteristics, a new comprehensive understanding of terrorism as a baseline definition is put together as: "(1) politically inspired (2) violence committed by (3) sub-state actors that seek to (4) communicate a message by selecting a (5) symbolic and (6) civilian target(s) that are, in principle, (7) interchangeable" (Armborst, 2010, p. 423).

The definitions for terrorism and political violence often lead to researchers seeing them as mutually exclusive categories, rather than considering the characteristics that overlap (Armborst, 2010). When considering that political violence is seen as "the force used by a group with a political purpose or motivation, that is often designed to secure resources and access or alter path to power" (Fjelde & Østby, 2014, p. 92) the similarities become evident. The main difference between the definition for terrorism and political violence remains the interchangeability of the targets (Armborst, 2010). In terrorism there are numerous potential targets whereas the targets for political violence are limited by its objectives.

The risks of terrorism are similar to the associated risks of natural disasters. They are unpredictable, infrequent, result in clusters of deaths and comprise of single events that cause substantial losses (Viscusi, 2009; Mabasa, 2010; Wallace et al., 2014). However, unlike natural disasters, terrorism has human intent and intelligence as one of the major driving factors. As such, these events are not random in nature (Major, 2002; Wallace et al., 2014).

Although not random, the events are very difficult to model. Possible explanations for terrorist acts may simply be a rational decision, or it could be a way in which oppressed individuals build esteem. It may even be a way to ensure group cohesion, but it may also just be a reason that gives individuals self-permission to use violence as an action (Leistedt, 2013).

The behaviour associated with terrorism acts are complicated and the methods and motives have high variability. The individual psychopathology currently remains a mystery, but it has been shown that there is a need for these individuals to belong to a group to define their social status (Leistedt, 2013). Terrorist networks, regardless of its ideologies, frequently aim to implement a proper business model. This ensures smooth economic and organisational operations, which makes the network take on the form of an organism that aims to survive. This adaption-for-preservation aspect introduces a new challenge to modelling activities and networks. It requires a deep understanding of human psychology of the group to model their distinctive aspects of human intent and human intelligence (Bonabeau, 2002). It requires knowledge of how the network will adapt and grow, which will differ between genders, groups, cultures, ideologies and countries (Leistedt, 2013).

Quantifying the risk and damage caused by a terrorist attack is difficult. There are numerous potential targets, modes of attack as well as weapons which can be used. There can be multiple attacks in a short window of time (Major, 2002; Leistedt, 2013; Wallace et al., 2014; Chakravarty, 2015). In order to accurately model this risk, the efficiency of the counterintelligence gathering, and the terrorist resources also have to be considered (Major, 2002). The more resources a terrorist organisation has at its disposal, the greater the probability for detection before an attack occurs (Major, 2002). The impact of a terrorist attack is not limited to property damage and causalities. The media attention and decision making occurring afterward may create knock-on effects within the economy (Andersen, 2005; Viscusi, 2009; Mabasa, 2010; Mountz and Hiemstra, 2014).

Modelling these events comes with additional challenges, as the data pertaining to attacks are often not readily available. The challenge becomes even bigger for thwarted attacks, as information pertaining to it is classified to protect national security (Chakravarty, 2015). The complexities in building predictive models for terrorism is the main reason why cost-effective terrorism insurance is mostly unavailable and usually excluded from general insurance policies (Major, 2002; Wallace et al., 2014).

To compensate for this void in the insurance market, a few countries have created platforms where these risks can be insured (Chakravarty, 2015). The first terrorism insurance pool was created in Spain in 1941 and was known as Consorcio (Mabasa, 2010; IFTRIP, 2017). Israel followed a few years later in 1961 and formed a fund in terms of the law. This fund is known as the Property Tax and Compensation Fund Law and the Victims of Hostile Action Law. In 1979 South Africa became the first country in Africa, to create the terrorism insurance pool known as SASRIA (Mabasa, 2010; IFTRIP, 2017).

Most of the insurance pools were created in the aftermath of the 9/11 Twin Tower collapse in New York (Mabasa, 2010; Airmic Technical, 2013; IFTRIP, 2017). This single event brought to light the financial ramifications that come with the terror and disorder of a terror attack. The sheer financial impact of the event on the surrounding infrastructure and economy had been underestimated by most countries up to that moment.

The exact structure of these insurance pools varies from country to country, with most being funded by the government that created it. Most were created to insure property, but some offer comprehensive cover that not only insures property but offers life, business interruption and health insurance (IFTRIP, 2017). At the moment Australia, Austria, Bahrain, Belgium, Denmark, France, Germany, India, Indonesia, Israel, Namibia, Netherlands, Northern Ireland, Russia, South Africa, Spain, Sri Lanka, United Kingdom and United States of America all have insurance pools that carry the risk of terrorist attacks within their borders (Mabasa, 2010; Airmic Technical, 2013; Wallace et al., 2014; IFTRIP, 2017).

Individual countries have taken great pains to address this problem within their own borders. Ultimately an international terrorism scheme may be a better approach to model and insure the risk of terrorism as this increases the amount of data that can be made available to make this product more affordable (Chakravarty, 2015). The various definitions of terrorism, data availability through diplomatic channels, law structures and scope of coverage per region are all challenges that comes with this approach (Chakravarty, 2015).

The International Forum for Terrorism Risk (Re)Insurance Pools (IFTRIP) was created in 2015 as a method to address this. It allows for various terrorism pools to collaborate with one another by creating a platform where experience, expertise and risk management tools can be combined (IFTRIP, 2017).

# Chapter 3: Potential conditions that may affect the likelihood of unrest events

A large body of research has been dedicated to determining the underlying conditions that cause unrest events. The reason for this is simple, if it is known which circumstances increase or decrease the likelihood of unrest events, then the probability of unrest events can be adjusted by monitoring and intervening in these circumstances, to ensure that they remain at acceptable levels. In the literature this has been found to be easier said than done with economic, demographic, country-specific, socio-economic as well as internationally related conditions playing a role in unrest events (Renn et al., 2011; Linke et al., 2015). The historical conditions, cultural backgrounds and social circumstances of a region affect the motivation and incidence rate of protests (Bellemare, 2014; Weinberg & Bakker, 2014; Albertus et al., 2016). This has resulted in the relationships observed in various regions differing from one another.

The conditions that have been investigated in the literature as potential unrest triggers are divided into six categories, namely economic factors, demographic factors, country, socio-economic, international and other factors. Each of these six categories are discussed in this chapter. A table has been created to summarise the categories and factors with their main relationships for an easy overview. This summary, provided in Table 3, can be found at the end of Chapter 3.

## 3.1 Economic factors

Economic factors such as inflation rates and economic growth are important in describing the economic climate of a country. The impact of these factors on political violence events are important as the economic climate has a large impact on the lives of the people living within a country.

Economic hardship can translate to unrest events (Ponticelli & Voth, 2011). It is believed that improving economic conditions decreases the level of social unrest (Parvin, 1973). The impact of poor economic conditions may make political violence more appealing to the individuals who are most affected by it (Burrows and Harris, 2009; Caruso and Schneider, 2011). General unrest is not always an immediate response to economic conditions and can come as a delayed response to it (Ponticelli & Voth, 2011).

However, the impact that economic growth has is even more complicated. High levels of economic growth and development may cause political violence due to unreasonable expectations not being met and greater levels of inequality causing a larger gap between rich and poor. In addition, it may increase inflation, thereby decreasing purchasing power for the people within the country acting as an additional cause for social unrest (Caruso & Schneider, 2011; Renn et al., 2011). This was seen during the Arab Spring, in the Middle East, where inflation was the main factor that drove the citizens' dissatisfaction with their governments, that served as a catalyst (Dewey et al., 2012).

This contrast of high and low levels of economic growth on unrest events is illustrated in Figure 4. The expected relationship between the number of unrest events and economic growth differs based on the rate at which the economy grows.

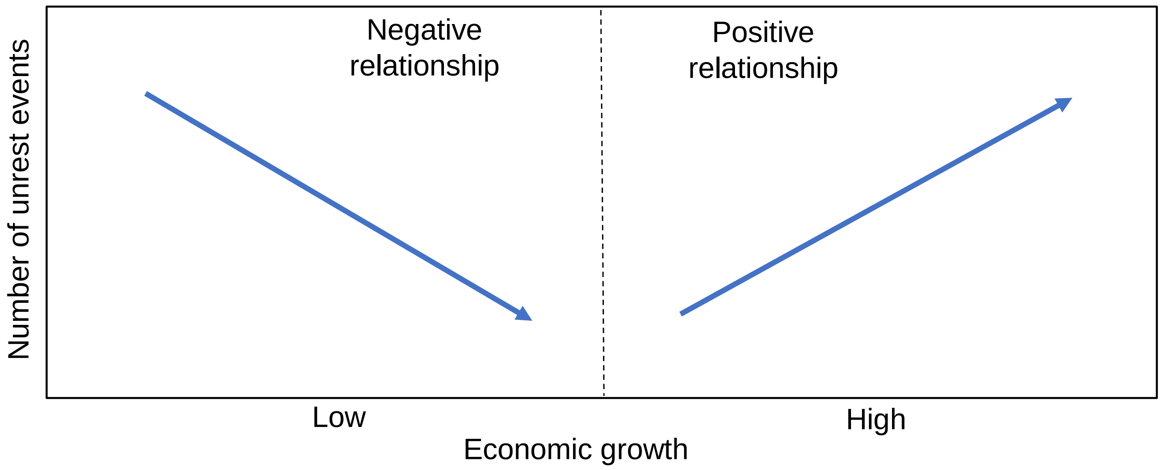



*Figure 4: A schematic comparison of the expected relationship between the number of unrest events and the economic growth rate for both low and high growth rates. Adapted from: (Parvin, 1973; Caruso & Schneider, 2011)*

Different economic indicators are used to test these relationships, each measuring a different aspect of economic growth. The main indicators considered in the literature are the Gross Domestic Product (GDP) per capita, the real GDP per capita, GDP, the GDP growth as well as the real GDP growth.

GDP per capita and the real GDP per capita are negatively correlated with rioting, urban unrest and terrorism in numerous studies (DiPasquale & Glaeser, 1998; Caruso & Schneider, 2011; Buhaug and Urdal, 2013; Hendrix and Haggard, 2015; Weinberg & Bakker, 2014). However, these GDP per capita indices have displayed a positive association with protests (Fox & Bell, 2016). Buhang and Urdal (2013), who initially found a significant negative relationship between GDP per capita and social disorder, saw that GDP per capita became trivial when the effects of economic shocks were accounted for.

GDP and its relationship with unrest is expected to have a negative relationship with each other, but this does not hold for war torn countries. In this case increased levels of GDP led to increased levels of unrest (Ponticelli & Voth, 2011).

GDP growth per year and the real GDP growth both presented a negative relationship with the number of unrest events (Caruso & Schneider, 2011; Ponticelli & Voth, 2011; Dewey et al., 2012; Fox & Bell, 2016). It was found that economic contraction increased the probability of civil wars and terrorism within the country (Jensen & Young, 2008). A clear example of this is the Arab Spring. Stagnated growth acted as a destabilising factor that fuelled the population discontent, leading to the occurrence of large-scale protests (Dewey et al., 2012).

A possible reason why studies investigating the role that economic growth has on unrest events have yielded mixed results is because researchers have been testing for linear trends. This does not account for the change in the relationship that has been described in Figure 4. Better results may be obtained by dividing the economic data into two groups, namely low and high economic growth. Another possible explanation is that it is not economic growth that increases the probability of social unrest but rather a sudden change in the economic situation that triggers unrest events (Caruso & Schneider, 2011). Finally, it may be that political instability causes lower levels of economic growth (Alesina et al., 1996).

## 3.2 Demographic factors

The demographic factors that may act as potential triggers includes urbanisation, population size and growth and ethnic heterogeneity. Changes in these demographics can put a lot of strain on the resources that are available in certain regions and create an imbalance within the systems that are in place. By looking at these aspects individually a clearer understanding of its role in political violence events can be attained.

### 3.2.1 Urbanisation

The role of urbanisation is country specific. When countries are considered individually, the more populated regions experience increased levels of unrest and riots (Hendrix & Haggard, 2015; DiPasquale & Glaeser, 1998). A reason for this may be that it is easier to organise political unrest in densely populated areas as information can be spread more effectively (Linke et al., 2015; DiPasquale & Glaeser, 1998; Klomp and de Haan, 2013). This is known as the information sharing phenomena (Linke et al., 2015; DiPasquale & Glaeser, 1998; Klomp and de Haan, 2013).

When considering the level of urbanisation in different countries this premise did not always hold (Parvin, 1973; Fox & Bell, 2016). A study considering 71 countries from 1972 to 2007 found that the urban population size did not have a significant effect in modelling the number of social unrest events (Weinberg & Bakker, 2014). Furthermore, since the end of World War 2, all the major political unrest events and violent revolutions have occurred in agricultural driven countries (Parvin, 1973).

### 3.2.2 Population Size and Population Growth

The role of population size and growth have shown varying results. A study performed by DiPasquale & Glaeser in 1998 showed a strong positive correlation between the size of a country's population and the number of riots that occurred within the country. In 2009, Raleigh and Hegre found a positive relationship between the local population size and the risk of conflict events in 14 Central African countries.

A possible explanation for this is that population growth puts additional pressure on service delivery, education and health care in the region. If the growth is slow, then the additional pressure can be absorbed more easily than during periods of rapid growth. When the additional pressure is not absorbed it may cause dissatisfaction which may then result in unrest (Østby, Urdal, Tadjoeddin, Murshed and Strand, 2011).

In contrast to the previous studies that found positive relationships, Weinberg & Bakker (2014) found a negative relationship between the number of protests and the population size. In a different study, Buhang and Urdal (2013) considered 55 major cities in Sub-Saharan Africa and Asia and found little to support that the probability of social disorder increased as a result of a large city size or local population growth.

This study was supported by Hendrix & Haggard (2015) who concluded that general city population growth did not have a significant impact on the number of urban unrest events. It has been found that rapid population growth in urban areas is negatively correlated to the frequency of protests (Fox & Bell, 2016).

### 3.2.3 Ethnic heterogeneity

The impact that ethnic heterogeneity has on political violence events are country specific (Levinsohn, 2002). It was seen that some countries have a higher propensity for protests due to increased levels of fractionalisation, increasing the number of interests that must be considered when decisions are made (Dewey et al., 2012; Klomp & de Haan, 2013). Naturally different ethnic groups have different ethnic norms, and this can be seen in the propensity and manner in which they voice their dissatisfaction with a matter (DiPasquale & Glaeser, 1998; Renn et al., 2011; Linke et al., 2015).

In some African countries, there are power hierarchies that exist between the various ethno-regional communities which determine its relative importance and influence within the region. In turn this may result in inequality as the political landscape and ethnic hierarchy can become interwound, biasing the state's decisions. (Fjelde & Østby, 2014). Any hierarchical shifts can play a significant role in the type of political violence events that are orchestrated at different times (Raleigh, 2014). Social tensions between different ethnic groups may exacerbate the number of unrest events (Weinberg & Bakker, 2014).

A negative parabolic relationship (with a maximum turning point) has been found between the likelihood of civil wars and the level of ethnic diversity (Elbadawi and Sambanis, 2000; Jensen & Young, 2008). This suggests that the level of ethnic diversity alone does not result in unrest unless an economic or political rivalry is created and driven by the level of diversity as major groups compete with each other (Elbadawi & Sambanis, 2000; Jensen & Young, 2008).

## 3.3 Country specific factors

Country specific factors include the type of regime, government policies, level of corruption and a nation's dependence on natural resources. All these factors can influence the incidence of unrest events within a country in different ways. Most of these results are country dependent and it is important to consider each country individually to obtain a more comprehensive understanding of the impact on political violence events.

### 3.3.1 Type of regime

There are two main regimes in the world, dictatorships and democracies (Buhaug & Urdal, 2013). Any regimes that fall between these two extremes are known as hybrid regimes. Hybrid regimes are usually countries where the government in power is not seen as completely legitimate. Most of the countries in sub-Saharan African fall into this category (Fox & Bell, 2016). In addition, there is a discrepancy between the expectations that its citizens have and the actual performance of the country and this is what drives the risk of unrest events. Hybrid regimes usually have higher levels of protests than democratic countries (Weinberg & Bakker, 2014; Fox & Bell, 2016).

Based on the literature, the distribution between the type of regime and the number of unrest events takes the form of a negative parabolic curve, with the maximum turning point within the hybrid regime. This relationship is depicted in Figure 5 (Linke et al., 2015; Weinberg & Bakker, 2014; Fox & Bell, 2016).

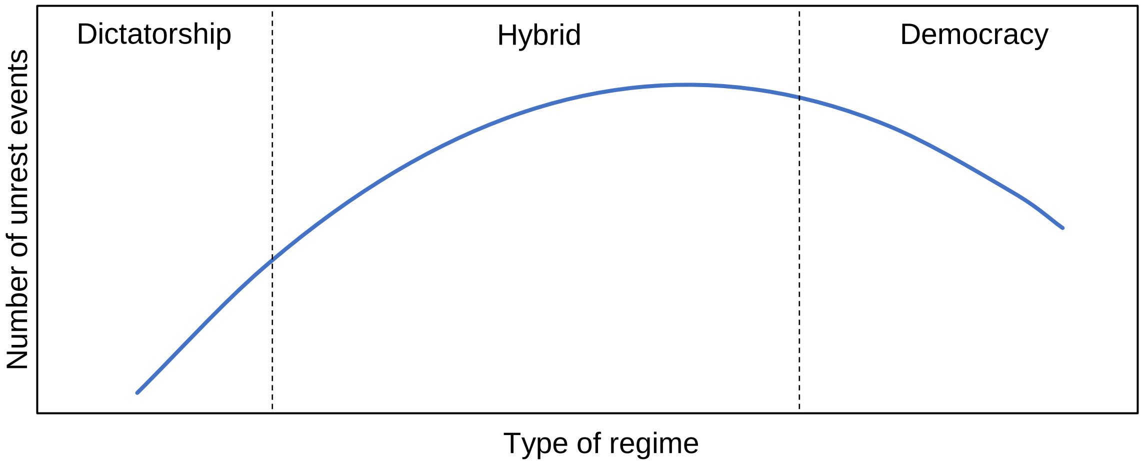



*Figure 5: A schematic illustration of the relationship between the type of regime and the expected number of unrest events in the region. Adapted from: (Dewey et al., 2012; Buhaug & Urdal, 2013; Linke et al., 2015; Weinberg & Bakker, 2014; Fox & Bell, 2016).*

Between these three groups dictatorships have the lowest levels of social disorder and riots, whereas democratic countries experience much higher levels of unrest, (Dewey et al., 2012; Buhaug & Urdal, 2013; Fox & Bell, 2016). This is due to an increased level of freedom of speech available to citizens in democratic countries.

In dictatorships there are harsh punishments to opposers of the rule of law (Buhaug & Urdal, 2013). In democratic countries there are multiple political parties and civil society organisations (like trade unions) that are independent of the ruling government (Fox & Bell, 2016). These societies can not only organise a protest but can give its support to a cause. It has been argued that the effect of a protest is much larger when a civil society organisation supports the cause (Klomp & de Haan, 2013; Alexander & Pfaffe, 2014). The number of opposition parties plays a role in the social unrest and may lead to conflict escalation (Khmelko & Pereguda, 2014).

### 3.3.2 Government policies

Government policies may not only lead to social unrest but may play an integral role in the incidence, timing, appearance and escalation of conflict in these scenarios (Renn et al., 2011; Khmelko & Pereguda, 2014). This may be as a result of a component of the population opposing the proposals and using social unrest to voice this. Some protests coincide with the timing of legislative meetings while a delay is present in many of the general protests where it is not an immediate response to economic and government fiscal decisions (Ponticelli & Voth, 2011).

There are a number of fiscal policies that affect the level of stability within a country. The stability can be enhanced by increased per capita police expenditure and increased levels of government expenditure. (DiPasquale & Glaeser, 1998; Ponticelli & Voth, 2011). Democratic countries who were experiencing mass protests prior to upcoming elections were prone to increase government spending in the year before elections were scheduled to take place. These steps may improve the citizens perspective of the country, gain support for the elections and may decrease the number of mass protests (Ponticelli & Voth, 2011; Klomp & de Haan, 2013). Unfortunately, this approach leads to increases in the budget deficit (Klomp & de Haan, 2013). In the short term this relationship may increase the stability within a region, but it does not ultimately quell mass protests.

On a national level, it was found that countries with higher levels of unrest were more prone to being indebted and that attempts at financial consolidation further increased the level of unrest within the country (Ponticelli & Voth, 2011). Fiscal policies such as tax increases, budget cuts and lowering government subsidies have played a role in increasing levels of instability within the country (Ponticelli & Voth, 2011). Countries should be especially cautious when decreasing food subsidies as this may enhance the rate at which citizens become dissatisfied (Bellemare, 2014). The Arab Spring is a clear example of this. The financial crisis forced the Arabic countries to decrease housing, fuel and food subsidies, creating a high level of dissatisfaction for the individuals reliant on these subsidies (Dewey et al., 2012).

### 3.3.3 Corruption

Corruption is widely described as a factor that may play a role in conflict escalation of political violence events (Renn et al., 2011; Khmelko & Pereguda, 2014). Objective factors like corruption and poverty play a role in the cognitive processes of the individuals who partake in political violence events (Armborst, 2010). These processes then affect the reaction deemed appropriate for each of the grievances, altering the probability of violence to be used (Armborst, 2010). This intricate relationship may be a reason why a conclusive deterministic relationship between corruption and political violence has not yet been found (Armborst, 2010; Renn et al., 2011; Dewey et al., 2012).

### 3.3.4 Dependence on natural resource

Natural resources like oil, diamonds, gold, timber and clean water play a significant role within nations (Le Billon, 2004; Burrows & Harris, 2009). Strategic relations are built between countries to ensure that they have access to the resources that they require. However, such security cannot always be ensured, and wars can breakout as individuals and nations fight about the value that is attached to the resources (Jensen & Young, 2008). The scarcity of the resources may open the door for criminal networks and terrorists to finance their operations through black market trafficking (Le Billon, 2004; Burrows & Harris, 2009). A significant positive relationship was found between a region's dependence on natural resources and the level of violence in the region (Jensen & Young, 2008). The dependence on natural resources had an impact on the duration of violent behaviour that was displayed within the region (Jensen & Young, 2008).

## 3.4 Socio-economic factors

Some of the main socio-economic factors considered are homeownership, the age dependency ratio, unemployment and poverty, income growth, segregation and inequality, migration, food prices and human development and education. All these factors may play a role in causing or escalating political violence events (Renn et al., 2011). Many of these factors are country specific.

### 3.4.1 Homeownership

The relative homeownership rate in the US has a significant negative correlation with riot occurrence. The incentive to start fires and destroy property when the people themselves own property in that neighbourhood is considerably lower. However, once a riot has started homeownership has little impact (DiPasquale & Glaeser, 1998).

### 3.4.2 Age dependency ratio

The age dependency ratio compares the proportion of the population that is either above or below working age to the proportion of the population that is of working age (between 15-65 years of age) (Burrows & Harris, 2009). This ratio determines the pressure that an aging or young population places on the productive population. The pressure is created due to the difficulty for a country to maintain or enhance its economic growth when it has an ageing population, and/or a significant portion of the population is younger than the legal working age (Klomp & de Haan, 2013). The knock-on effects created by this pressure may result in social unrest events.

### 3.4.3 Unemployment and poverty

National unemployment has a positive relationship with the incidence of political violence (Caruso & Schneider, 2011). It may increase the number of riots, xenophobia and terrorist attacks as it lowers the opportunity cost of time (DiPasquale & Glaeser, 1998; Asagba, 2008; Caruso & Schneider, 2011). It’s not only the level of unemployment that causes unrest events but changes in unemployment rates can induce social disturbances (Caruso & Schneider, 2011). Depending on the country, unemployment rates of different ethnic groups may have a varying impact on the incidence of unrest events (DiPasquale & Glaeser, 1998).

When youth unemployment decreases or is kept low it has been found that social unrest is less appealing (Daras and Mazis, 2015). Unemployed youths have a lot of time and energy. When it is used to organise protests, they have the ability to mobilise the community as a whole and mount dramatic protests that may undermine the legitimacy of the country’s political landscape (Alexander & Pfaffe, 2014). During the Arab Spring in Egypt, the unemployed youths were on the forefront of the unrest (Dewey et al., 2012).

Unemployment and poverty go hand in hand. There is little evidence that poverty causes increased levels of protests, riots or other forms of political violence (DiPasquale & Glaeser, 1998; Caruso & Schneider, 2011). However, it is widely suggested that poverty may drive dissatisfaction and frustration (DiPasquale & Glaeser, 1998; Armborst, 2010; Caruso & Schneider, 2011; Linke et al., 2015). If this is the case, then poverty may enhance the incidence of unrest within a region as a knock-on effect (DiPasquale & Glaeser, 1998; Caruso & Schneider, 2011).

Poverty can have economic impacts. Fewer individuals paying taxes lower the levels of national income. This has knock-on effects, as the country’s ability to suppress political violence is weakened (Fearon and Laitin, 2003; Jensen & Young, 2008). An impression of poverty can be created when workers are exploited, laid off, lose employee benefits or are unpaid and any one of these scenarios may result in social unrest events occurring (Renn et al., 2011).

### 3.4.4 Income growth

Income growth has different impacts on social unrest. As expected, low incomes can lead to dissatisfaction being voiced through unrest (Renn et al., 2011; Linke et al., 2015). The rate and direction of income growth plays an important role in the stability within a country. Negative income growth is linked to increased levels of instability (Weinberg & Bakker, 2014). There is an optimal level between positive income growth and unrest events. Below the threshold, a negative relationship is found, and after the threshold is reached the opposite relationship applies (Parvin, 1973). Figure 6 illustrates the resultant distribution between the income growth rate and the number of unrest events when the aforementioned relationships are combined. The change seen after the optimal income growth rate can be due to knock-on effects in income inequality, inflation, food prices and the level of poverty in the region.

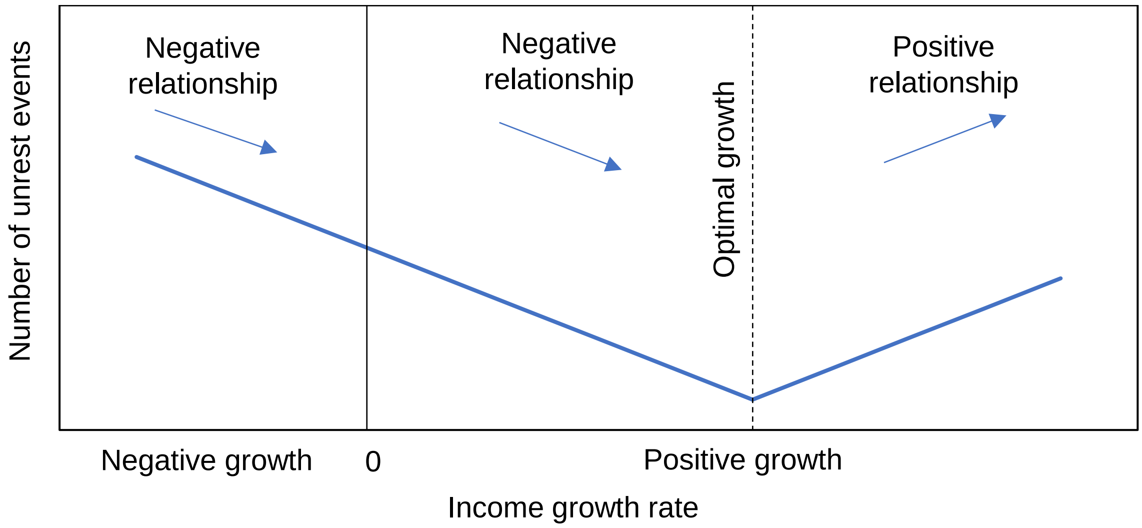



*Figure 6: An illustrative distribution between the income growth rate and the number of unrest events. Adapted from: (Parvin, 1973; Renn et al., 2011; Linke et al., 2015; Weinberg & Bakker, 2014)*

### 3.4.5 Segregation and inequality

Segregation within a society can manifest itself on a social, cultural or racial level. It is supposed that urban unrest is motivated by social exclusion and segregation rather than racial segregation (Malmberg, Andersson and Östh, 2013). This principle is supported by data from the urban unrest in Sweden in 2009, where more cars were burnt in protest in residential areas that experienced higher levels of isolation (Malmberg et al., 2013).

Location is not the only factor that can cause segregation between individuals and groups of people. Inequality, often measured by the GINI coefficient, can do the same. There are two broad categories of inequality, namely vertical and horizontal inequality. Vertical inequality causes variation between individuals and households, while horizontal inequality causes variation between different groups of people (Fjelde & Østby, 2014). These groups may refer to ethnicity, religious alliance, nationality or any other difference between individuals. In Sub-Saharan Africa the stronger that either one of these two forms of inequality become, the higher the probability of violent communal conflicts in the region is (Fjelde & Østby, 2014).

Ethnic discrimination in the labour market can lead to political violence events (Caruso & Schneider, 2011). Tension due to unequal access across groups may destabilise a region Østby et al., 2011). In sub Saharan Africa armed conflicts in these regions are even more likely when the largest ethnic group has less access than the other groups (Fjelde & Østby, 2014). This effect is not limited to Sub-Saharan Africa. Similar results were observed in 26 developing countries (Østby, 2008).

Numerous studies concluded that income inequality contributes to political unrest levels (Parvin, 1973; Alesina and Perotti, 1996; Caruso & Schneider, 2011). However, income inequality is not considered to be a cause of violent conflicts (DiPasquale & Glaeser, 1998; Østby, 2008). Violent conflicts are formed as a result of a collective force rather than an individual force. Therefore, horizontal inequality plays a bigger role (Østby, 2008). In the 1980's France gained relative public peace by addressing inequality in poorer areas by exerting a lot of social control by deploying social workers and educators in these areas (Cesari, 2018).

### 3.4.6 Migration

It has been found that an inflow of migrants into a certain country results in an outflow of the locals from that country (Cushing and Poot, 2004; Stillwell et al., 2014). There are two different types of migration patterns that exist. National migration occurs within the borders of a country, whereas international migration occurs across country borders (Cushing & Poot, 2004).

There are two main factors that drive migration: forced migration and forward-looking migration (Cushing & Poot, 2004; Stillwell et al., 2014). Forced migration is caused by factors like war, famine as well as political unrest (Stillwell et al., 2014). These migrants are forced to move to a safer place and usually do not have the correct paperwork. As a result, they often require the help of human smugglers to migrate from one country to the another (Papadopoulos and Fratsea, 2013). These migrants often show selective behaviour towards countries that have favourable migration policies (known as the migration phenomena) and may target and exploit these countries (Cushing & Poot, 2004).

Forward-looking migration on the other hand is driven by individuals intending to maximise the well-being of their households. This may be achieved through working holidays, international education, retirement migration and temporary or full-time work schemes (Cushing & Poot, 2004). The brain-drain caused by this results in a weakening of the original country (Mountz & Hiemstra, 2014). Globalisation has further intensified this problem as professional migration has become easier and there are more incentives for these professionals to make use these opportunities (Cushing & Poot, 2004). Figure 7 shows the factors that are affected by changes in migration patterns, as well as the potential changes that can be brought about in the characteristics of unrest events.

Conflict between the migrants and the local community is inevitable when locals feel threatened and insecure due to the presence of the migrants (Mountz & Hiemstra, 2014; Daly, 1996). It is important to note that no relationship between elevated levels of migration and the number of riots within a region have been found (DiPasquale & Glaeser, 1998). The recent wave of xenophobic attacks in South Africa can be an indication that a relationship may indeed exist.

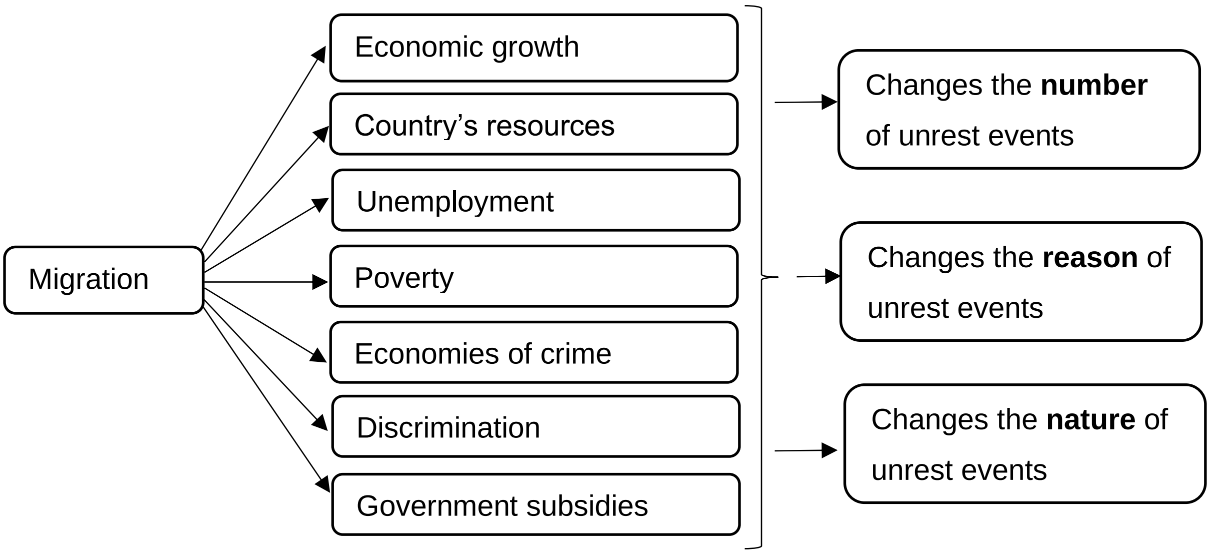



*Figure 7: Factors that may be affected by changes in migration patterns and the resultant changes that these factors can have on unrest event characteristics. Adapted from: (Daly, 1996; Cushing & Poot, 2004; Papadopoulos & Fratsea, 2013; Mountz & Hiemstra, 2014; Rânceanu and Marghescu, 2015)*

### 3.4.7 Food prices

Increases in food prices threaten the livelihood of many people. This can easily create displeasure within a society (Dewey et al., 2012). A significant positive correlation exists between food prices and the incidence of social unrest and riots within the region (Bellemare, 2014; Hendrix & Haggard, 2015; Weinberg & Bakker, 2014). This relationship is affected by type of regime that governs the country and is more prominent in democratic countries than in autocratic countries (Hendrix & Haggard, 2015).

When data from multiple countries are considered together, these relationships no longer hold. First world countries have higher food prices than third world countries but experience fewer unrest events. Weinberg et. al. (2014) found that it wasn't necessarily high food prices that lead to unrest events but rather the short-term volatility in food prices which had a positive relationship with the number of unrest events.

### 3.4.8 Human development and Education

Declining child mortality and increased levels of education have all resulted in higher levels of human development (Kuhn, 2012). Education driving individuals to have higher expectations for the future helps in managing unrest. The availability of education opportunities not only lowers the level of political unrest within a region, but changes the way dissatisfaction is voiced (Parvin, 1973; Armborst, 2010; Dewey et al., 2012).

However, human development may increase the pressure on governments. Ultimately it is seen as the government's responsibility to bring these aspirations to life. In addition, as human development increases so does the proportion of the population exposed to media and engaging in political discussions. This combination causes problems when expectations are not met, and the government is presumed to be the cause (Kuhn, 2012).

## 3.5 International factors

International factors are factors that can influence a country irrespective of the country or region that it originates from. These factors include technological developments, mass media and globalisation. Many of the technological developments have changed the manner and speed at which information can be shared, making it easier for people to voice their opinions, organise protests, or influence debates in different regions (Tufekci and Wilson, 2012; Wolfsfeld, Segev and Sheafer, 2013).

### 3.5.1 Technological developments and the internet

The introduction of technology as well as technological developments are bringing about huge changes within the labour market. As technology takes the place of humans the demand for skilled workers increase and the demand for non-skilled workers decrease (Caruso & Schneider, 2011; Renn et al., 2011). This alteration increases employment and salary inequality between skilled and non-skilled workers (Caruso & Schneider, 2011). Other knock-on effects include increased unemployment and poverty in the region (Renn et al., 2011).

During the Arab spring, protesters communicated over the phone, internet and through social media (Tufekci and Wilson, 2012; Howard and Hussain, 2013). Internet usage was statistically significant in determining an individual's involvement in protest activity during this time. It is also believed that cell phone penetration may have played a role in the way these protests evolved (Dewey et al., 2012). The effect of technological developments is not limited to directly impacting communication between groups but increases the rate and ease of information sharing (mobilisation theory). This can cause further knock-on effects on mass media.

### 3.5.2 Mass media and Communication

The main role of mass media is to inform the public about various issues. The manner in which the news is covered can escalate a social movement or bring its legitimacy into question. The inclusion of graphic images has a large effect in framing the public's perception and gaining attention (Arpan et al., 2006).

Ponticelli and Voth found no evidence that the spread of mass media facilitated an increase in the number of mass protests in a study of 26 EU countries from 1919-2009 (Ponticelli & Voth, 2011). Methods of communication do not cause unrest events. There must be some sort of economic, social or political grievance which is the driving factor behind the unrest (Dewey et al., 2012). Therefore, it is extremely important to consider the political context of the region before the role of social media and mass media is investigated (Wolfsfeld, Segev and Sheafer, 2013).

Modes of communication facilitate discussions between individuals, making it easier to organise social movements and spread information at a faster rate (Dewey et al., 2012; Howard & Hussain, 2013). The content that is posted on social media is not completely controlled by governments (Tufekci & Wilson, 2012). It can be used by governments to monitor and predict civil unrest. The information from social media allows them to resolve, thwart or manage planned events (Dewey et al., 2012). However, this method is only able to predict protests a few weeks in advance, but it cannot be used to predict the number of protests in the distant future (Ramakrishnan et al., 2014).

Governments may however, censor social media platforms. The level of censorship on social media may play a role in the incidence of unrest events. An agent-based model found that in the absence of social media censorship there were fewer riots and larger periods of peace between riot outbursts (Casilli and Tubaro, 2012). Social media has made it easier to entice the international community with a specific social movement which may further help the cause as it may cause the movement to spill over to other countries (Dewey et al., 2012).

### 3.5.3 Globalisation

Globalisation has bonded countries together through different economic and cultural ties. It has improved human development and increased interconnection and interdependency between countries (Tuathail, 1999; Goldin and Vogel, 2010).

Globalisation has created an extremely complicated world. Not only has it brought about increased levels of systemic risk, but it has resulted in a governance gap with new systemic risks being created. Most of these are clearly visible in geopolitics. This governance gap has been created by countries being obliged to adhere to international law while prescribing the local laws. This results in a country not having total control over the laws that govern them and diminishing control over its economic destiny (Tuathail, 1999; Goldin & Vogel, 2010).

Globalisation effects are both geostrategic and geopolitical (Tuathail, 1999). Globalisation increases the rate of human development and intensifies the exchange of professionals. This causes increased levels of inequality and alters migration patterns in different countries (Cushing & Poot, 2004; Goldin & Vogel, 2010).

Globalisation played a significant role in the global financial crisis that occurred in 2007-2008 (Goldin & Vogel, 2010). What started off as a US banking and stock exchange crisis, ended in a global crisis due to increased levels of interconnection and interdependency between different countries (Nesvetailova and Palan, 2008). Large numbers of protests occurred all over the world, as the implications of the financial crisis was felt across the world.

## 3.6 Other factors

Researchers have gone even further to identify factors that may shed light on incidents of political violence. Factors from the size of the government to oil price increases, level of trade openness, natural disasters, access to water, failed infrastructure and exchange rate fluctuations have all been investigated.

Relationships with various aspects of political violence were identified with five of these factors. The size of the government has been shown to be somewhat positively correlated to rioting, suggesting that it plays a role in conflict escalation of social unrest events (DiPasquale & Glaeser, 1998; Khmelko & Pereguda, 2014). Countries with increased levels of trade openness in their economies had increased levels of unrest (Hendrix & Haggard, 2015). Access to water during droughts, failed infrastructure and natural disasters like earthquakes and hurricanes are also known to trigger social unrest within regions (Renn et al., 2011). No evidence was obtained to support the premise for the two other factors, namely oil price increases and exchange rate fluctuations.

There is no definitive evidence suggesting oil price increases lead to an increase in the number of urban unrest events (Hendrix & Haggard, 2015). There appears to be a slightly stronger trend that increases in the oil price may decrease the number of urban unrest events, however the trend is not strong enough to be conclusive (Hendrix & Haggard, 2015). Exchange rate fluctuations were found not to be significant in estimating urban unrest either (Hendrix & Haggard, 2015).

## 3.7 Summary of the relationships found in the literature

The different impact of all these different variables on political violence events can be difficult to compare and understand due to the vastness of the topics covered. To simplify the information for easier understanding and comprehension Table 3 was created to summarise the categories and factors with their main relationships for an easy overview.

*Table 3: Summary of all the relationships that were identified in the literature study.*

| | Factor | Relationships Found with Political Violence Events: Positive Relationship | Negative Relationship | Mixed Results | Country Specific | Other |
|---|---|---|---|---|---|---|
| Economic | Economic Growth | - | - | ✓ | - | - |
| | Inflation | ✓ | - | - | - | - |
| Demographic | Urbanisation | - | - | - | ✓ | - |
| | Population Size and Growth | - | - | ✓ | - | - |
| | Ethnic Heterogeneity | - | - | - | ✓ | - |
| Country | Type of Regime | - | - | - | - | Negative Parabolic |
| | Government Policies | - | - | - | - | Multiple |
| | Corruption | - | - | ✓ | - | - |
| | Dependence on Natural Resource | ✓ | - | - | - | - |
| Socioeconomic | Homeownership | - | - | - | - | Initial Impact |
| | Age Dependency Ratio | - | - | - | - | Knock-On |
| | Unemployment | ✓ | - | - | - | - |
| | Poverty | - | - | - | - | Knock-On |
| | Income Growth | - | - | - | - | Positive Parabolic |
| | Segregation | - | - | ✓ | - | - |
| | Vertical Inequality | - | - | ✓ | - | - |
| | Horizontal Inequality | ✓ | - | - | - | - |
| | Migration | - | - | - | - | Knock-On |
| | Food Prices | - | - | - | ✓ | - |
| | Human Development and Education | - | - | - | - | Knock-On |
| International | Technological Developments | - | - | - | - | Knock-On |
| | Mass Media and Modes of Communication | - | - | ✓ | - | - |
| | Globalisation | - | - | - | - | Knock-On |
| Other factors | Size of The Government | ✓ | - | - | - | - |
| | Oil Price Increases | - | - | - | - | No Relationship |
| | Level of Trade Openness | ✓ | - | - | - | - |
| | Natural Disasters | ✓ | - | - | - | - |
| | Access to Water | ✓ | - | - | - | - |
| | Failed Infrastructure | ✓ | - | - | - | - |
| | Exchange Rate Fluctuations | - | - | - | - | No Relationship |

# Chapter 4: Methodology

The research methodology used to model social unrest is described in this chapter. It starts with a description of how the economic, socio-economic and social unrest database for South Africa was created. This is followed by a description of the statistical analysis and modelling approaches that were used in this study.

## 4.1 Creating a database for social unrest and economic and socio-economic factors

Based on the literature the data required for this study was vast. As no databases could be found that contained a number of the variables described within the literature review, a database had to be created. The process that was followed is described below. It starts by describing the sources from which the economic, socio-economic and social unrest data were obtained. This is followed by the data capturing process and the limitations of the data.

### 4.1.1 Data source selection

Due to the volume of economic, socio-economic and social unrest variables numerous data sources were required. The data sources that were selected for the economic and socio-economic variables are described first. This is followed with the data source that was selected for the social unrest data.

#### 4.1.1.1 Economic and socio-economic variables

The literature study, described in Chapter 3, proposed a host of economic, demographic, country-specific, socio-economic, international and other variables that could play a role in unrest events. The range of variables that were proposed were so vast that multiple data sources were required, as the array of variables in a single database do not cover all the categories that were mentioned in the literature.

Data from eight accredited organisations, all with publicly available data, were combined to create a data set covering the necessary variables. These organisations include Transparency International (1995-2016), the Freedom house (2017), the Heritage Foundation (2017), Statistics South Africa (Stats SA) (2013-2016, 2018), the United Nations Development Program (UNDP) (2017), the Food and Agriculture Organisation of the United Nations (FAO) (2017-2018), the Republic of South Africa’s Department of Energy (1997-2016), the South African Reserve Bank (SARB) (2017) and the World Bank (2017-2018). Table 4 shows which of the economic and socioeconomic variables are included in the new database and from where the data were obtained. The full list of the 80 references that were used for the creation of this data base is given in the reference list for data sources.

*Table 4: Economic and socioeconomic variables included in the new database and the data source it was obtained from.*

| | | Data Included | Data Source | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | World Bank | FAO | Freedom House | Heritage Foundation | SARB | Stats SA | Transparency Int | UNDP | RSA Department of Energy |
| Eco-nomic | Economic growth | ✓ | ✓ | ✓ | - | - | ✓ | - | - | - | - |
| | Inflation | ✓ | ✓ | - | - | - | - | ✓ | - | - | - |
| Demo-graphic | Urbanisation | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Population Size and Growth | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Ethnic Heterogeneity | ✗ | - | - | - | - | - | - | - | - | - |
| Country-specific | Type of Regime | ✗ | - | - | - | - | - | - | - | - | - |
| | Government Policies | ✓ | ✓ | - | - | ✓ | ✓ | - | - | - | - |
| | Corruption | ✓ | - | - | - | - | - | - | ✓ | - | - |
| | Dependence on Natural Resource | ✗ | - | - | - | - | - | - | - | - | - |
| | Homeownership | ✗ | - | - | - | - | - | - | - | - | - |
| Socio-economic | Age Dependency Ratio | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Unemployment | ✓ | ✓ | - | - | - | ✓ | - | - | - | - |
| | Poverty | ✓ | ✓ | - | - | - | - | ✓ | - | - | - |
| | Income Growth | ✓ | ✓ | - | - | - | ✓ | - | - | - | - |
| | Segregation | ✗ | - | - | - | - | - | - | - | - | - |
| | Vertical Inequality | ✓ | ✓ | - | - | - | - | ✓ | - | - | - |
| | Horizontal Inequality | ✗ | - | - | - | - | - | - | - | - | - |
| | Migration | ✗ | - | - | - | - | - | - | - | - | - |
| | Food Prices | ✓ | - | ✓ | - | - | - | - | - | - | - |
| | Human Development and Education | ✓ | ✓ | - | - | - | - | ✓ | - | ✓ | - |
| Inter-national | Technological Developments | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Mass Media and Modes of Communication | ✓ | - | - | ✓ | - | - | - | - | - | - |
| | Globalisation | ✗ | - | - | - | - | - | - | - | - | - |
| Other factors | Size of the Government | ✗ | - | - | - | - | - | - | - | - | - |
| | Oil Price Increases | ✓ | - | - | - | - | - | - | - | - | ✓ |
| | Level of Trade Openness | ✓ | - | - | - | ✓ | - | - | - | - | - |
| | Natural Disasters | ✗ | - | - | - | - | - | - | - | - | - |
| | Access to Water | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Failed Infrastructure | ✗ | - | - | - | - | - | - | - | - | - |
| | Exchange Rate Fluctuations | ✓ | - | - | - | - | - | - | - | - | ✓ |

The data obtained from these eight organisations compromised of both national and provincial data. The national data were captured in monthly, quarterly and annual categories, based on data availability. For provincial data only the annual data that was available for all nine provinces were captured. This is due to the limited availability of information in the public domain. A breakdown of the sources for national and provincial data that were collected can be seen in Table 5.

*Table 5: A breakdown of the national and provincial data that was collected and the time intervals between consecutive data points for the economic and socio-economic variables.*

| | | National | | | Provincial (Annually) |
|---|---|---|---|---|---|
| | | Annually | Quarterly | Monthly | |
| Economic | Economic Growth | ✓ | - | - | - |
| | Inflation | ✓ | ✓ | ✓ | - |
| Demographic | Urbanisation | ✓ | - | - | - |
| | Population Size and Growth | ✓ | - | - | ✓ |
| Country | Government Policies | ✓ | ✓ | ✓ | - |
| | Corruption | ✓ | - | - | - |
| Socio-economic | Age Dependency Ratio | ✓ | - | - | - |
| | Unemployment | ✓ | ✓ | - | - |
| | Poverty | ✓ | - | - | - |
| | Income Growth | ✓ | ✓ | - | - |
| | Vertical Inequality | ✓ | - | - | ✓ |
| | Food Prices | ✓ | ✓ | ✓ | - |
| | Human Development and Education | ✓ | - | - | ✓ |
| International | Technological Developments | ✓ | - | - | - |
| | Mass Media and Modes of Communication | ✓ | - | - | ✓ |
| Other factors | Oil Price Increases | ✓ | ✓ | ✓ | - |
| | Level of Trade Openness | ✓ | - | - | - |
| | Access to Water | ✓ | - | - | ✓ |
| | Exchange Rate Fluctuations | ✓ | ✓ | ✓ | - |

#### 4.1.1.2 Social unrest

There are only two databases that collect data about political violence, social unrest and/or protest events in South Africa. These two databases are the Armed Conflict Location & Event Data Project (ACLED) and the Institute of Safety and Security (ISS). The format of both databases is described below. Thereafter, the databases are compared with one another and the data selection is explained.

##### 4.1.1.2.1 Armed Conflict Location & Event Data Project

ACLED collects all the political violence events that are reported in newspapers for more than 60 countries in Africa and Asia. This includes Burundi, Cambodia, Kenya, India, Mali, Myanmar, Nepal, Sierra Leone, South Africa and Vietnam.  The definition of political violence that is used by ACLED is quite general and includes, but is not limited to, assassinations, normal social unrest/protests, xenophobia, battles and any events that relate to political parties or their members. ACLED's full database can be found on their website (http://www.acleddata.com/).

The specific database pertaining to South Africa starts in 1997 and is consistently being updated. This database has a lag of less than a month. The date, the number of fatalities and the exact GPS location of each event is recorded in chronological order. Additional information such as the type of event (which ranges from assassinations, battles, civilian killings, protests, riots and recruitment activities) as well as the parties involved, and a brief description of the event is also available. This brief description gives the context of the political violence event that occurs as it usually describes other factors like police presence, injuries, property damage and the underlying reason for the political violence event. The exact format of ACLED's data set can be seen in Figure 8.

| GWNO | EVENT_ID_CNTY | EVENT_ID_NO_CNTY | EVENT_DATE | YEAR | EVENT_TYPE | ACTOR1 | ACTOR2 | COUNTRY | ADMIN1 | ADMIN2 | ADMIN3 | LOCATION | LATITUDE | LONGITUDE | SOURCE | NOTES | FATALITIES |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 560 | 1SAF | 85780 | 05/01/1997 | 1997 | Remote violence | BAT: Boere Attack Troop | Civilians (South Africa) | South Africa | North West | Bojanala | Rustenburg | Rustenburg | -25.66667 | 27.25000 | Reuters News | The Boere Attack Troop, a far right group, detonates three explosives near a Mosque, post office, and a local shop. Extensive property damage, three children and an adult killed. | 4 |
| 560 | 2SAF | 85781 | 09/01/1997 | 1997 | Riots/Protests | Protesters (South Africa) | | South Africa | North West | Bojanala | Rustenburg | Rustenburg | -25.66667 | 27.25000 | Reuters News | The NUM, the mining union which the Rustenburg Mosque bombers belong, hold a mass protest to their actions. | 0 |
| 560 | 3SAF | 85782 | 14/01/1997 | 1997 | Strategic development | Unidentified Armed Group (South Africa) | Police Forces of South Africa (1994-) | South Africa | KwaZulu-Natal | eThekwini | Ethekwini | Amanzimtoti | -30.05000 | 30.88333 | Reuters News | A group of unknown men place a bomb in a bank. The local police are able to defuse the explosive. | 0 |
| 560 | 4SAF | 85783 | 01/02/1997 | 1997 | Battle-No change of territory | Unidentified Armed Group (South Africa) | Police Forces of South Africa (1994-) | South Africa | Gauteng | City of Johannesburg | City of Johannesburg | Johannesburg | -26.20227 | 28.04364 | Reuters News | A group of robbers steal 50 guns from a police station. | 0 |
| 560 | 5SAF | 85784 | 06/02/1997 | 1997 | Riots/Protests | Protesters (South Africa) | Police Forces of South Africa (1994-) | South Africa | Gauteng | City of Johannesburg | City of Johannesburg | Soweto | -26.26667 | 27.86667 | Reuters News | A group of demonstrators blocked roadways in protest of the threat of civic services to be cut due to non-payment and increasing rent. Police disperse the crowd with tear gas and rubber bullets. A seven year-old-boy is trampled to death amid the chaos, and others die | 4 |
| 560 | 6SAF | 85785 | 06/02/1997 | 1997 | Riots/Protests | Protesters (South Africa) | | South Africa | Gauteng | City of Tshwane | City of Tshwane | Pretoria | -25.74486 | 28.18783 | BBC Monitoring Service | COSATU members protest outside the Swaziland High Commission in Pretoria, demanding the release of arrest trade union members taken during a recent protest for democracy in that country. | 0 |
| 560 | 7SAF | 85786 | 07/02/1997 | 1997 | Riots/Protests | Protesters (South Africa) | | South Africa | Western Cape | City of Cape Town | City of Cape Town | Cape Town | -33.92528 | 18.42389 | Reuters News | Demonstrators gather during a speech by President Mandela to protest what they see as a lack of AIDS research by the government. | 0 |

*Figure 8: An example of ACLED's data base for public unrest (Raleigh et al., 2010).*

#### 4.1.1.2.2 Institute of Safety and Security

The second available database, ISS, only considers incidents of protest or public violence events that occurred in South Africa from 2013 and have been reported in newspapers. Their database, which can be found on their website (https://issafrica.org/crimehub/maps/public-violence), is updated consistently, however, there is a lag of 3 months. The public violence data are plotted on a map and each event that occurred is represented by a dot on the exact location where the event took place. The data pertaining to each event is displayed when a person clicks on the respective dot, as shown in Figure 9. The ISS's records are more detailed than ACLED's data as it not only records the date, location and number of fatalities but also classifies the underlying reason for the protest. It also records the number of fatalities, whether it was a legal or illegal protest, whether it was peaceful or violent, the size of the protest, whether there was any police intervention as well as the number of people who were arrested.

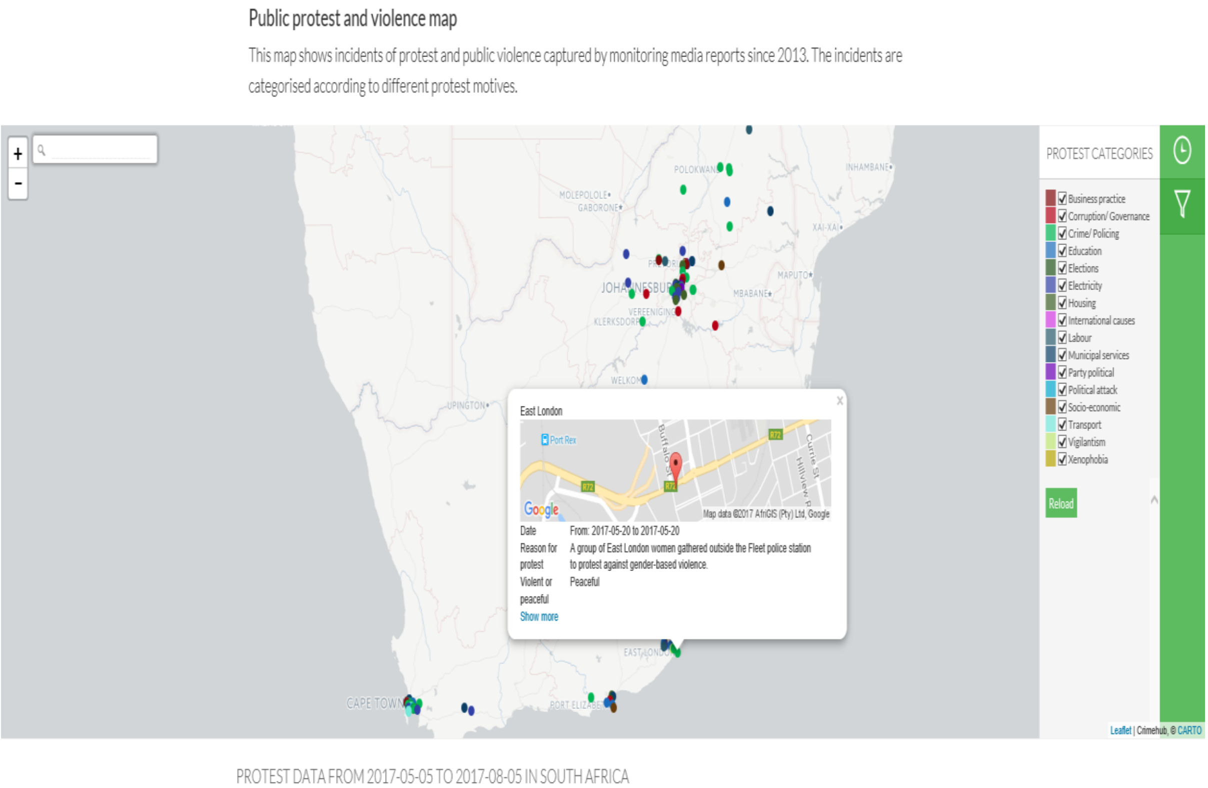



*Figure 9: Example of ISS's data base for public unrest (ISS Crime Hub, 2017).*

#### 4.1.1.2.3 Social unrest data selection

Although the ISS's data are more detailed than ACLED's data the quantity of data available is insufficient, as it only consists of 4 years' data. For this reason, ACLED's database was selected as the data source to ensure sufficient data availability in later analysis.

The ACLED's data set cannot be used in its original form as it is not completely categorised and includes political violence events that are not considered in this study. For this reason, the data pertaining to protests was extracted and then looked at further. The brief description of each protest event was used to extract the additional information in a format that is compatible with statistical analysis software.

The specific data that was looked at includes the number of injuries, whether there was any property damage, the type of protest, the size of protest, whether the protest was legal or illegal, violent or peaceful, whether the protesters committed any crimes and what they were, the number of arrests made, whether the police intervened and the methods used by them and whether it was an ongoing and/or contagious protest. This additional extracted information results in the adjusted ACLED's database to closely resemble the ISS database. For this study, the definitions for a protest and a social unrest event is the same and these two words are used interchangeably.

### 4.1.2 Data capturing and conversion to social unrest data

The data capturing was done with the help of nine actuarial mathematics undergraduate students from the University of Pretoria. These nine students were carefully selected out of 30 students. Each student received a sample data set which contained 12 entries of which two were completed and the other 10 was left for each student to complete. The completed sample data sets were checked for accuracy and completeness while considering the amount of time that it took the students to complete the 10 entries. The nine students who best completed the sample data set were then selected. Each of these students was assigned with between 800-1000 data points. The number of data points was also divided based on the accuracy and speed with which the student completed the sample data set. A training session was held with all the students. During the training session, the assumptions and categories were explained to each student hereafter the students received the data set with their assigned data entries. Once all the students completed their assigned work, the data were combined. After this reasonability and consistency checks were performed on the combined data. Additional information about the training session, data capture, the data consolidation process and the reasonability checks can be found in Appendix A.

To ensure that the data were extracted consistently, and that the final product was in a usable format, categories relating to the type of political violence event and reason for protest as well as several assumptions were made. The political violence data were categorised as either a full protest, political assassinations, lone wolf attacks, disputes, gang violence, robbery and/or trespassing, murder, prison and detention centre events, xenophobic attacks, ambiguous entries or double entries. The list of possible reasons for a protest included corruption, crime/anti-crime, education, elections, labour strike, land, municipal service and national causes. The two main assumptions that were used frequently pertained to property damage and violence and are briefly described below. A description of these categories and assumptions is described in Appendix A.

The definition for property damage, that was used in this study, includes anything other than life insurance and healthcare insurance products that can currently be insured in the South African market. In general terms this ranges from infrastructure, property, vehicles, equipment, containers and loss of production to stock. There is however an exception to this. Burning tyres in isolation is not considered as "property damage" although tyres are technically insurable in South Africa. This is because we assume that majority of the tyres that are burnt have been thrown out due to wear and tear before being used in protests. Therefore, these tyres are no longer insured items. This assumption is also logical as it is more likely that protesters will burn the entire car in the heat of the moment than to take the time to find bricks to balance the car and then remove the tyres from the nearby cars before being able to torch them.

Unless stated otherwise it was assumed that a protest was violent if there were any fatalities, police intervention, any violent crimes are committed including arson, burning tyres, throwing stones and/or riots or if there was property damage. If nothing in this regard is mentioned, it is assumed that the protest was peaceful.

### 4.1.3 Limitations of the data

This study makes use of historic data to determine the likelihood of future social unrest events. This creates a good starting point to determine the risk involved, although historic data does not always create a reliable image of what the future may be like. Furthermore, there are time lags in some of the economic data which may lead to errors in the trends and values that are estimated in this study.

There are several different definitions for political violence and social unrest. Only one definition for each of these will be used. As a result, any interpretation of the results as well as any comparison between studies must be done with caution as the results are dependent on the exact definition used.

As both ISS and ACLED only record the daily information pertaining to the protests that occurred, it is impossible to track the hourly movement of a protest. In this most of the detail such as the triggers of violence and the spread of contagion is lost, making it extremely difficult to analyse and model these aspects accurately.

The amount of detail regarding each event also differs. Some events are well documented including the number of persons involved, injured and the exact property damage that occurred, where other accounts are ambiguous or only mention that there was property damage or injuries but never specifying the exact numbers. This was somewhat mitigated by removing ambiguous events (i.e. one protest had three completely different accounts of the events that took place) and by recording the presence of injuries in cases where the exact number of injuries was not specified.

The data are dependent on newspaper reports and this opens up a lot of potential questions regarding the reliability of the description of each protest as well as which protests appear in the newspapers. The description is dependent on the journalist that reports the incident as well as their location when the protests occurred. When large protests occur, they usually fill more than just one street and then the journalist's account will be dependent on the street they were stationed at and the witnesses that were present in the other streets. For smaller protests journalists may not actually be present at the protest and then the entire written account will be based on hearsay. In addition, there may be bias in the written account (this may be from the journalist, newspaper or political climate within the country at the time) that then leads to either under or over reporting of the incident. This risk is somewhat mitigated as several different newspapers are used to find articles relating to protests. A further risk worth mentioning is "fake news" which has recently been identified as a big problem. This risk is somewhat mitigated by using reputable newspapers in searching for articles. There is however, still the possibility that entries based on fake news accounts may have been included in the databases.

As only the protests that appear in reputable newspapers are used to create the initial database it is important to mention that not all protests will be recorded. As a result, the databases only contain a subset of all the social unrest events that occur in South Africa. On a day where there are multiple protests not all protests will appear in the media as the media usually reports the events that may bring the largest media attention which are usually those where there was violence, police intervention, a nationwide reason and/or those that occur in major cities. The portion of the data that is likely not to be included in the database are peaceful protests, small protests or protests regarding a local problem (e.g. the installation of a swimming pool) that easily go unnoticed in the presence of high volumes of violent protests or are reported in local newspapers that are not always recognised or are difficult to access.

It is also extremely difficult to determine the size of the portion of protest events that is not recorded by ISS and ACLED. The only comparative value is the number of public violence events that is recorded in each financial year (1 April to 31 March) by CrimeSA. But there are multiple problems with using these as comparable values. Firstly, no definition for public violence is given which opens a lot of speculation to what is considered as public violence and what is not. Secondly, the first recorded value for CrimeSA is for the 2004/2005 financial year which leaves at least eight years of data that cannot be verified using this method. Furthermore, there are many questions raised about the reliability of the figures published by CrimeSA. Due to this, comparisons between CrimeSA, ACLED and ISS are futile.

In the data from ACLED we see a big increase in the number of reported protests from 2011 to 2012. This increase may be as a result of the number of protests greatly increasing in 2012. However, it may also be as a result of other factors such as better reporting or internet access increasing communication speed and access to newspapers. The exact reason will not be known but it is important to take note of this abnormality within the data.

The economic data that is recorded are annual figures, as the daily and monthly figures for most of these indices are not readily available. This causes a lot of information to be lost as daily or even monthly comparison between the economic factors and the number of protests is not possible. As such there are at most 20 values for each of these variables that can be compared with one another. This lack of data may result in spurious regression being observed in further analysis.

Because of some of the assumptions made and as a result of definitions there are several natural and constructed correlations that are likely to exist. For example, the number of fatalities, the number of injuries and property damage will naturally be correlated with violence and illegal protest. This is because the occurrence of either of these is usually in the presence of violence which, by law, is considered an illegal protest.

There is no way of truly knowing the number or extent of injuries that occurred because of protests. The number of injuries is most likely understated as only the most serious injuries gain attention and are then reported. Furthermore, in cases where individuals get injured while participating in illegal protests there is a motive for these individuals to delay seeking immediate medical attention as this may result in them being arrested.

## 4.2 Modelling techniques

The risks involved with political violence are very similar to catastrophe risks. The difference is that political violence contains an element of human intelligence and human intent (Major, 2002). These characteristics makes it inappropriate to use the central limit theorem, extreme value theorem and the assumption of independent increments as probabilities are insufficient (Major, 2002). In this chapter the models that are applicable to political violence will be looked at in more detail. For completeness, it describes both the models that were used as well as other methods that were available. These models include agent-based modelling (ABM), Game theory, Dynamic Bayesian Networks, Time-Series modelling, linear regression and lagged regression models, the Epidemic-like model that uses Markov chains and finally Scenario analysis.

### 4.2.1 Agent based modelling

Each individual on earth is unique and their interactions with one another are even more so (An, 2012). The behaviour and decisions of individuals are not only affected by their abilities, beliefs and aspirations. It may also be affected by social norms and the reputation of key individuals (An, 2012). Individual behaviour cannot be aggregated when attempting to model the behaviour of a group of people. The way in which an individual makes a decision is very different to the way that groups of people make a decision (An, 2012). ABM is one of the few models that accounts for this. It takes individual social interactions into account to shed light on how they shape both the group behaviour and the social structure that surrounds it (Renn et al., 2011).

ABM is a non-linear bottom-up technique using independent individuals (agents) to establish how decisions are made and interactions performed (Bonabeau, 2002; Renn et al., 2011; An, 2012; Park et al., 2012; Mei et al., 2015). Computerised simulation has been used for applications in political sciences and geopolitics (Bonabeau, 2002; Daras and Mazis, 2014; Daras & Mazis, 2015). The agents are defined based on the real world, allowing for irrational and subjective behaviour, adaption, memory, communication and motion to name a few (Bonabeau, 2002; Viscusi, 2009; An, 2012; Renn et al., 2011; Park et al., 2012; Mei et al., 2015).

ABM is conceptually complicated. It is a flexible technique that can be combined with other modelling techniques (neural networks, game theory, etc.) for more realistic models (Bonabeau, 2002; An, 2012). Decision models (microeconomic, cognitive, experience models, etc.) can be incorporated to mimic the decision-making process. The flexibility of ABM makes it useful in modelling different aspects of social unrest (Bonabeau, 2002). Limited knowledge on individual and group interactions in social unrest scenarios is currently a limiting factor.

### 4.2.2 Game theory

Game theory can be used to quantify the risk of social unrest events. This modelling approach uses rational decisions and strategic interactions. It requires a lot of information pertaining to hazards, risks, locations, financial resources and likelihood of the events happening, to name a few. By using a variety of assumptions for each of the contributing factors a zero-sum game can be set up. This is a model where one group aims to maximise damage and another group aims to minimise it (Major, 2002).

Variations of this approach are agent-based game theory, that removes some of the assumptions used in the zero-sum game model and opponent theory, that considers only one rival (Bonabeau, 2002; Engle, 2008).

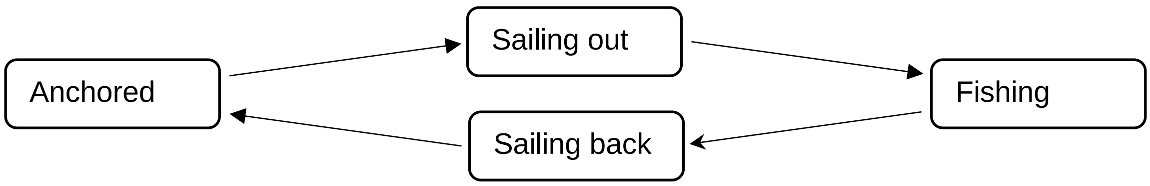
### 4.2.3 Dynamic Bayesian networks

Dynamic Bayesian networks are based on graphical models that are created by combining different layers (Dabrowski and de Villiers, 2015a; Dabrowski and de Villiers, 2015b). These layers may include detection, feature selection/recognition and tracking. It is a multi-agent generative model that can consider different behavioural activities and model complicated actions (Dabrowski and de Villiers, 2015a; Dabrowski and de Villiers, 2015b). This method is usually used for detection and recognition (Dabrowski and de Villiers, 2015a; Dabrowski and de Villiers, 2015b). The possible use of Dynamic Bayesian networks for modelling human behaviour is illustrated by looking at the method that is used to identify pirate vessels in the Gulf of Aden. An application for this modelling technique in the context of political violence events is described in Appendix B.

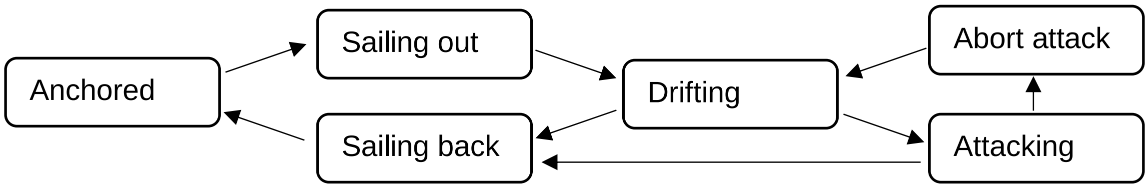
This idea is created under the assumption that there are only three types of vessels that can be found in this region of the ocean, namely a transport vessel, a fishing vessel and a pirate vessel. The movement characteristics of each of the vessels, which forms the basis of the model that detects suspicious behaviour, is summarised in Figure 10. A transport vessel (Figure 10a) sails from one port to another, only moving between two states, either being anchored or sailing. A fishing vessel (Figure 10b) leaves the coastline, sails out to a location (where there are fish), stops to catch fish and returns to the coastline. These vessels move between three states; anchored, sailing and catching fish. The movements of a pirate vessel (Figure 10c) differ considerably from the other two vessels. It leaves the coast and sails out to a location close to the ship routes, once there it may drift, attack, abort an attack or sail back to the coastline depending on the events that unfold. (Dabrowski & de Villiers, 2015a; Dabrowski & de Villiers, 2015b)

The assumptions regarding the movement patterns of the different vessels form the basis of the Dynamic Bayesian Network's ability to identify any suspicious behaviour. It requires real time tracking and comparison to each of the assumptions. Selected features, such as drifting or attacking will then be detected by the model as possible piracy. The model categorises the vessels into three categories, decreasing the number of vessels that need to be monitored to only those vessels that exhibit suspicious behaviour. If closer inspections of these vessels show that the behaviour is not actually suspicious, then it can be recategorised. This allows authorities to advise other vessels in the region about the risky areas or to send resources to the vessels are under treat. (Dabrowski & de Villiers, 2015a; Dabrowski & de Villiers, 2015b)


(a) Movement of a transport vessel




(b) Movement of a fishing vessel




(c) Movement of a pirate vessel

*Figure 10: Diagrams that describe the movements of transport, fishing and pirate vessels (Dabrowski & de Villiers, 2015b).*

### 4.2.4 Time series modelling

Time series models observe chronological data over time to identify any long term, seasonal, cyclical or irregular patterns within the data that are not necessarily visible to the naked eye. These trends are then used to build a model that imitates the behaviour within the data. In addition to chronological data, time series models only require consistency in the duration between observations, making it very versatile (Hamilton, 1994). Furthermore, it can be used to create both univariate and multivariate models with forecasting abilities. One of the limitations, however, is that it may not always be able to account for the demographic or social trends that may be present within the data (Cushing & Poot, 2004).

This technique was chosen because of its flexibility in building models. A brief overview of the models used within this study is given below. An overview of the various univariate models is discussed first. This is followed by the Vector Auto Regressive (VAR) model, a multivariate timeseries model. This is concluded with a description of the model diagnostics that are used to validate the time series models.

#### 4.2.4.1 Univariate time series models

The univariate model which is chosen is determined by the underlying characteristics within the data. The first step is to determine whether the data are stationary or non-stationary. If it is stationary, then an autoregressive moving average (ARMA) model is used (Hamilton, 1994). If the data are non-stationary, then it is further sub-categorised as either trend-stationary, a unit root process or neither.

A process that is classified as trend-stationary contains a deterministic trend which may be caused by a constant mean, linear, parabolic, exponential or seasonal trend within the data (Hamilton, 1994). These trends can be removed with the use of dummy variables, logarithms and cosine and sine functions or a combination of these techniques (Hamilton, 1994). Once the trend is removed, the process becomes stationary and modelled accordingly. The name given to the model depends on the adjustment(s) that were made. These include linear trend, log liner trend, exponential and seasonal models (Hamilton, 1994). In certain instances, these models are improved by the addition of smoothing functions, which incorporate data from more than one point in time. These models include the seasonal exponential trend smoothing model, linear exponential smoothing model, dampened trend exponential smoothing model and the Additive Winters model, amongst others.

A unit root process is made stationary by taking the first or successive differences of the unit root process. The ARMA model is then applied to the resultant stationary process. Finally, if the non-stationary process is classified as neither trend-stationary nor a unit root process, then the autoregressive integrated moving average (ARIMA) model is used (Hamilton, 1994).

#### 4.2.4.2 Vector Auto Regressive model

The equation for a VAR (1) model is depicted by Equation 1.

$$\boldsymbol{Z_t} = \boldsymbol{c} + \boldsymbol{\Phi_1 Z_{t-1}} + \boldsymbol{a_t} \tag{1}$$

where $\{\boldsymbol{Z}_t\}$ is a multivariate time series process, $\boldsymbol{c}$ is a $n \times 1$ vector of constants, $\boldsymbol{\Phi}_1$ is a $n \times n$ matrix containing the autoregressive coefficients and $\boldsymbol{a}_t$ is a $n \times 1$ vector of white noise terms (Hamilton, 1994, p. 259).

The equation for the expanded matrix form of the VAR (1) model is shown in Equation 2.

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{n,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} \emptyset_{11} & \emptyset_{12} & \cdots & \emptyset_{1n} \\ \emptyset_{21} & \emptyset_{22} & \cdots & \emptyset_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \emptyset_{n1} & \emptyset_{n2} & \cdots & \emptyset_{nn} \end{pmatrix} \begin{pmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ \vdots \\ Z_{n,t-1} \end{pmatrix} + \begin{pmatrix} a_{1,t} \\ a_{2,t} \\ \vdots \\ a_{n,t} \end{pmatrix} \qquad (2)$$

The first component of $\boldsymbol{Z_t}$ (for the first of the $n$ equations) is depicted by Equation 3.

$$Z_{1,t} = c_1 + \emptyset_{11} Z_{1,t-1} + \emptyset_{12} Z_{2,t-1} + \cdots + \emptyset_{1n} Z_{n,t-1} + a_{1,t} \qquad (3)$$

**4.2.4.3 Model diagnostics**

If the correct model is fitted then the residuals (difference between the actual and predicted values) should be uncorrelated, have a normal distribution, have a zero mean, be identically distributed and have a constant variance. To determine whether this is true, model diagnostics which include Durbin-Watson, Ljung-Box test for white noise and the unit root test are performed. A brief overview of each of these tests are described below.

The Durbin Watson test specifically tests for the presence of autocorrelation at lag 1. The null and alternative hypotheses for this test are:

$$H_0: \rho_1^{residual} = 0$$

$$H_A: \rho_1^{residual} \neq 0$$

If the test statistic is close to 2, then the residuals at lag 1 are uncorrelated (Durbin and Watson, 1950; Durbin and Watson, 1951)

The Ljung-Box test is used to determine whether the residuals form a white noise sequence. The null and alternative hypotheses for this test are:

$$H_0: \rho_1^{residual} = \rho_2^{residual} = \ldots = \rho_K^{residual} = 0 \qquad \text{for } k = 1, 2, \ldots, K$$

$$H_A: \text{At least one } \rho_K^{residual} \neq 0 \qquad \text{for } k = 1, 2, \ldots, K$$

If the residual series is white noise, then it means the sequence is identically distributed with zero mean and constant variance, both of which are constant over time (Ljung and Box, 1978).

The Augmented Dicky-Fuller Unit Root Test is used to determine whether a series is stationary or whether a unit root is present. The data are stationary if $H_0$, shown below, is rejected (Dickey and Fuller, 1979). This test makes use of the following null and alternative hypotheses are:

$$H_0: d=1$$

$$H_A: d=0$$

The residuals compatibility with the Normal distribution can be verified using hypothesis testing. Kolmogorov-Smirnov test, Cramer-von Mises test and Anderson-Darling test are three of the hypothesis tests that are often used to test for normality. The specific null hypothesis depends the applicable hypothesis test used (Yazici and Yolacan, 2007).

The final model diagnostic is the Granger-causality Walt test. This test is specifically used for the VAR model. It tests for causal relationships between variables to ensure that only variables that cause a change in the value of the dependant variable are included in the model (Hamilton, 1994).

### 4.2.5 Linear regression and lagged regression models

Linear regression and lagged regression models are based on the general linear model. The equation for a general linear model is showed in Equation 4 (Wackerly et al., 2008).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \qquad (4)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are unknown parameters, $x_1, x_2, \dots, x_n$ is a set of independent variables and assume known values, $Y$ is dependent on the set $x_1, x_2, \dots, x_n$ and ε is a random error term with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$ (Wackerly et al., 2008).

The lagged regression model makes use of an adjusted form of the general linear equation. This equation allows the independent variables to use both current and lagged (past) values. Equation 5 shows the adjusted equation that is used for the lagged regression model.

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \dots + \beta_n x_{t-n} + \varepsilon_t \qquad for\ t = 1, 2, \dots, T\text{-}n \qquad (5)$$

where $T$ is the maximum number of points in time, $n$ is the lag length, $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are unknown parameters, $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-n}$ is a set of independent variables and assume known values. $Y_t$ is the value of the dependent variable $Y$ at time t and is dependent on the set $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-n}$. $\varepsilon_t$ is a random error term with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma^2$ (Wackerly et al., 2008).

### 4.2.6 Epidemic-like model using Markov chains

Markov chains are based on the main assumption that the conditional probability of an event to occur is only dependent on current conditions. This allows for the conditional probability to change along with changing conditions and makes this technique very useful in scenarios, like social unrest, and financial markets, where past data are not always the best indicator of future behaviour. (Nizamani, Memon and Galam, 2014).

Within communities, hatred and public outrage spreads in the same way in which rumours or epidemics do. An epidemic or contagion model can therefore be used to build a public violence model, by treating hatred and public outrage as a virus. Differential equations based on this model are created and then used to determine the conditional probability of public outrage/unrest events. (Nizamani et al., 2014)

This concept is illustrated in Figure 11, where individuals are subdivided into one of only five states (upset, violent, sensitive, immune and relaxed). The movements between states are limited to those illustrated in Figure 11. These transitions form the basis of the differential equations which determines the transition probabilities (the likelihood that and an individual will move from one state to another) between different states. These transition probabilities can then be used to determine the approximate time frame when individuals enter and exit the violent category. Thus, approximating both when spats of public violence are likely to occur and how long they will continue. (Nizamani et al., 2014)

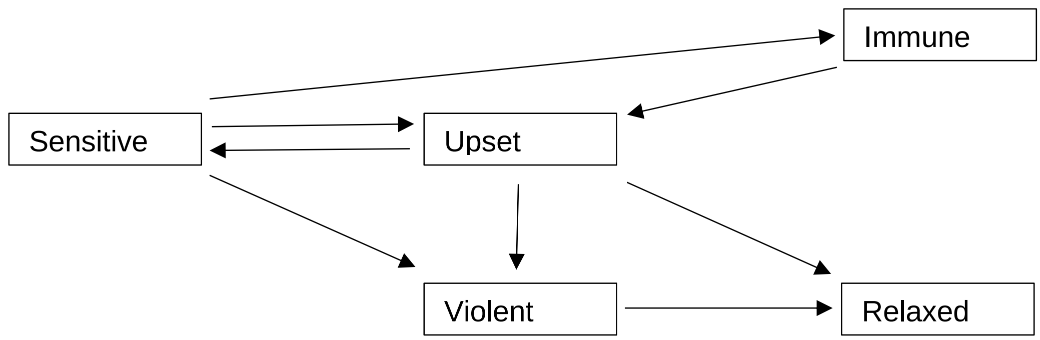



*Figure 11: A stochastic model that can be used to model the way in which public outrage leads to eruptions of public violence (Nizamani et al., 2014).*

### 4.2.7 Scenario analysis

Scenario analysis does not extrapolate from the past but rather attempts to imagine how the future will be (Levinsohn, 2002; McCreight, 2013). It is particularly beneficial for low frequency high impact risk, where there is not a lot of data to work from (Andersen, 2005; McCreight, 2013). This method helps with public policies and understanding the hidden and obvious dynamics which are usually underestimated, making it particularly useful in conflict simulation and war games (McCreight, 2013). Scenario analysis techniques attempt to ensure that the simulated situation is as realistic as possible. This ensures that the results are close to what would occur in real life as it incorporates real-time pressure and distractions that people would be under (McCreight, 2013).

### 4.2.8 Comparison and selection of modelling approach

In general, there are two approaches that can be used to model different types of political violence. The first is a bottom-up approach which starts with the individual components of human intent and intelligence and then works its way up until a global idea is created. The other is a top-down approach which starts off with a global image and works down from there. Each has their own advantages and disadvantages which can make them very good solutions for different types of problems. For this reason, the approaches need to be analysed so that the best approach for modelling political violence in South Africa can be chosen.

The advantage of a bottom-up approach is that it allows the researcher to model scenarios where there is very little or no data. It also allows the researcher to model parts of the problem, gaining insight into the inner workings of the problem at hand. However, a good basis of assumptions is vital for the success of the model. In scenarios where there is enough literature, this is not a problem because there is a good starting point from which assumptions can be made. In this case the literature has shown that there is very little consistency in the results from studies that were performed in different countries. Furthermore, very little research has been done on the relationships that exist in South Africa. Therefore, there is no clear starting point that can be used to make the assumptions. This may lead to subjectivity in the simulated data as the data simulated will only be as good as the assumptions that were made in each of the models.

A deep understanding of numerous different networks will be required to ensure that the model accurately portrays the actual interactions that would take place. The model will also have to be monitored to ensure that the assumptions remain relevant while accounting for changes in human behaviour and the introduction of new actors. Any personal data that is used and not simulated may be seen as unethical or unlawful, as it would infringe on people's right to privacy. Furthermore, this method will require large amounts of computing power, which is very time consuming. A proposed approach that can be used for future bottom-up modelling of political violence can be seen in Appendix B.

The advantage of a top-down approach is that it makes use of existing data and does not require as many assumptions as a bottom-up approach. It is however more of a black box approach, which does not give a lot of insight into the core of the problem. The results are also only as good as the data that is available. As South Africa has data available for a few of the political violence categories, the top-down approach was a natural and simple starting point to gain insight in the matter.

The definition for political violence is extremely broad and the dynamics of the different political violence categories are all distinctly different and therefore political violence cannot be modelled as a whole (Alexander, 2010). In South Africa social unrests, riots and communal conflicts are quite common, which has allowed researchers to build a database for these types of events, namely ACLED. One of the most complete data bases in South Africa is for social unrest. Therefore, the focus from this point on is social unrest in South Africa. A time series approach, as described in Section 4.2.4, will be used to model the data.

## 4.3 Data analysis

The data analysis that was performed on the protest, economic and socioeconomic data, described in Table 5, made use of different approaches. The data were first analysed by making use of Spearman's rank correlation and trendline analysis. Many of the studies described in the literature made extensive use of Spearman's rank correlations and it is important to allow for this data to be compared to what was observed in the literature. The trendline analysis allows the data to be visualised in a more intuitive manner and gives insight into additional effects that are present in the data.

The next phase of data analysis made use of modelling techniques. These techniques describe the data and determine how the various variables influence the number of protests, both for different time frames and different models. The methods used to support these analyses are discussed below in two different sections

### 4.3.1 Spearman's rank correlation and trendline analysis

Spearman's rank correlation determines whether a monotonic relationship exists between two non-parametric variables (Wackerly, Mendenhall and Scheaffer, 2008). SAS 9.4 was used to calculate the Spearman's rank correlation coefficient. This was used as the basis to compare the direction and strength of the relationships that exist between protests and the economic and socio-economic variables. Spearman's rank correlation was performed on national and provincial data to determine whether differences exist between the national and provincial relationships. Cross correlations were also analysed but will not be discussed.

National social unrest, economic and socio-economic data are available for multiple time intervals. All the monthly, quarterly and annual data were collected. Spearman's rank correlation was performed on each of these three-time intervals to determine whether the relationships are sensitive to changes in the data frequency used.

The economic and socio-economic variables for all nine provinces in South Africa are not as readily available as the national data. Only annual socio-economic data were obtainable for all nine the provinces. This makes it difficult to test the independent relationships within each province. For this reason, the socio-economic and corresponding protest data for each of the nine provinces were combined into one data set and analysed accordingly. This analysis was used to shed light on the volatile relationships within the provincial data that may not be apparent within the national data. This data set was analysed using Spearman's rank correlation to determine the provincial relationships.

Trendline analysis made use of scatter graphs drawn for all the variables where data were available for multiple time intervals. Using Microsoft Excel linear, logarithmic, quadratic, power and exponential trendlines, based on best fit, were plotted on the scatterplots to find the trendline best describing the relationship between variables. These trendlines were compared with one another to determine whether changes in data frequency affect the relationships present. The same was done to compare the relationships between national and provincial data.

### 4.3.2 Modelling the incidence of social unrest events

Daily, monthly, quarterly and annual protest models were created to enhance the understanding of the behaviour of South African social unrest events. The number of monthly and quarterly protests were analysed using four different models: univariate time series, linear regression, lagged regression and VAR (1). Each of these models considers different interactions between the incidence of social unrests events and various economic and socio-economic variables, allowing greater insight into the interactions that exist. The number of daily and annual protests were only modelled with univariate time series models due to the unavailability of daily economic data and limited annual data points. Due to the small sample sizes, out of sample testing was not performed on any of these models.

SAS version 9.4 was used for the multiple regression, linear regression and VAR (1) models. While the univariate time series models were fitted by making use of the SAS version 9.4 Time Series Forecasting System.

### 4.3.3 Testing models forecasting abilities

The out-of-sample procedures to test both the long term and short-term forecasting abilities of the models are briefly described below. The main difference between these two techniques is the number of forecasts produced prior to recalibration.

A one-step forecasting procedure was used to test the model's short-term forecasting abilities (Tashman, 2000). A portion of the data are removed and referred to as the hold out sample, the remainder of the data are referred to as the in-sample data. A model is created based on the in-sample data. This model is used to forecast one step into the future (Tashman, 2000). The actual data point that corresponds to the point that has just been forecasted is added to the in-sample data set. Using the "new" in-sample data set, that contains the additional datapoint, the model is recalibrated, and a forecast is produced using the recalibrated model. This procedure of adding a data point, recalibrating the model and forecasting the next data point is repeated for each data point in the hold out sample (Tashman, 2000).

An n-step forecasting procedure, where n is equal to the length of the hold out sample, was used to test the model's long-term forecasting ability. As is done in the one-step forecasting procedure, a portion of the data are removed from the data set. A model is created using the in-sample data and *n* forecasts, one for each data point in the hold out sample, are created using this model (Tashman, 2000). With this process the hold out sample is always kept separate from the in-sample data used to create a model and the model is not recalibrated (Tashman, 2000).

Once the forecasts have been created the two techniques are validated in a similar fashion. The forecasted values are compared to the actual values and the error, absolute error, Mean Absolute Error (MAE) and Root Mean Absolute Error (RMAE), depicted by Equation 6, 7, 8 and 9 respectively, are calculated (Salazar-Moreno, López-Cruz and Sánchez Cruz, 2019).

$$Error = \sum_{t=1}^{N}(A_t - F_t) \tag{6}$$

where $N$ is the size of the hold out sample, $A_t$ is the actual value at time $t$ and $F_t$ is the forecasted value at time $t$.

$$Absolute\ error = \sum_{t=1}^{N}|A_t - F_t| \tag{7}$$

where $N$ is the size of the hold out sample, $A_t$ is the actual value at time $t$ and $F_t$ is the forecasted value at time $t$.

$$MAE = \frac{\sum_{t=1}^{N} |A_t - F_t|}{N} \tag{8}$$

where $N$ is the size of the hold out sample, $A_t$ is the actual value at time $t$ and $F_t$ is the forecasted value at time $t$.

$$RMAE = \sqrt{MAE} = \sqrt{\frac{\sum_{t=1}^{N} |A_t - F_t|}{N}} \tag{9}$$

where $N$ is the size of the hold out sample, $A_t$ is the actual value at time $t$ and $F_t$ is the forecasted value at time $t$.

All the data pertaining to 2016 were removed from the database. This portion, which constituted 5% of the data, formed the hold out sample used to analyse the long term and short-term forecasting abilities of the models. There was a large change in the incidence of protests in the last five years data. Therefore, a sufficient amount of time after this change in the incidence of protests had to be given so that the models could adjust to the new protest levels. As a result, the hold out sample could not be increased.

Validation was performed using a combination of SAS 9.4 and Microsoft Excel. SAS 9.4 was used to create the models, while Microsoft Excel was used to calculate the error, absolute error, MAE and RMAE of each model.

### 4.3.4 Modelling subcategories of protests

Five subcategories of protests were evaluated to determine whether the natural trends within these subcategories were different to those seen in all the protests. The subcategories that were selected considered the incidence of violence, protests with property damage as well as the three main reasons for protests; which were education, labour and municipal service-related protests.

The statistical analyses performed on the subcategories were similar to what was done for all the protests. The interactions between the subcategories and the economic and socio-economic variables were evaluated using Spearman’s rank correlation. Each of the subcategories were modelled using univariate time series, multiple regression and VAR (1) models. Due to the availability of data, daily, monthly, quarterly and annual univariate time series models were created. Linear regression and VAR (1) models were only created for monthly and quarterly data. Due to the small sample sizes, out of sample testing was not included in modelling any of the protest subcategories.

# Chapter 5: Results and Discussion

This chapter comprises of three parts. The first section examines and analyses the social unrest database for South Africa. The second part discusses Spearman's rank correlation and trendline analysis to look at the relationships that are present between protests and the various economic and socio-economic variables. In the third part, protests are modelled. The model validation of these models is described in the fourth part. The final part looks at the characteristics of subcategories of protests.

## 5.1 Description of the social unrest database

The ACLED database contains the political violence information that pertains to South Africa. This includes, but is not limited to, the social unrest data that is of interest in this study. By looking at the different categories present in the ACLED database and then specifically at its social unrest information, a lot of insight can be gained.

### 5.1.1 Description of ACLED political violence data

In the original ACLED database there are a total of 8,897 recorded political violence events that occurred in South Africa from the start of 1997 to the end of 2016 (Figure 12). Prior to 2012, the highest number of political violence events that where recorded in a year was 413. Between 2011 and 2012 there was a tremendous shift in the number of political violence events recorded. The value increased more than threefold from 319 in 2011 to 1,248 in 2012. Since this shift, the number of recorded events per year has remained extremely high, with all the recorded annual values being higher than 1,100, with the highest recorded value of 1,520 political violence events being seen in 2015.

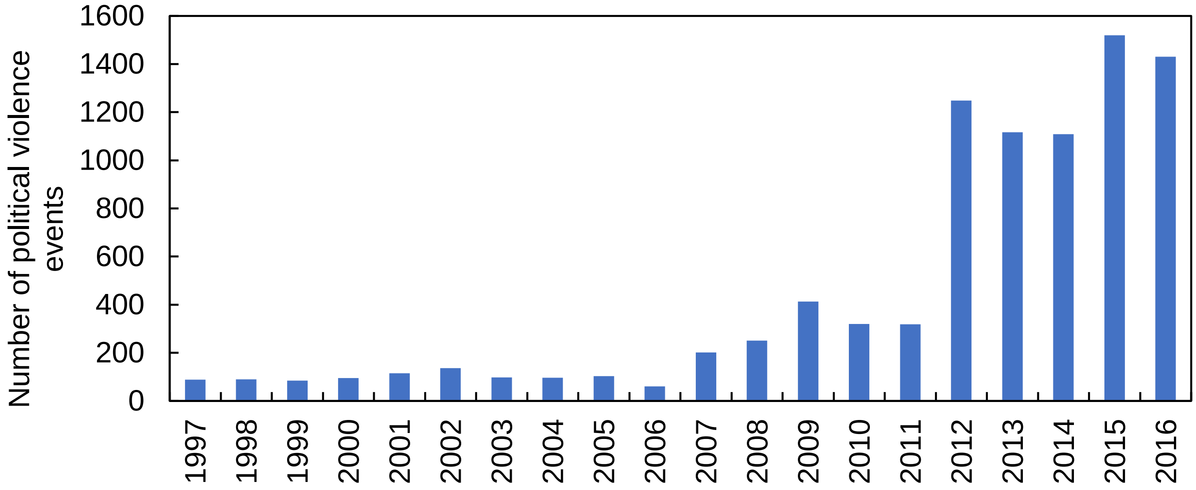



*Figure 12: The number of political violence events in South Africa per year that have been recorded by ACLED.*

Of the 8,897 political violence events, 7,172 (80.6%) were full protests (or social unrest events), 494 (5.6%) were related to mob justice and/or gang violence, while a total of 295 (3.3%) political killings or clashes were recorded. The exact categorisation of the ACLED data set, among the eleven subcategories of political violence, is shown in Table 6. The full definitions for each category in Table 6 is given in Appendix C.

*Table 6: The categorisation of ACLED's political violence events recorded in South Africa.*

| Category | Number of Events | Portion of Data (%) |
|---|---|---|
| Protest | 7172 | 80.6 |
| Political Killings and Clashes | 295 | 3.3 |
| Lone Wolf and Revenge Attacks | 74 | 0.8 |
| Disputes | 119 | 1.3 |
| Gang Violence and Mob Justice | 494 | 5.6 |
| Robbery and Trespassing | 16 | 0.2 |
| Murder | 161 | 1.8 |
| Prison Events | 37 | 0.4 |
| Ambiguous | 202 | 2.3 |
| Double Entry | 173 | 1.9 |
| Xenophobia | 154 | 1.7 |

Only the protests, as defined the Appendix A, are used for the remainder of this dissertation. All the other categorisations of political violence therefore been removed from the data set. Figure 13 shows the proportion of the annual data related to full protests. The data that related to the 10 other political violence subcategories, have been removed.

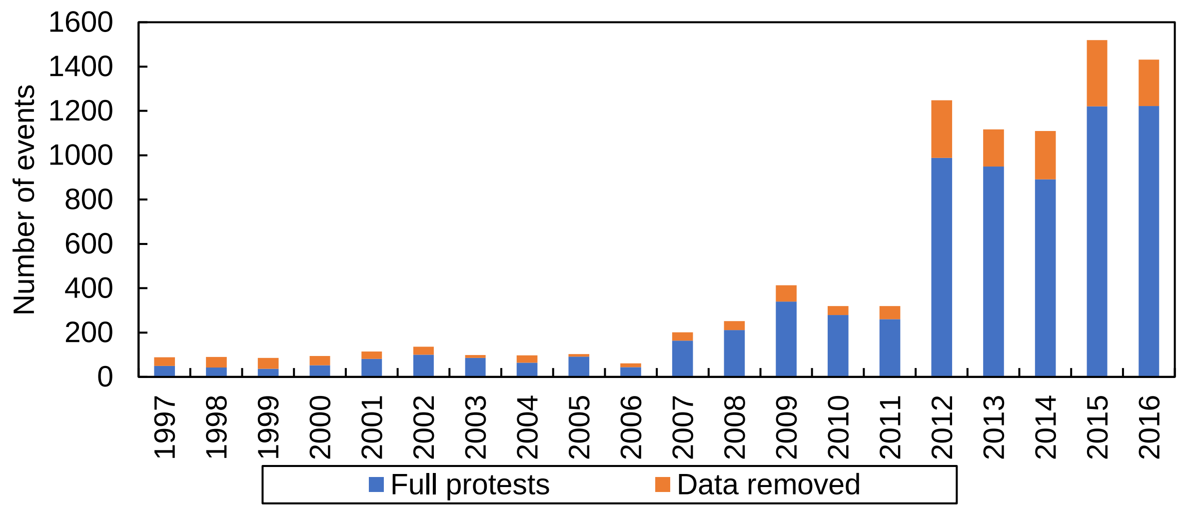



*Figure 13: The number of political violence events in South Africa per year split by full protests and other forms of political violence.*

### 5.1.2 Description of Social unrest data

The social unrest section of the ACLED database shows a steady increase in the number of protests per year over the past 20 years, with an enormous increase occurring in 2012. This is similar to the annual behaviour observed when the whole political violence data set was considered.

Figures 14, 15 and 16 show the number of protests per quarter, month and day, respectively. Since the start of 2012 there have been more protests that occur on any given day and there have been fewer days where no protests were recorded. The impact of this increase in both the frequency and regularity of protests is clearly evident in both the quarterly and monthly graphs (Figure 14 and 15).

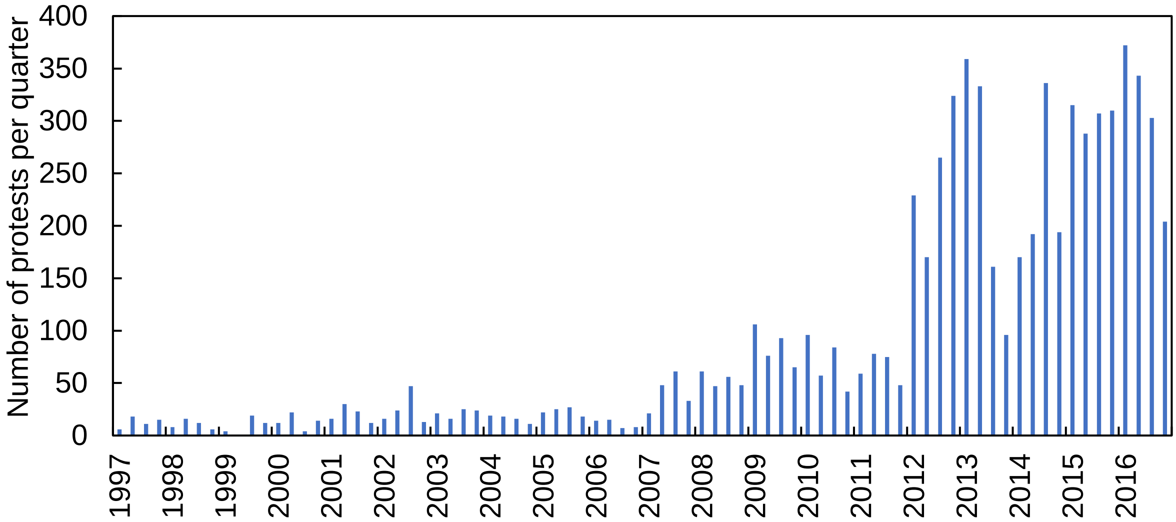



*Figure 14: The number of protests in South Africa per quarter.*

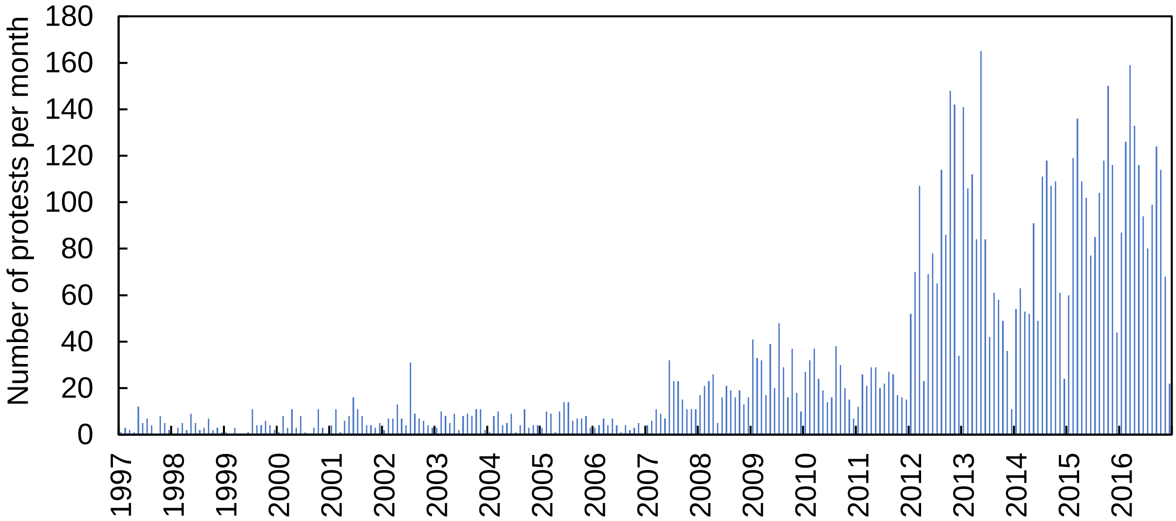



*Figure 15: The number of protests in South Africa per month.*

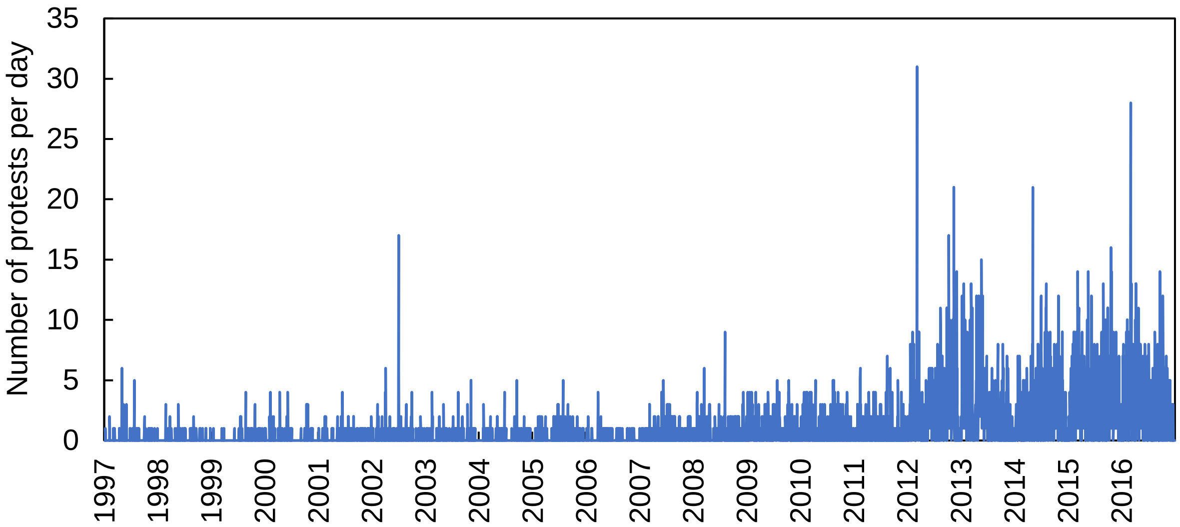



*Figure 16: The daily number of protests in South Africa.*

Over the 20-year time frame from 1997 to 2016, there was an average of 358.6 protests per year (average of 0.98 protests per day). In Figure 17 it is clear that the number of protests per day follows an exponentially decreasing function, with fewer protests per day being more probable. During this time, on 63.1% of the days no recorded protests occurred, 16.7% of the days had one recorded protest, 7.3% of the days had two recorded protests, while 12.8% of the days had three or more recorded protests.

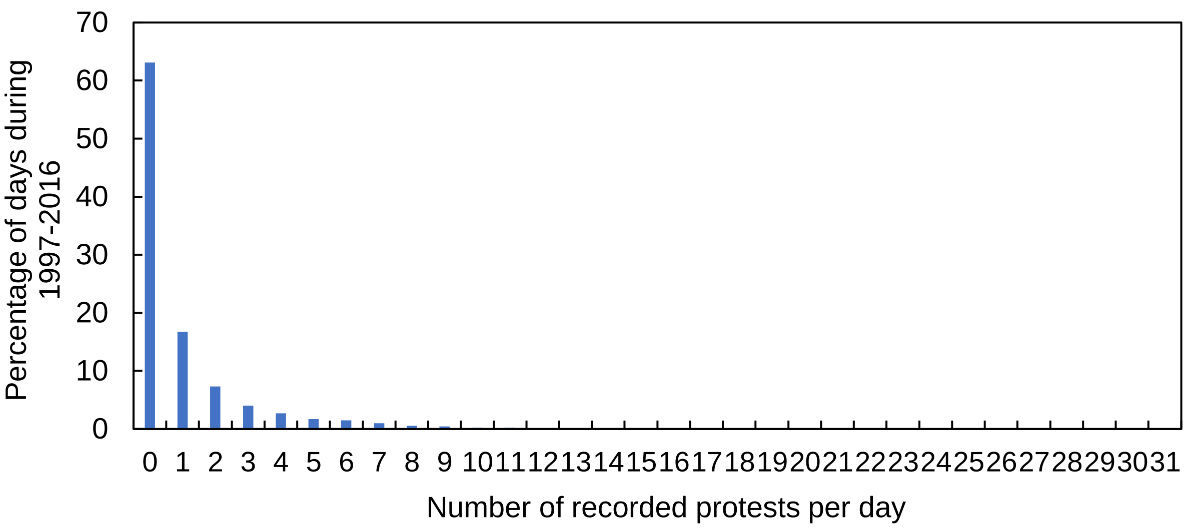



*Figure 17: The distribution of the number of protests recorded in South Africa for each day from 1 January 1997 to 31 December 2016.*

There was a large increase in the recorded number of protests that occurred in the last five years (from 2012-2016). During these five years there was an average of 1054.2 protests per year (an average of 2.89 protests per day). This is nearly three times higher than what was seen over the 20-year time frame. The distribution of the number of protests per day, shown in Figure 18, also has an exponentially decreasing function, but its decay is much slower than what was seen in Figure 17. In this period, only 21.4% of the days had no recorded protests, 19.5% of the days had one recorded protest, 15.8% of the days had two recorded protests, while 43.2% of the days had three or more recorded protests. These figures, in comparison to those seen in Figure 17, is indicative of a change in the patterns of protests in South Africa.

To put all of this into context, there were only 390 days where no protests were recorded. If all the no protest days were to be consecutive, this would amount to 1 year and 25 days without any protests. While an average of 3.67 protests per day would be recorded on the remaining 3 years, 11 months and 6 days.

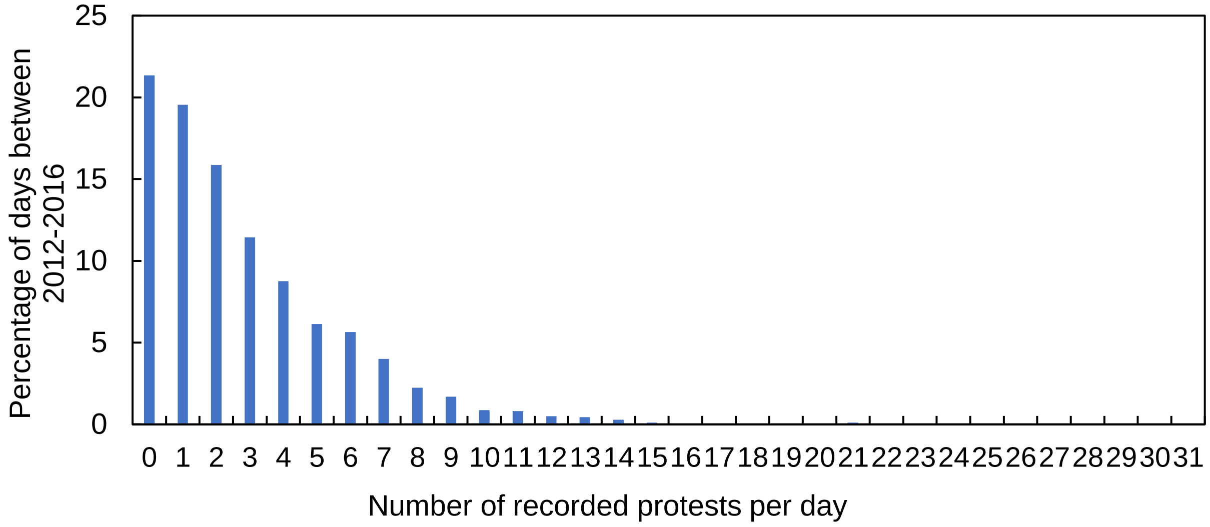
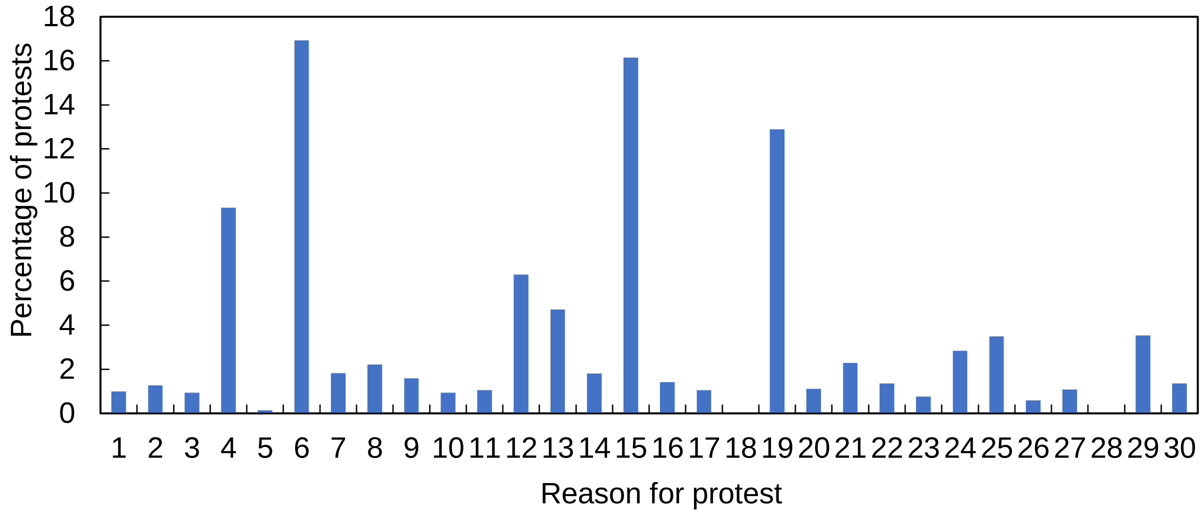



*Figure 18: The distribution of the number of protests recorded in South Africa for each day from 1 January 2012 to 31 December 2016.*

The highest number of recorded protests on one day was on the 7th of March 2012 when COSATU held a nationwide protest against e-tolls and labour brokering. This gave rise to 31 protests occurring on the same day. The second highest was 28 protests on the 5th of March 2016. These were all election protests which occurred on the voter registration day.

The percentage of protests which are attributed to each of the 30 different reasons to protest, is shown in Figure 19. Over the 20-year time frame (1997-2016) the three biggest reasons for protests were education (16.9%), narrowly followed by labour strikes (16.2%) and municipal services (12.9%). Labour related protests could be seen throughout the year. This can be explained by wage discussions that occur all year round, every year. In contrast to this clusters were observed in protests related to education and municipal services.

It was often seen that a protest in one region triggered multiple protests in different regions, all protesting for the same reason. One reason for these trends can be as a result of promises made before elections, creating similar expectations in different regions. Clusters of protests may occur if these expectations are not achieved in the expected timeframe. A second way to explain these trends is through the grievance-based approach. One group may start protesting as a result of a perceived injustice. As the grievance gains media attention other individuals may notice that they are experiencing the same injustice, causing clusters of protests.




| | | | |
|---|---|---|---|
| 1. Business practice | 2. By law enforcement | 3. Corruption | 4. Crime |
| 5. Democracy | 6. Education | 7. Elections | 8. Electricity |
| 9. Environmental | 10. Xenophobia | 11. Healthcare | 12. Housing |
| 13. International causes | 14. Jobs | 15. Labour strike | 16. Land |
| 17. Mismanagement | 18. Mob justice | 19. Municipal service | 20. National |
| 21. Political | 22. Roads | 23. Sanitation | 24. Transport |
| 25. Unspecified | 26. Various | 27. Water | 28. Vigilantism |
| 29. Other | 30. Racism | | |

*Figure 19: The percentage of South African protests which are attributed to each of the 30 different reasons for protests.*

Violent behaviour was recorded in 46.9% of the protests, whereas property damage was documented in 14.4% of protests. Overall this means that 30.7% of violent protests resulted in property being damaged or destroyed. Additional analysis of the social unrest database can be seen in Appendix C.

## 5.2 Correlation and trendline comparison

Spearman's rank correlation and trendline analysis are two methods that can be used to determine the relationships that are present between protests and the various economic and socio-economic variables. This allows for better understanding of how the variables interact with one another over the 20-year time frame (1997-2016). Spearman's rank correlations are discussed first, followed by trendline analysis for the different variables.

### 5.2.1 Spearman's rank correlation

Spearman's rank correlation determines whether a monotonic relationship exists between two variables. The presence of a monotonic relationship does not necessarily mean that a change in the one variable causes a change in the other variable. Both national and provincial data were examined to create a holistic image of the relationships within South Africa. Spearman's rank correlation for the national data are described first. This is followed by Spearman's rank correlation that was performed on the provincial data of all nine the South African provinces.

#### 5.2.1.1 Spearman's rank correlation: National data

The national data were evaluated for three different time periods, namely monthly, quarterly and annual. This was done for two reasons. The first is that the relationships that exist may change as a result of the data frequency. It is important to identify such changes to get a holistic understanding of South African protests. The second reason is the availability of data. When considering monthly data there are 12 times more data points available than when annual data are used. The problem however is that monthly data for economic and socio-economic data are seldom available. To get an idea of the interactions between the other variables, it is essential to consider quarterly and annual data.

The economic and socio-economic variables are divided into two groups. The first group is made up with the variables that were available for multiple time intervals (shown in Table 7). While the second group consists of the variables where only annual data were available (shown in Table 8).

The first group is discussed first. Spearman's rank correlation coefficients for the variables for multiple time intervals data were available, is shown in Table 7. The relationships that were observed are described below.

The Food Price Index (FPI) and the Consumer Price Index (CPI) variables are both related to the inflation levels within a country. In South Africa it was found that changing food prices were associated with the incidence of social unrest. At a 1% level, the FPI has a statistically significant positive relationship with the number of monthly, quarterly and annual protests. A statistically significant positive relationship, at the same level, was also found between CPI and the incidence of monthly, quarterly and annual protests. Indicating that inflation as a whole is associated with higher levels of social unrest within the country.

It was also found that South Africa's fiscal policies are associated with the incidence of protests within the country. At a 1% level, a statistically significant positive relationship was found between total government revenue and number of monthly, quarterly and annual protests. The same relationship was present between government expenditure and the number of monthly, quarterly and annual protests. The positive relationship between government revenue and protests is not strange as increased levels of revenue may be linked to increased levels of tax and higher expectations regarding service delivery and infrastructure. This may fuel discontent within the nation when or if these expectations are not met.

Poticelli and Voth (2011) found that democratic countries experiencing mass protests prior to elections were prone to increase government spending. This was done to improve the citizens perspective of the country, thus decreasing the number of protests and gaining support for the upcoming elections. The positive relationship between government expenditure and protests is indicative that this plan may not be as effective as thought.

A possible reason for the positive relationship becomes evident when analysing the relationship that exists between protests and the government's surplus. A statistically significant negative relationship exists between these two variables. This suggests that it may not be high levels of expenditure that fuels discontent but rather the imbalance between revenue and expenditure that may be responsible for the dissatisfaction in South Africa. Another possible explanation is that there may be a relationship between the size of the government's surplus and the country's economic prosperity. It may thus act as a proxy for the country's financial stability.

The price of petrol, diesel and paraffin are dependent on the exchange rate, Brent crude oil price as well as the government's tax laws. A statistically significant positive relationship was found between the number of monthly, quarterly and annual protests and the wholesale price of both petrol and paraffin. This indicates that the combined increase in these three factors are associated with increased levels of protests. Similar results were observed between protests and pre-tax diesel prices, indicating that the results remain even when the effects of government tax laws are removed. A possible explanation for the link between petrol, diesel and paraffin prices and protests is that increases in the fuel price may lead to increased delivery costs. This may in turn result in increased levels of inflation, which may fuel discontent as the cycle may cause the cost of living to grow faster than the salary growth rate resulting in higher levels of poverty. A statistically significant positive relationship was also found between protests and the Rand/Dollar exchange rate, which is expected as increases in this exchange rate can result in fuel price increases.

*Table 7: Spearman's rank correlation coefficient of the number of monthly, quarterly and annual protests, economic and socioeconomic variables for South Africa.*

| | Monthly | Quarterly | Annual |
|---|---|---|---|
| FPI | 0.69391* | 0.70955* | 0.72632* |
| CPI Headline Index (2016=100) | 0.85077* | 0.88832* | 0.92481* |
| Total Government Revenue (R Mil) | 0.7569* | 0.87478* | 0.91729* |
| Total Government Expenditure (R Mil) | 0.84545* | 0.89296* | 0.9218* |
| Total Government Surplus (R Mil) | -0.36224* | -0.69222* | -0.73985* |
| Basic Diesel Price | 0.78225* | 0.81934* | 0.89474* |
| Exchange Rate Rands/US$ | 0.67321* | 0.69331* | 0.76992* |
| Inland Petrol Prices | 0.83794* | 0.86695* | 0.90376* |
| Inland Paraffin Price | 0.79682* | 0.83104* | 0.90226* |
| Income Growth | - | -0.40465* | -0.52275** |
| Total Unemployment Rate | - | 0.27536** | 0.12185 |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

The quarterly income growth rate displayed a statistically significant negative relationship with protests, at a 1% level. When annual data were analysed, the statistical significance of the negative relationship decreased from a 1% level to a 5% level. Thus, higher salary growth was associated with lower levels of protest. Higher salary growth is more likely to either grow in line with inflation or above it. This may improve individual spending ability, in turn improving the standard of living. Low salary growth does the opposite. As additional strain is placed on an individual's spending ability the standard of living is lowered, which may lead to increased levels of poverty. Any one of these knock-on effects of low salary growth may lead to dissatisfaction and may give rise to grievance-based protests.

A positive relationship between quarterly overall unemployment and protests was found at a 5% level of statistical significance. A positive relationship was also present between the annual overall unemployment figures and number of annual protests, this relationship is however, not statistically significant.

Only annual data were available for the remainder of the variables discussed in this section. The Spearman's rank correlation coefficients between annual protests and the main economic and socioeconomic variables, can be seen in Table 8. Spearman's correlations for additional annual variables are depicted in Appendix D.

*Table 8: Spearman's rank correlation coefficient for the number of annual protests and annual economic and socioeconomic variables for South Africa.*

| | | N | All protests |
|---|---|---|---|
| Economic | GDP Growth Rate | 20 | -0.38496*** |
| | GDP Per Capita Growth Rate | 20 | -0.36692 |
| Demographic | Population Density | 20 | 0.9218* |
| | Annual Population Growth Rate | 20 | -0.28872 |
| | Annual Rural Population Growth Rate | 20 | -0.41955*** |
| | Annual Urban Population Growth Rate | 20 | -0.48271** |
| Country-specific | Corruption Perception Index | 20 | -0.66894* |
| | Government Subsidies (% Of Government Expense) | 20 | 0.72632* |
| Socio-economic | Education Completed (At Least Primary School) | 14 | 0.86813* |
| | Human Development Index | 19 | 0.50461** |
| | Access to Electricity (% Of Population) | 20 | 0.89282* |
| | Access to Improved Sanitation (% Of Population) | 19 | 0.90877* |
| | Female Unemployment Rate | 20 | -0.18346 |
| | Male Unemployment Rate | 20 | 0.46617** |
| | Youth Unemployment Rate | 20 | 0.28722 |
| International | Access to The Internet (% Of Population) | 20 | 0.91429* |
| | Access to Cell Phones (% Of Population) | 20 | 0.9203* |
| | Access to Fixed Telephone (% Of Population) | 20 | -0.90977* |
| | Freedom of The Press | 20 | 0.78141* |
| | Economic Freedom Index | 20 | -0.66918* |
| Other | Access to Improved Water Sources (% Of Population) | 19 | 0.90877* |
| | Trade Freedom | 20 | 0.58304* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

The GDP growth rate and the per capita GDP growth rates are two indicators that are commonly used to describe the economic landscape of a country. Negative associations with protests were observed with the annual growth rates for both these variables. However, a statistically significant relationship was only present between the GDP growth rate and annual number of protests. Increased levels of economic growth usually result in the population being exposed to more opportunities. This can result in lower levels of dissatisfaction, decreasing the incentive to protest. Economic growth also has knock-on effects, these include lowering unemployment, increasing salary growth and investments in infrastructure, thereby further lowering individuals' incentive to protest.

Population density showed the strongest relationship to the number of protests of all the demographic variables, with a positive relationship at a 1% level of statistical significance. Increased population density results in increased proximity between individuals. This increased proximity can increase ease of planning protests and may increase the possibility of finding individuals with similar grievances (mobilisation theory). Both factors may cause an increase in protest numbers.

A negative association was found between overall population growth and protests. However, the relationship was not statistically significant. When rural and urban population growth rates were considered separately, statistically significant negative relationships, at a level of 5% and 10% respectively, were observed.

The birth and death rate of the country is incorporated into the overall population growth rate. The urban and rural population growth rates also account for birth and death rates, but it also considers migratory patterns from urban to rural areas and visa-versa. New parents may have a lower propensity to protest because they have to provide and look after the young children. Similarly, migrants, especially foreign migrants, may also be less inclined to protest. This reluctancy to protest may play a part in the incidence of protests and may result in the negative relationship between population growth and protests.

Young children, especially babies, do not protest. A lagged effect between the overall population growth rates and protests may also be present. However, due to a lack of data, as this lag can be very long, the presence of a lagged relationship cannot be reliably tested. This lagged relationship may vary from what was observed in this study.

The Corruption Perception Index works on a sliding scale with 0 representing a country being perceived as being highly corrupt and 100 representing a country that is deemed "free" of corruption. At a 1% level of statistical significance, a negative relationship was found between this index and the number of annual protests. Increased levels of corruption can induce dissatisfaction in individuals who are disadvantaged as a result of it.

Government subsidies, as a percentage of government expenses, showed a statistically significant positive relationship with protests, at a 1% level of significance. This is contrary to what is expected, as increased levels of subsidies should lead to lower levels of poverty, decreasing dissatisfaction. However, this is only the case when subsidies are large enough. Furthermore, as this percentage of government expenses increases for subsidies, money is taken away from other crucial services such as education, healthcare and infrastructure. This can cause a whole new reason for grievance-based protests.

Education and the Human Development Index showed statistically significant positive relationships with the number of protests at a 1% and 5% level of statistical significance respectively. This shows that increased levels of education in South Africa is associated with increased levels of protests. Education brings future aspirations. Individuals may become disgruntled when these expectations are not met, and the government is deemed to be the responsible party. Further associations between socio-economic variables and protests were also observed.

It was seen that sanitation, electricity and access to water were all positively correlated with the number of protests, at a 1% level of statistical significance. The positive relationships are counter intuitive. By using the grievance-based approach, an increase in service delivery should lead to fewer people being disgruntled and therefore negative relationships are expected. However, at the end of Apartheid numerous promises were made about sanitation, access to water and electricity. There has been a delay in the delivery of these promises (as the infrastructure had to be built) and this delay may result in the individuals who are still waiting to become disgruntled. Grievances may also be caused by a mismatch between individuals' expectations and the reality of the services being provided. Examples may include improved sanitation constituting the installation of the bucket system and having to pay for water and electricity when individuals are unable to so.

Subcategories of unemployment rendered mixed and inconclusive results. A non-statistically significant positive association was present between the annual overall unemployment figures and number of protests. The same results were observed for the youth unemployment figures. When the overall unemployment figures where spilt by gender an anomaly arose. At a 10% level of statistical significance, a positive relationship was found between male unemployment and protest incidence. A negative association was observed between female unemployment and the number of protests, this relationship is however, not statistically significant

Associations between number of protests and modes of communication were also observed, with all the relationships being statistically significant at a 1% level. Both internet and cell phone access were positively correlated with the number of protests, consistent with the mobilisation theory. It was also found that freedom of the press (which measures 0 for a country with complete freedom of speech and publications and 100 having no freedom of press whatsoever) was positively correlated with protests. However, a negative relationship was observed between protests and fixed telephone subscriptions. This can possibly be attributed to the decreased demand for fixed telephone connections as mobile phones become more popular.

The economic freedom index, released by the Heritage organisation, works on a sliding scale between 0 and 100 (where 100 indicates complete economic freedom, and 0 the opposite). This indicator considers numerous aspects of a country, including its fiscal health, judicial effectiveness, government integrity and labour freedom. Economic freedom and protests showed a statistically significant negative correlation at a 1% level. Indicating that increased levels of economic freedom are associated with fewer protests in the country. Improvements in the economic factors taken into consideration may result in citizens having more confidence in the government.

Trade freedom, also released by the Heritage organisation, indicates a positive relationship with the number of protests, at a 1% level of statistical significance. This is the similar to what was seen in literature.

**5.2.1.2 Spearman's rank correlation: Provincial data**

The economic and socio-economic variables for all nine provinces in South Africa are not as readily available as the national data, as shown in Table 5. Due to this limitation, only the annual data were used in the analysis. The socioeconomic and corresponding protest data for each of the nine provinces were analysed to determine provincial relationships. The Spearman's rank correlation coefficients between the provincial protests and socio-economic variables for which data could be obtained can be seen in Table 9.

*Table 9: Spearman's rank correlation coefficient of the number of provincial protests and the provincial economic and socio-economic variables for South Africa.*

| | Provincial protests |
|---|---|
| Population Size | 0.67801* |
| Proportion of South Africa's Population in Province (%) | 0.58945* |
| Households with Access to Landlines (%) | 0.14663*** |
| Households with Access to Mobile Phones (%) | 0.7215* |
| Households with Access to Piped or Tap Water in Their Dwellings (%) | 0.31632* |
| Households with Access to Improved Sanitation (%) | 0.6014* |
| Households That Have No Toilets or Use Bucket System (%) | -0.64335* |
| Households Connected to Main Electricity Supply (%) | 0.11564 |
| Persons with Formal Schooling (%) | 0.83584* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

There are two different demographic measures that are available for provincial data, the population size per province and what proportion of South Africa's population live in each of the nine provinces. At a 1% level of statistical significance, both variables are positively correlated to the number of provincial protests. These relationships are consistent with the mobilisation theory which states the larger the population size becomes, the easier it becomes to organise a protest which in turn results in more protests occurring.

Modes of communication works in the same way as population size. At a 1% level of statistical significance, a positive correlation has also been found between households with access to cell phones and provincial protests. This is equivalent to the relationship seen in the national data. A similar positive relationship was seen between households with access to landlines and provincial protests. This relationship, however, is only significant at a 10% level of statistical significance. This contrasts with what is observed at a national level where, at a 1% level of statistical significance, a negative relationship was observed.

Socio-economic measures like water, sanitation, electricity and education are usually explained with a grievance-based approach, rather than the mobilisation theory which is used for demographic factors and modes of communication. It was found that provincial protests were positively correlated to the proportion of households in the province that have access to tap water, those that have access to improved sanitation as well as persons with formal schooling, at a 1% level of statistical significance. At the same level, a negative correlation was found between provincial protests and the proportion of households in the province that did not have toilets and had to make use of the bucket system. These results correspond to the relationships that were observed in the national data. However, the relationship between protests and electricity supply changed. A positive association was still present between the two, but it was not statistically significant.

### 5.2.2 Comparison of trendline analysis

In the Spearman's rank correlations, very few differences are observed due to the time period considered or whether national or provincial data were analysed. Trendlines have been drawn to identify the impact that these changes have on the patterns in protests. The impact of the data frequency is discussed first. Hereafter the national and provincial trends are compared with one another.

The monthly, quarterly and annual trendlines between protests and seven economic and socio-economic variables (income growth, CPI, government expenditure, inland petrol prices, FPI, government revenue, and government surplus) are depicted in Figure 20. Of these seven variables, the literature proposes linear relationships for all of the variables except the relationship between income growth and the number of protests. This relationship was described as parabolic (Parvin, 1973). Similar results were seen in South Africa's data, as parabolic trendlines best explained the relationship between income growth and the number of quarterly and annual protests (Figure 20a).

All the other relationships depicted in the literature, were described by monotonic relationships. In South Africa, monotonic relationships were consistently observed in the monthly, quarterly and annual data for three of the economic and socio-economic variables: CPI, government expenditure and inland petrol prices (Figure 20b, c, d respectively). The three remaining variables, FPI, government revenue and government surplus, were not consistently described by either a monotonic or parabolic relationship (Figure 20e, f, g respectively). For these three variables the relationship between the data were dependent on the timeframe being considered.

When considering the quarterly and annual data of the FPI, an increasing monotonic relationship is observed. Thus, as the FPI increases, the number of protests increase at a specific rate. This is not the case when considering monthly data. Here a negative parabolic trendline describes the general pattern of the data. This shows that increases in food price inflation are linked to increases in the number of protests, but once a threshold is reached the number of protests stabilises and then start to decrease slightly. In this case the grievance-based argument can still possibly explain the reason for protesting but the various income brackets in the country may explain the concave shape. Low income homes, which make up a large portion of South African households, have very few resources to shield themselves from the effects of small increases in FPI. As a result, even small increases in FPI can cause a lot of dissatisfaction. As the FPI increases more people become unable to shield themselves from the effects, resulting in more and more people becoming unhappy. This continues until the effect is filtered through all the income brackets. At this point, the grievance is felt throughout which then causes the number of protests to plateau.

The changes that were seen in the relationship between protests and government revenue was similar to those seen in the FPI. Increasing monotonic relationships were seen for the quarterly and annual data, while a negative parabolic relationship was observed in the monthly data. Government surplus behaved differently to the FPI and government revenue. Positive parabolic relationships best explained the monthly and annual data. Whereas the quarterly data's trendline was monotonically decreasing.

The change in data frequencies allows different relationships to be identified between the number of protests and economic and socio-economic variables. This is due to the differences in the seasonal trends and volatility within the data. Changes in the relationships may also be observed if national and provincial data are analysed. The national and provincial trendlines (access to cell phones, education, water and sanitation, population size and access to landlines) are depicted in Figure 21.

Cell phone access and the number of protests did not show significant differences in trends between the national and provincial data (Figure 21a). Positive parabolic trendlines with similar characteristics have been found for both. Similar trendlines were observed for access to education (Figure 21b). Positive parabolic trends could also be seen in both national and provincial data for access to water and access to improved sanitation (Figure 21c, d respectively). The difference here had to do with the curvature of the parabolas. For both these variables the national trendline has a steeper increase than the provincial trendline.

Vast differences in the national and provincial relationships are observed for access to a landline and the population size (Figure 21e, f respectively). A decreasing monotonic trend explains the national relationship between protests and access to landlines. While a concave parabolic relationship is seen in the provincial data. The national relationship between protests and population size is explained by a positive parabolic curve where an increasing monotonic function explains the provincial relationship.

When Spearman's rank correlations are considered, very few differences are observed in the relationships between protests and economic and socio-economic variables as changes in the data frequency are made. Similarly, the use of either national or provincial data yielded few changes. When trendlines are considered it is clear that this is not the whole story. It has been seen that the relationships between protests and three economic variables namely, FPI, government revenue and government surplus changed as the data frequency was altered. The use of either national or provincial data also showed changes in the relationships. The trends between protests and two variables, namely telephone access and population size varied greatly depending on the type of data considered. This suggests that the relationships between protests and economic and socio-economic variables are sensitive to these changes in the data frequency and changing from national to provincial data.

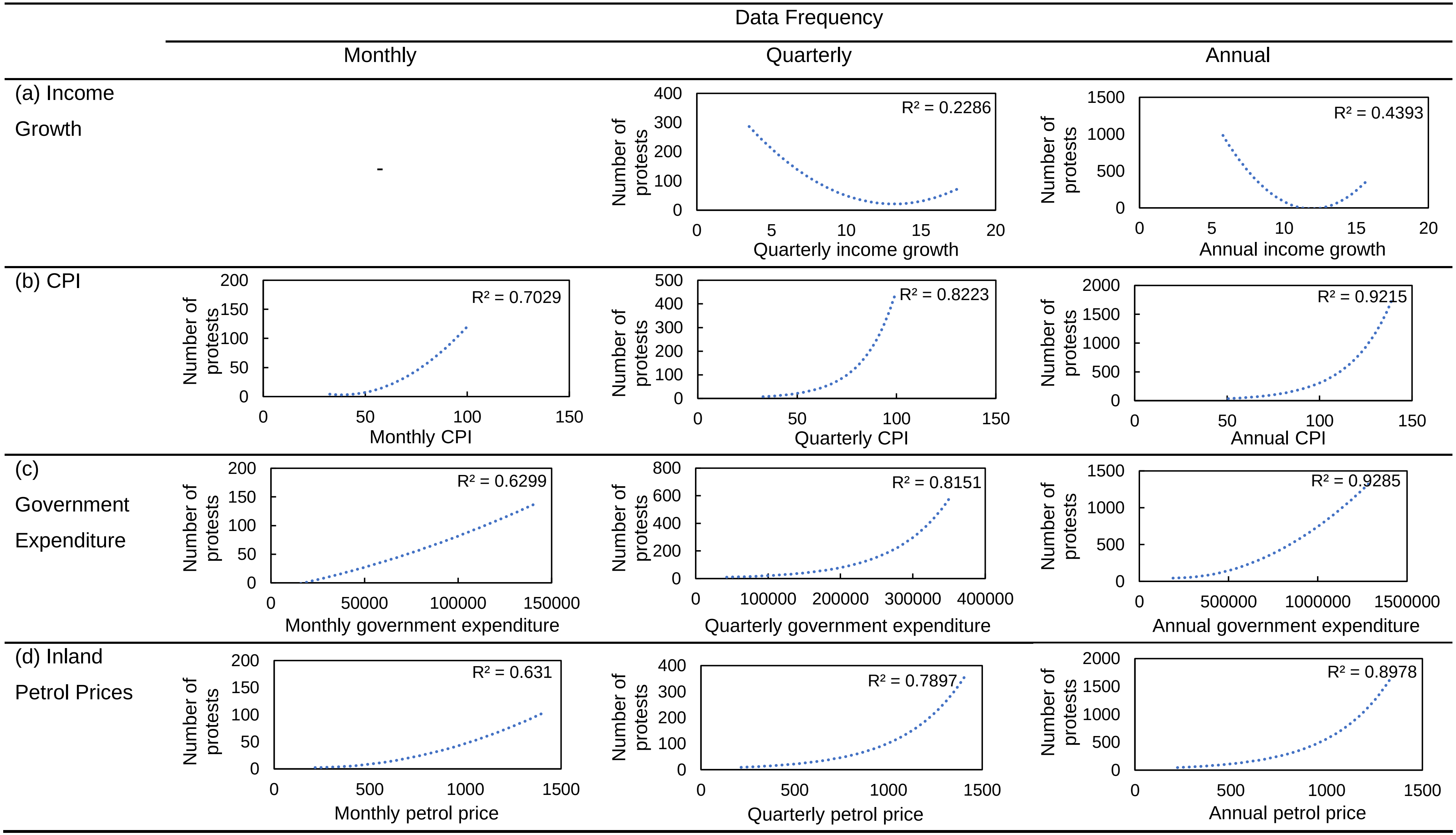



*Figure 20: Trendline comparison between protests and economic and socio-economic variables for different time intervals.*

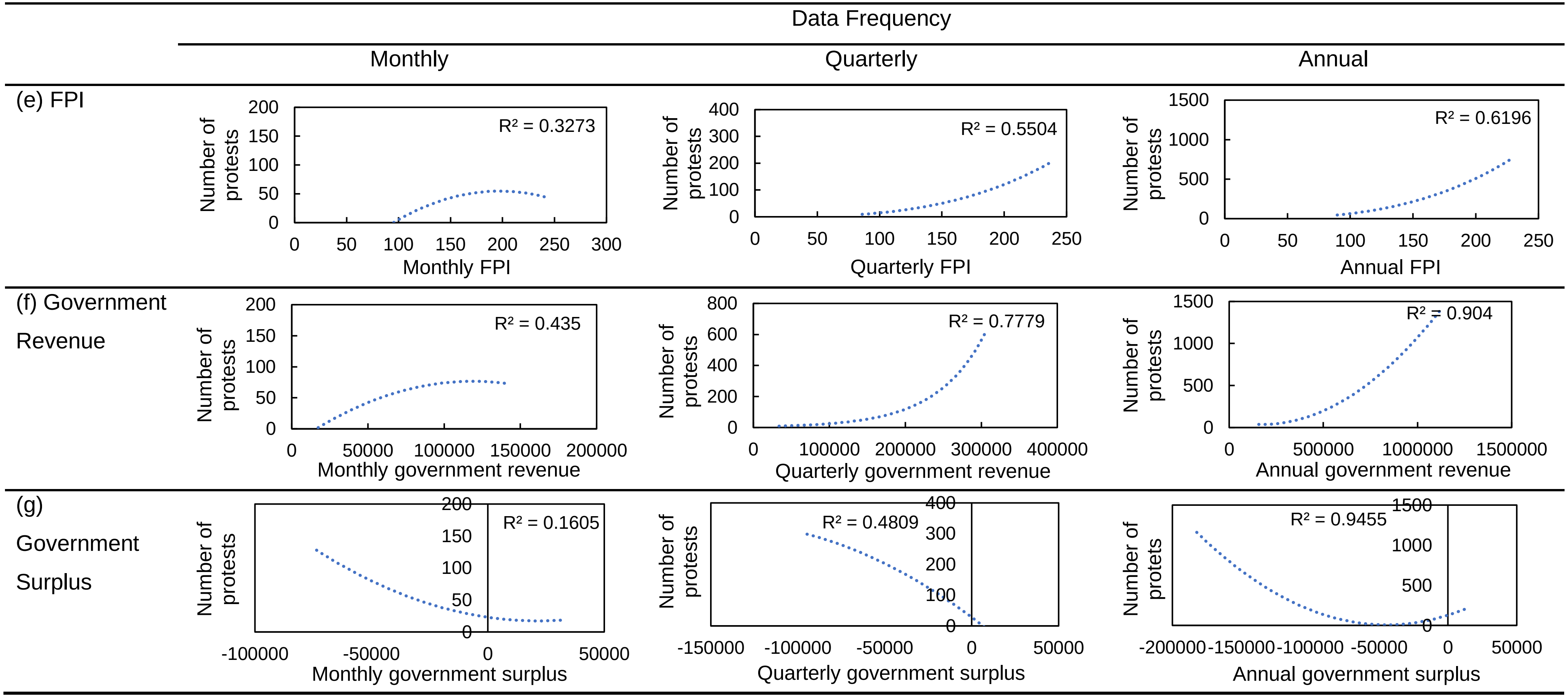



Figure 20 cont.: Trendline comparison between South African protests and economic and socio-economic variables for different time intervals.

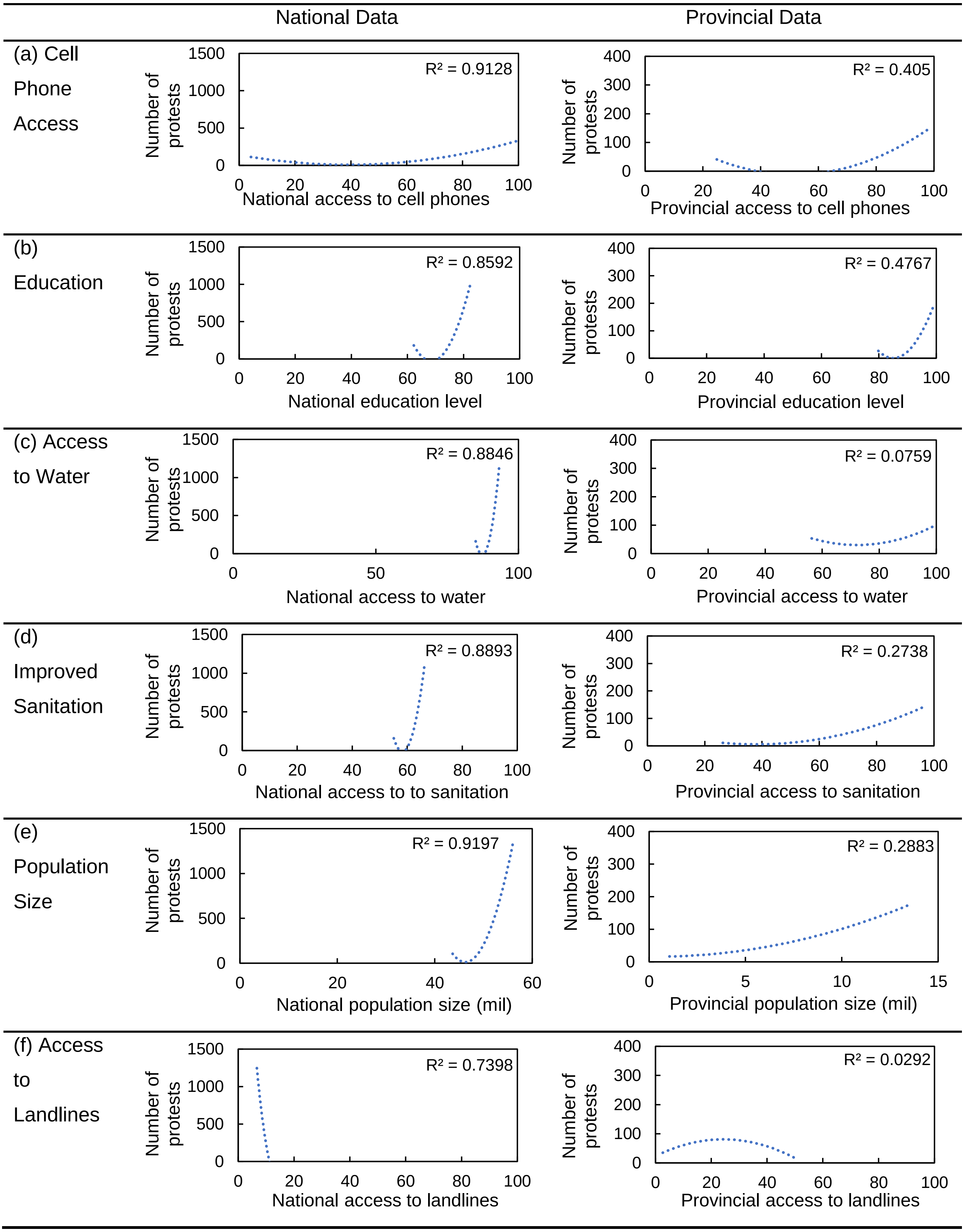



Figure 21: National and provincial trendline comparisons for South Africa.

## 5.3 Modelling the number of protests

To better understand South African protests, models that attempted to recreate the behaviour were created by considering the number of daily, monthly, quarterly and annual protests. The number of daily and annual protests were only modelled using the univariate time series modelling technique. This was due to the scarcity of daily economic and socio-economic data and the limited number of annual data points. The monthly and quarterly data bases balanced the availability of economic data while containing sufficient data points. This allowed for additional modelling techniques to be used. As such the monthly and quarterly incidence of social unrest were modelled using univariate time series, linear regression, lagged regression and VAR (1) models.

This section starts by looking at a univariate time series model for the number of daily protests. This is followed by the four monthly models. The monthly univariate time series is described first, followed by the linear regression, lagged regression and VAR (1) models and is concluded with a comparison of the four monthly models. Hereafter, the four quarterly models are described in the same manner as the monthly models. Thereafter, the univariate time series model for the number of annual protests is given. The section is concluded with a summary and comparison of all ten models.

### 5.3.1 Modelling the number of daily protests

Daily economic and socio-economic data are not readily available. Therefore, the model for the number of protests per day only considers social unrest data. The number of daily protests increased exponentially from 1997 to 2016. This was not the only trend that could be observed, there were also seasonal trends present. Wednesdays had the highest incidence of protests, while the lowest incidence occurred on Sundays.

A seasonally exponential smoothing model accounts for long term and seasonal trends, making it appropriate to model the daily number of protests. The model parameters for this model, depicted in Table 10, are all statistically significant. This model explains 49.95% of the variation present in the number of daily protests ($R^2$ = 0.49951).

The prediction error series that is created by this model is stationary, however, autocorrelation and partial autocorrelation are present within the residuals. This means that the model does not remove all the trends that are present within the data, making it an inappropriate modelling technique. For this reason, the model's equation has not been provided.

*Table 10: The parameter estimates for the seasonal exponential smoothing model for the number of daily protests.*

| Model Parameter | Estimate | Std. Error | T Value | Prob>\|T\| |
|---|---|---|---|---|
| Level Smoothing Weight | 0.14642 | 0.0043 | 33.9123 | <.0001 |
| Seasonal Smoothing Weight | 0.03469 | 0.0024 | 14.6441 | <.0001 |
| Smoothed Level | 0.27182 | - | - | - |
| Smoothed Seasonal Factor 1: Sunday | -2.32668 | - | - | - |
| Smoothed Seasonal Factor 2: Monday | 1.13756 | - | - | - |
| Smoothed Seasonal Factor 3: Tuesday | 1.00825 | - | - | - |
| Smoothed Seasonal Factor 4: Wednesday | 1.73120 | - | - | - |
| Smoothed Seasonal Factor 5: Thursday | 0.74534 | - | - | - |
| Smoothed Seasonal Factor 6: Friday | -0.33324 | - | - | - |
| Smoothed Seasonal Factor 7: Saturday | -1.91367 | - | - | - |

### 5.3.2 Modelling the number of monthly protests

The four methods used to model the number of monthly protests is described in this section. The univariate case, where no economic variables are considered, is described first. This is followed by three models that incorporate economic data, namely the linear regression model, a lagged regression model and a VAR (1) model. Thereafter, the four models are compared with one another.

#### 5.3.2.1 Univariate Time Series model

Two trends were observed in the monthly data. The first is a linear trend, with the number of protests increasing each month. The second is a constant seasonal trend. Historically, March had the most recorded protests while December had the fewest. The estimate for December is much lower than the estimates for the other months, indicating that the protest incidence is considerably reduced during this period.

The Additive Winters Method model, which incorporates both trends, is an ideal model for this data. The model's parameter estimates are depicted in Table 11. It has one parameter, the trend smoothing weight, which is not statistically significant. All of the other parameters were statistically significant. This model explains 81% of the variation that is present within the data ($R^2 = 0.810$).

The model diagnostics reveals that the residuals are stationary, but they are not uncorrelated. This means that the model does not remove all the trends that are present within the data, making it an inappropriate modelling technique. For this reason, the model's equation has not been provided.

*Table 11: The parameter estimates for the Additive Winters Method model for the number of monthly protests.*

| Model Parameter | Estimate | Std. Error | T Value | Prob>\|T\| |
|---|---|---|---|---|
| Level Smoothing Weight | 0.39035 | 0.0345 | 11.317 | <.0001 |
| Trend Smothing Weight | 0.00100 | 0.0050 | 0.2008 | 0.8410 |
| Seasonal Smoothing Weight | 0.45981 | 0.0623 | 7.3796 | <.0001 |
| Smoothed Level | 86.55648 | - | - | - |
| Smoothed Trend | 0.35994 | - | - | - |
| Smoothed Seasonal Factor 1: January | -2.51312 | - | - | - |
| Smoothed Seasonal Factor 2: Febuary | 23.89473 | - | - | - |
| Smoothed Seasonal Factor 3: March | 33.28908 | - | - | - |
| Smoothed Seasonal Factor 4: April | 2.52928 | - | - | - |
| Smoothed Seasonal Factor 5: May | 9.81561 | - | - | - |
| Smoothed Seasonal Factor 6: June | -15.32249 | - | - | - |
| Smoothed Seasonal Factor 7: July | -10.57674 | - | - | - |
| Smoothed Seasonal Factor 8: Augest | 8.40724 | - | - | - |
| Smoothed Seasonal Factor 9: September | 13.02323 | - | - | - |
| Smoothed Seasonal Factor 10: October | 15.00278 | - | - | - |
| Smoothed Seasonal Factor 11: November | -17.82846 | - | - | - |
| Smoothed Seasonal Factor 12: December | -58.68866 | - | - | - |

#### 5.3.2.2 Linear regression model

South Africa's fiscal policies play an integral part in understanding and explaining changes in the monthly protest patterns. In the regression model the government revenue and surplus were both significant predictors for the number of monthly protests. None of the other economic and socio-economic variables were identified as significant predictors in this model.

The model using government revenue and surplus explained 62.79% of the variation that was present in the number of monthly protests ($R^2 = 0.6279$). This model's analysis of variance and the model's parameters are shown in Table 12 and 13 respectively. This is followed by the model's equation which is depicted by Equation 10.

*Table 12: Analysis of variance for the monthly linear regression model.*

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 2 | 231016 | 115508 | 199.10 | <.0001 |
| Error | 236 | 136913 | 580.13841 | | |
| Corrected Total | 238 | 367929 | | | |

*Table 13: Model parameters for the monthly linear regression model of number of monthly protests.*

| Variable | Parameter Estimate | Standard Error | Type II SS | F Value | Pr > F |
|---|---|---|---|---|---|
| Intercept | -19.38466 | 2.99875 | 24242 | 41.79 | <.0001 |
| Government Revenue | 0.00096962 | 0.00005465 | 182599 | 314.75 | <.0001 |
| Government Surplus | -0.00114 | 0.00009434 | 84296 | 145.30 | <.0001 |

$$P_M = -19.38466 + 0.00096962R - 0.00114S \tag{10}$$

where $P_M$, $R$ and $S$ represents the number of monthly protests, government revenue and government surplus respectively.

This model only makes use of fiscal variables. This is useful for two reasons. The projected values for government revenue and surplus can be found in the budget and could be used to calculate the number of monthly protests. The second reason is that as it only considers two economic variables it is very easy to see how changes in revenue and government surplus may influence the level of social unrest within the country. This could also be used by the government when setting up a budget, as it may enable them to set up their revenue and expenditure such that it keeps social unrest at a reasonable level. An example and explanation of SAS procedure used to create a regression model is described in Appendix E.

#### 5.3.2.3 Lagged regression model

The lagged regression model for the number of monthly protests can be seen as an extension of the linear regression model just described. The lagged regression model made use of six variables with varying lagged time intervals. Five of those were factors that may be influenced by the government's fiscal policy namely; CPI, FPI, government revenue, expenditure and surplus. The sixth variable is the internet indicator variable. This was equal to 0 when less than 10% of the population makes use of the internet and denoted by 1 otherwise. All other variables were not significant predictors in this model.

The analysis of variance and the model's parameters are depicted in Table 14 and 15 respectively. This is followed by the model's equation which is depicted by Equation 11. The interactions that are seen in Table 15 are more complicated than the arguments given in Section 5.2. This is a result of the cross-correlations that are prevalent. This lagged regression model explains 81.38% of the variation in the number of monthly protests ($R^2 = 0.8138$).

This model considers an array of variables that can be influenced by changes in the government's fiscal policy. Once more showing that it is an integral part in understanding monthly protest patterns. In addition to the insurers, these interactions may be used by the government when setting up their fiscal policies to a balance the budget and protest incidence. An example and explanation of SAS procedure that was performed to create a lagged regression model is described in Appendix E.

*Table 14: Analysis of variance for the monthly lagged regression model.*

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 13 | 292393 | 22492 | 71.60 | <.0001 |
| Error | 213 | 66906 | 314.11439 | | |
| Corrected Total | 226 | 359299 | | | |

*Table 15: Model parameters for the monthly linear regression model.*

| Variable | Parameter Estimate | Standard Error | Type II SS | F Value | Pr > F |
|---|---|---|---|---|---|
| Intercept | -26.80081 | 2.94819 | 25958 | 82.64 | <.0001 |
| CPI(t-2)* | 10.64259 | 4.04476 | 2174.69161 | 6.92 | 0.0091 |
| FPI(t-3)* | -0.78167 | 0.30574 | 2053.20426 | 6.54 | 0.0113 |
| FPI(t-5)* | -0.76009 | 0.30198 | 1990.08387 | 6.34 | 0.0126 |
| FPI(t-9)* | -0.70761 | 0.28899 | 1883.30166 | 6.00 | 0.0152 |
| FPI(t-12)* | -1.11516 | 0.29227 | 4572.90054 | 14.56 | 0.0002 |
| Government Expenditure(t-7)* | 0.00059278 | 0.00015692 | 4482.21010 | 14.27 | 0.0002 |
| Government Revenue(t-8) | 0.00157 | 0.00019224 | 20990 | 66.82 | <.0001 |
| Government Revenue(t-11) | -0.00041032 | 0.00019153 | 1441.63153 | 4.59 | 0.0333 |
| Government Surplus(t-6) | -0.00069460 | 0.00009649 | 16277 | 51.82 | <.0001 |
| Government Surplus(t-8) | -0.00112 | 0.00018085 | 12075 | 38.44 | <.0001 |
| Government Surplus(t-10) | -0.00068894 | 0.00009876 | 15285 | 48.66 | <.0001 |
| Government Surplus (t-11) | -0.00051457 | 0.00017657 | 2667.80060 | 8.49 | 0.0039 |
| Internet Access(t-2) | -27.46936 | 5.47106 | 7918.47027 | 25.21 | <.0001 |

*change in value from one month to the next

$$\begin{aligned} P_M(t) = &-26.8 + 10.64DC(t-2) - 0.78DF(t-3) - 0.76DF(t-5) - 0.71DF(t-9) \\ &- 1.12DF(t-12) + 0.0006DE(t-7) + 0.0016R(t-8) - 0.0004R(t-11) \\ &- 0.0007S(t-6) - 0.0011S(t-8) - 0.0007S(t-10) - 0.0005S(t-11) \\ &- 27.47I(t-2) \qquad (11) \end{aligned}$$

where $P_M$, $DC$, $DF$, $DE$, $R$, $S$ and $I$ represents the number of monthly protests, the change in CPI, change in FPI, change in government expenditure, government revenue, government surplus and the internet indicator respectively. $I = 0$ when internet usage was less than 10% of the population, otherwise $I = 1$.

### 5.3.2.4 VAR (1) model

A VAR (1) model, which incorporated FPI, petrol price and internet usage, was created for the number of monthly protests (SAS output with explanation can be found in Appendix F). The model parameters for this VAR (1) model is depicted in Table 16. This is followed by the model's equation, shown in Equation 12. While the FPI, regional inland petrol price and an internet indicator variable, were all statistically significant parameters, the intercept of this model was not. This model explained 76.55 percent of the variation present in the number of monthly protests, ($R^2 = 0.7655$).

*Table 16: The model parameter estimates of the monthly VAR (1) model.*

| Equation | Parameter | Estimate | Standard Error | T Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| $Protests_M(t)$ | CONST1 | 2.95222 | 5.01414 | 0.59 | 0.5566 | 1 |
| | XL0_1_1 | 10.51751 | 4.91779 | 2.14 | 0.0335 | Internet(t) |
| | AR1_1_1 | 0.55293 | 0.05505 | 10.04 | 0.0001 | $Protests_M(t-1)$ |
| | AR1_1_2 | -0.14726 | 0.05217 | -2.82 | 0.0052 | FPI(t-1) |
| | AR1_1_3 | 0.03992 | 0.00890 | 4.49 | 0.0001 | Inland Petrol(t-1) |

$$P_M(t) = \begin{cases} 2.9522 + 0.5529P_M(t-1) - 0.147FPI(t-1) + 0.0399IP\ (t-1)\ \ if\ \ INT(t) = 0 \\ 13.4697 + 0.5529P_M(t-1) - 0.147FPI(t-1) + 0.0399IP\ (t-1)\ if\ \ INT(t) = 1 \end{cases} \quad (12)$$

where $P_M$, $FPI$, $IP$ and $INT$ represent the number of monthly protests, Food price index, Inland petrol prices and internet usage respectively. $INT = 0$ when internet usage was less than 10% of the population, otherwise $INT = 1$.

By analysing the model diagnostics, it is seen that this model is quite a good fit for the data. The Granger-causality Wald test identified a causal interaction from FPI, regional inland petrol price and the internet indicator variable to the number of monthly protests. The Durbin-Watson test statistic value for this model was equal to 2.04911. This is sufficiently close to 2 indicating that the autocorrelation at lag 1 is close to zero. There are also very few residuals that have significant autocorrelations, depicted by a + or – in Table 17, implying that the model fits well. Furthermore, the VAR (1) model is stationary as the modulus of all three indices, depicted in Table 18, are smaller than one. Therefore, further tests for spurious regression and co-integration are not required for this model.

*Table 17: The Schematic Representation of Cross Correlations of Residuals for the monthly VAR (1) model.*

| Variable/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Protests | +.- | ... | ... | ... | ..- | ... | +.- | +.- | ... | -.. | ... | ... | +.. |

+ is > 2*std error,  - is < -2*std error,  . is between these two values

*Table 18: The roots of the AR characteristic polynomial for the monthly VAR (1) model.*

| Index | Real | Imaginary | Modulus | Radian | Degree |
|---|---|---|---|---|---|
| 1 | 0.98093 | 0.02314 | 0.9812 | 0.0236 | 1.3515 |
| 2 | 0.98093 | -0.02314 | 0.9812 | -0.0236 | -1.3515 |
| 3 | 0.50684 | 0.00000 | 0.5068 | 0.0000 | 0.0000 |

**5.3.2.5 Comparison of the monthly models**

The univariate time series model was inadequate to model the number of monthly protests, as it struggled to control the autocorrelation within the residuals. The three remaining models fared much better, producing high $R^2$ values. The lagged regression model had the highest explanatory value, with a $R^2$ value of 0.8138. The VAR (1) model's $R^2$ value was marginally lower ($R^2 = 0.7655$), while the linear regression model produced a $R^2$ value of 62.79.

Government surplus, revenue and expenditure, CPI, internet usage, inland petrol price and FPI were identified as significant predictors, in at least one of the four monthly models. Many of these variables are influenced by the government's fiscal policy. This supports the premise that governments can influence the incidence of unrest events by altering its fiscal policies.

The exchange rate, diesel prices and paraffin prices were not identified as significant predictors in any of the monthly models. Spearman's rank correlation, however, found statistically significant relationships between the incidence of protests and these four variables. This shows that significant relationships between social unrest events and economic and socio-economic variables may not always translate to the variables being significant predictors of social unrest.

### 5.3.3 Modelling the number of quarterly protests

The description of the quarterly models for the number of quarterly protests follows the same structure as the models for the number of monthly protests. Four different models are described in this section. The univariate case is described first. This is followed the linear regression model, a lagged regression model and a VAR (1) model. The section is concluded with a comparison of the four models.

#### 5.3.3.1 Univariate Time Series model

Two trends were observed in the quarterly protest data. The first is a linear trend, with the number of protests increasing each quarter. The second is a constant seasonal trend. Historically the highest number of protests are recorded in the first quarter of the year, 1 January to 31 March. With the lowest number of protests being recorded in the final quarter of the year.

As a result of these two trends that are present in the data, the Additive Winters Method model was selected. This model explains 84.9% of the variation that is present within the data ($R^2 = 0.849$). The estimated model parameters for this data are depicted in Table 19. Here it can be seen that the level smoothing weight is a statistically significant variable, while both the trend and seasonal smoothing weights are insignificant.

The model diagnostics reveals that the residuals are stationary, but they are not uncorrelated. This means that the model does not remove all the trends that are present within the data, making it an inappropriate modelling technique. For this reason, the model's equation has not been provided.

*Table 19: The parameter estimates for the Additive Winter's Method model for the number of quarterly protests from January 1997 to December 2016.*

| Model Parameter | Estimate | Std. Error | T Value | Prob>\|T\| |
|---|---|---|---|---|
| Level Smoothing Weight | 0.82198 | 0.0821 | 10.0160 | <.0001 |
| Trend Smoothing Weight | 0.00100 | 0.0188 | 0.0532 | 0.9577 |
| Seasonal Smoothing Weight | 0.00100 | 0.0805 | 0.0124 | 0.9901 |
| Smoothed Level | 239.80583 | - | - | - |
| Smoothed Trend | 3.68344 | - | - | - |
| Smoothed Seasonal Factor 1: Quarter 1 | 12.37440 | - | - | - |
| Smoothed Seasonal Factor 2: Quarter 2 | 3.10703 | - | - | - |
| Smoothed Seasonal Factor 3: Quarter 3 | 5.03488 | - | - | - |
| Smoothed Seasonal Factor 4: Quarter 4 | -20.52655 | - | - | - |

### 5.3.3.2 Linear regression model

The variables that are significant predictors in quarterly linear regression model for quarterly protests is vastly different to those that were included in the monthly model. This may be attributed to seasonal trends and volatility within the different data sets. In this case internet usage, income growth, change in FPI and change in CPI were statistically significant predictors for the number of quarterly protests.

The multiple regression model, with these four variables, explain 68.86% of the variation that was present in the number of quarterly protests ($R^2 = 0.6886$). The model's analysis of variance and the model's parameters are depicted in Table 20 and 21 respectively. This is followed by the model's equation, shown in Equation 13.

*Table 20: Analysis of variance for the quarterly linear regression model.*

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 661385 | 165346 | 40.90 | <.0001 |
| Error | 74 | 299133 | 4042.33603 | | |
| Corrected Total | 78 | 960518 | | | |

*Table 21: The model parameters for the quarterly linear regression model.*

| Variable | Parameter Estimate | Standard Error | T Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | 82.73123 | 27.26287 | 3.03 | 0.0033 |
| Internet Usage | 139.91399 | 15.86750 | 8.82 | <.0001 |
| Income Growth | -9.25784 | 2.51981 | -3.67 | 0.0004 |
| Change in FPI | -1.68896 | 0.72334 | -2.33 | 0.0223 |
| Change in CPI | 44.38751 | 12.95982 | 3.43 | 0.0010 |

$$P_Q(t) = \begin{cases} 82.731 - 9.258R(t) - 1.689(F_t - F_{t-1}) + 44.388(C_t - C_{t-1})\ if I(t) = 0 \\ 222.643 - 9.258R(t) - 1.689(F_t - F_{t-1}) + 44.388(C_t - C_{t-1})\ if I(t) = 1 \end{cases} \quad (13)$$

where $P_Q$, $R$, $F$, $C$ and $I$ represents the number of quarterly protests, Income growth, change in FPI, change in CPI and internet usage respectively. $I = 0$ if internet usage <10% of the population, $I = 1$ otherwise.

### 5.3.3.3 Lagged regression model

The lagged regression model for the number of quarterly protests makes use of different variables than the linear regression model just described. The variables that were significant predictors included the government surplus, government expenditure, CPI and FPI. All other variables were not significant predictors in this model. The analysis of variance and the model's parameters are depicted in Table 22 and 23 respectively. This is followed by the model's equation, depicted by Equation 14. This model explains 76.92% of the variation in the number of quarterly protests ($R^2 = 0.7692$).

*Table 22: Analysis of variance for the quarterly lagged regression model.*

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 6 | 719181 | 119864 | 37.76 | <.0001 |
| Error | 68 | 215836 | 3174.06133 | | |
| Corrected Total | 74 | 935017 | | | |

*Table 23: The model parameters for the quarterly linear regression model.*

| Variable | Parameter Estimate | Standard Error | Type II SS | F Value | Pr > F |
|---|---|---|---|---|---|
| Intercept | -44.50887 | 14.54955 | 29704 | 9.36 | 0.0032 |
| Government Expenditure(t-3)* | 0.00135 | 0.00037665 | 40629 | 12.80 | 0.0006 |
| Government Surplus(t-2) | -0.00235 | 0.00035242 | 141661 | 44.63 | <.0001 |
| Government Surplus(t-4) | -0.00188 | 0.00035054 | 91776 | 28.91 | <.0001 |
| CPI(t-2)* | 36.13688 | 11.46776 | 31518 | 9.93 | 0.0024 |
| CPI(t-4)* | 43.03492 | 12.19360 | 39536 | 12.46 | 0.0008 |
| FPI(t-4)* | -1.59038 | 0.66913 | 17931 | 5.65 | 0.0203 |

*change from one quarter to the next

$$P_Q(t) = -44.5 + 0.001DE(t-3) - 0.002S(t-2) - 0.002S(t-4) + 36.137DC(t-2) + 43.035DC(t-4) - 1.59DF(t-4) \qquad (14)$$

where $P_Q$, $DE$, $S$, $DC$ and $DF$ represent the number of quarterly protests, change in government expenditure, government surplus, change in CPI and change in FPI respectively.

### 5.3.3.4 VAR (1) model

A VAR (1) model, which incorporated government surplus, petrol price, income growth and internet usage, was created for the number of quarterly protests. The model parameters for this VAR (1) model are depicted in Table 24, where after, the model's equation is shown by Equation 15. While the government surplus, regional inland petrol price, income growth and an internet indicator variable were significant parameters at a 10% level of statistical significance, the intercept was not. With the use of these variables 86.59 percent of the variation present in the number of quarterly protests could be explained ($R^2 = 0.8659$).

*Table 24: The model parameter estimates of the quarterly VAR (1) model.*

| Equation | Parameter | Estimate | Standard Error | T Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| $\text{Protests}_Q(t)$ | CONST1 | 18.52053 | 26.05250 | 0.71 | 0.4794 | 1 |
| | XL0_1_1 | 44.76332 | 20.08685 | 2.23 | 0.0289 | Internet(t) |
| | AR1_1_1 | 0.69406 | 0.08119 | 8.55 | 0.0001 | $\text{Protests}_Q(t-1)$ |
| | AR1_1_2 | 0.00083 | 0.00031 | 2.66 | 0.0096 | Surplus(t-1) |
| | AR1_1_3 | -3.23525 | 1.87446 | -1.73 | 0.0886 | Income Growth(t-1) |
| | AR1_1_4 | 0.05363 | 0.02949 | 1.82 | 0.0730 | Inland Petrol(t-1) |

$$P_Q(t) = \begin{cases} 18.52 + 0.69P_Q(t-1) + 0.001S(t-1) - 3.24R(t-1) + 0.05P(t-1) \; if I(t) = 0 \\ 63.28 + 0.69P_Q(t-1) + 0.001S(t-1) - 3.24R(t-1) + 0.05P(t-1) \; if I(t) = 1 \end{cases} \quad (15)$$

where $P_Q$, $S$, $R$ and $I$ represent the number of quarterly protests, Government surplus, Income growth, Inland petrol prices and internet usage respectively. $I = 0$ if internet usage is less than 10% of the population, otherwise $I = 1$.

As with the previous VAR (1) model, the model diagnostics indicated a good fit for the data. The Granger-causality Wald test identified a causal interaction from government surplus, regional inland petrol price, income growth and the internet indicator variable to the number of quarterly protests. The Durbin-Watson test statistics value for this model is equal to 1.95046. This is sufficiently close to 2 which indicates that the autocorrelation at lag 1 is close to zero. There are also very few residuals that have significant autocorrelations, depicted by a + or – in Table 25, implying that the model fits well. Furthermore, the VAR (1) model is stationary as the modulus of all 3 indices, depicted in Table 26, are smaller than one.

*Table 25: The Schematic Representation of Cross Correlations of Residuals for the quarterly VAR (1) model.*

| Variable/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Protests | +... | .... | ...- | .... | .... | .... | .-.. | .... | .... | .... | .-.. | .... | +..- |

+ is > 2*std error, - is < -2*std error, . is between these values

*Table 26: Roots of AR Characteristic Polynomial for the quarterly VAR (1) model.*

| Index | Real | Imaginary | Modulus | Radian | Degree |
|---|---|---|---|---|---|
| 1 | 0.96437 | 0.00000 | 0.9644 | 0.0000 | 0.0000 |
| 2 | 0.62985 | 0.00000 | 0.6298 | 0.0000 | 0.0000 |
| 3 | 0.50764 | 0.00000 | 0.5076 | 0.0000 | 0.0000 |
| 4 | -0.12887 | 0.00000 | 0.1289 | 3.1416 | 180.0000 |

At times the relationships that are observed within the VAR (1) model appear to be contrary to the relationships that were observed when looking at the correlations. Government surplus, in this VAR (1) model, is such an example.

A negative relationship was observed between the number of quarterly protests at time t and government surplus at time t, depicted in Figure 22. Similarly, a negative relationship was observed between the number of quarterly protests at time t+1 and government surplus at time t, depicted in Figure 23. Contrary to this, the parameter estimates within the VAR (1) model points at a positive relationship.

This is not a mistake in the VAR (1) model. The VAR (1) model uses both the number of protests at time t and the government surplus at time t to estimate the number of protests at time t+1. As such the relationship modelled is actually the relationship that exists between the change in the number of quarterly protests from time t to t+1 and government surplus at time t. This relationship, which is depicted in Figure 24, is in fact positive.

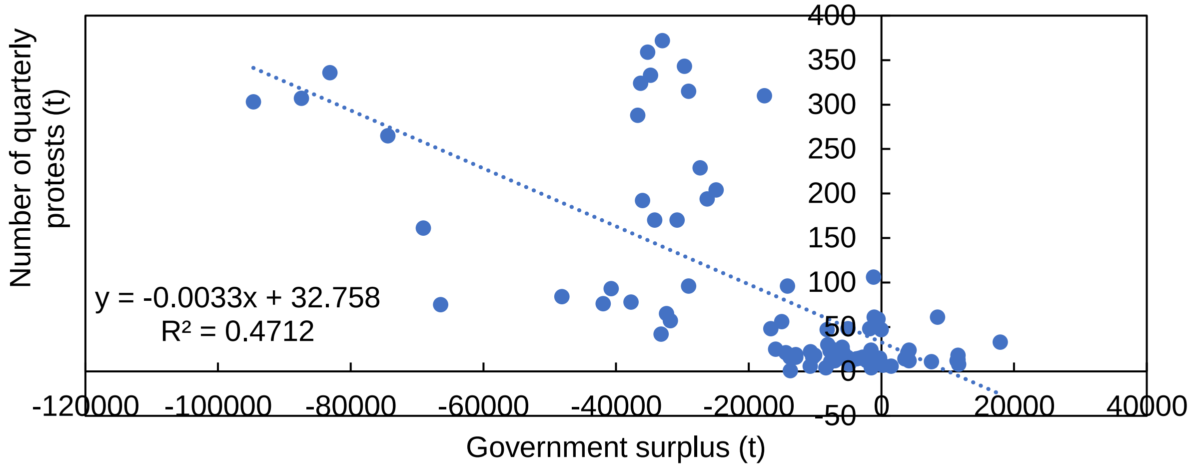



*Figure 22: The trendline between the number of quarterly protests at time t and government surplus at time t.*

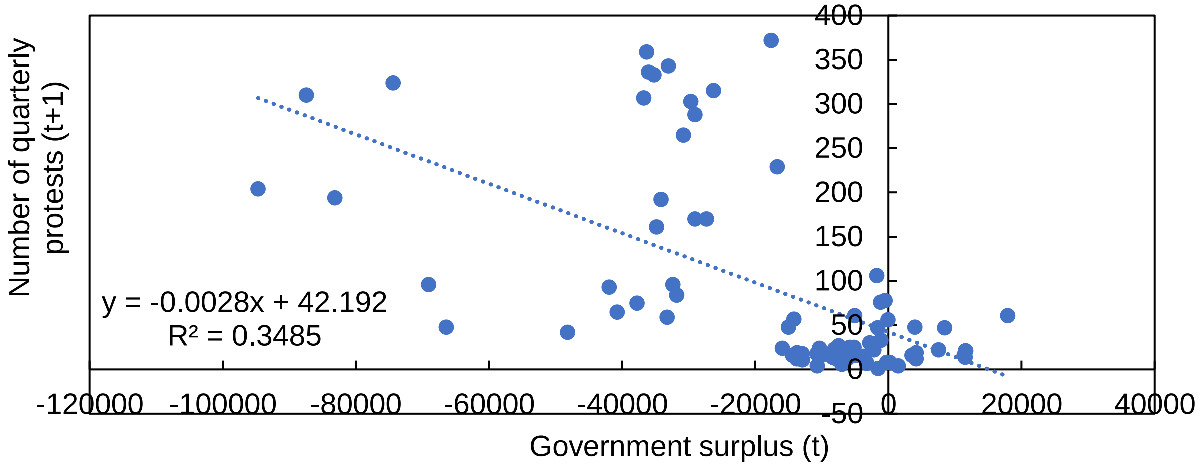



*Figure 23: The trendline between the number of quarterly protests at time t+1 and government surplus at time t.*

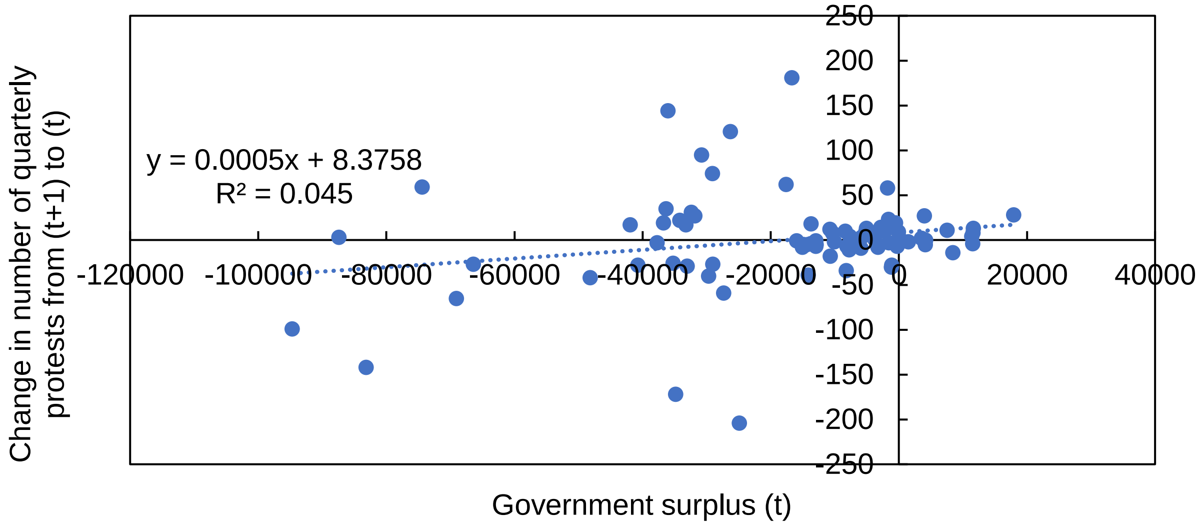



*Figure 24: The trendline between the change in the number of quarterly protests between time t+1 and t and government surplus at time t.*

#### 5.3.3.5 Comparison of the quarterly models

The univariate time series model struggled to control the autocorrelation within the residuals. As a result, it is an inadequate modelling technique for the number of quarterly protests. The three remaining models fared well, producing high $R^2$ values. Of these models, the VAR (1) model was the best at explaining the variation in the number of quarterly protests ($R^2 = 0.8659$). The lagged regression and linear regression models, with $R^2$ values of 0.7692 and 0.6886 respectively, also had high explanatory value.

Government expenditure and surplus, CPI, FPI, internet usage, inland petrol price and income growth were all identified as significant predictors, in at least one of the four quarterly models. While the exchange rate, government revenue, unemployment rate, diesel prices and paraffin prices were not identified as significant predictors in any of the quarterly models. Once more showing that significant relationships may not always result in the variables being significant predictors of social unrest.

### 5.3.4 Modelling the annual number of protests

There is an exponential trend present in the annual data. To account for this, the logarithm of protests was modelled. The estimated model parameters for this data are all significant and are depicted in Table 27, while the model's equation is shown by Equation 16. This model explains 87.21% of the variation that is present within the data ($R^2 = 0.8721$).

*Table 27: Model parameters for the regression model of number of annual protests.*

| Variable | DF | Parameter Estimate | Standard Error | T Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -387.71350 | 35.45778 | -10.93 | <.0001 |
| Year | 1 | 0.19580 | 0.01767 | 11.08 | <.0001 |

$$\ln(Protests_Y) = -387.71350 + 0.19580Y \qquad (16)$$

where $Y$ represents the year and $Protests_Y$ represent the number of protests per year.

The model diagnostics reveals that the residuals are stationary and uncorrelated. This model has a lot of explanatory power and it accounts for the trends that are present, making it a good model. Limited data, however, removes some of the validity of this model as it only considers 20 data points and may not be robust in the long term.

### 5.3.5 Comparing daily, monthly, quarterly and annual models with each other

All the daily, monthly, quarterly and annual models had power in explaining the variation observed in the protest data. A summary of the $R^2$ values produced by each of the protest models is shown in Table 28. The annual univariate model had the highest explanatory power ($R^2$ = 0.8721) and was closely followed by the quarterly VAR (1) model that had an $R^2$ value of 0.8659. While, the daily univariate time series model had the lowest explanatory value ($R^2$ = 0.4995).

*Table 28: Summary of the $R^2$ values produced by each of the South African protest models.*

| | Data frequency | | | |
|---|---|---|---|---|
| | Daily | Monthly | Quarterly | Annual |
| Univariate time series model | 0.4995 | 0.8100 | 0.8490 | 0.8721 |
| Linear regression model | - | 0.6279 | 0.6886 | - |
| Lagged regression model | - | 0.8138 | 0.7692 | - |
| VAR (1) model | - | 0.7655 | 0.8659 | - |

Irrespective of the high $R^2$ values produced by the univariate time series models, it is an inappropriate modelling technique. The daily, monthly and quarterly models were deemed inadequate as they were unable to control the autocorrelation within the residuals. While this was not a problem with the annual model, the limited data that the model was based on, made it an unsuitable model to make forecasts from.

The linear regression, lagged regression and VAR (1) models produced high $R^2$ values for both the monthly and quarterly models. The monthly lagged regression model was the best at explaining the variation within the number of monthly social unrest events. While the quarterly VAR (1) model best described the quarterly number of social unrest events.

Spearman's rank correlation identified various statistically significant interactions between social unrest and the economic and socio-economic variables. However, significant relationships did not always translate in the variables being significant predictors. Government surplus, revenue and expenditure, CPI, internet usage, inland petrol price and FPI were identified as significant predictors, in at least one of the four monthly models. All these variables, apart from government revenue, also had predictive power in the quarterly models. In addition to these six variables, income growth was also identified as a significant predictor in at least one of the four quarterly models.

To date researchers have tried to better understand factors that cause social unrest by analysing straight forward relationships that may exist between social unrest and various economic and socio-economic variables. The relationships that were identified by Spearman's rank correlation, trendline analysis and the various models showed that these relationships are anything but straight forward. They are not static in nature and may be influenced by the type of data and changes in data frequency. These complexities may explain some of the inconsistent relationships that have been found within the literature.

## 5.4 Testing the models forecasting abilities

In the field of insurance, it is important to evaluate the short and long-term forecasting abilities of the models that are used. This is done through model validation. Two types of model validation have been considered. The first, one-step forecasting procedure, examines the short-term performance of a model, while the n-step forecasting procedure tests the long-term performance of the model.

In Section 5.3, four different models (univariate time series, linear regression model, lagged regression model and VAR (1) model) were used to model and explain South African protests. Unfortunately, due to limited data availability and model specifications, model validation could not be performed on all the daily, monthly, quarterly and annual models. A description of the models that can be validated by each of these two forecasting procedures is given at the start of each technique.

The one-step forecasting procedure is described first. This is followed by the n-step forecasting procedure. Thereafter, a summary of the model validation is provided. An example and explanation of how the one-step and n-step procedures were performed on the VAR (1) model can be seen in Appendix G. Similarly, the one-step procedure for the lagged regression model is shown in Appendix H. Additional details on the forecasts produced by each model is described in Appendix I.

### 5.4.1 One-step forecasting procedure

The daily, monthly and quarterly univariate time series models described in Section 5.3 were all deemed to be inadequate. The annual univariate time series model was found to be adequate. There were however too few data points to forecast from. For these two reasons, neither one of the forecasting procedures were performed on any of the univariate time series models.

The monthly and quarterly linear regression models helped to explain the direct interactions and relationships that were present within the data. However, this model has limited forecasting abilities, as a forecast for every variable in the model would have to be produced before social unrest can be forecasted. Due to this, neither one of the forecasting procedures could not be performed on the two linear regression models.

Unlike the linear regression model, the lagged regression model makes use of previous (lagged) values. This gives the lagged regression model its forecasting abilities. The forecasting time frames for the monthly and quarterly lagged regression models are greater than one, which allows the model to be validated using the one-step forecasting procedure. Because the VAR (1) model is designed to produce forecasts, it can be validated using the one-step forecasting procedure.

As part of the one-step forecasting procedure process, forecasts were created for the monthly lagged regression model and monthly VAR (1) model. A combined model was also created to determine whether forecasts could be improved by combining these two models. This was purely for illustrative purposes and the forecasts for the combined model was equal to the average forecasts produced by the two original models. The one-step forecasting procedure was performed on the monthly lagged regression model, the monthly VAR (1) model and the new combined model. A summary of the error, absolute error, MAE and RMAE produced by these three models, is shown in Table 29.

The monthly lagged regression model overestimated the number of protests in 2016 by 27. In contrast, the monthly VAR (1) model and the combined models underestimated the number of protests by 60 and 17, respectively. Of the three models, the combined model had the smallest error over the 12-month period.

There was very little between the absolute error produced by the lagged regression model and the combined model. The lagged regression model had the lowest absolute error (195) and subsequently had the lowest RMAE (4.034). The absolute error for the combined model was 198, while the RMAE was 4.063. The monthly VAR (1) model had the highest absolute error (293) and RMAE (4.943).

Of the three-monthly models that have been validated, the monthly VAR (1) model performed the worst, producing the largest error, absolute error, MAE and RMAE. There was very little difference between the performance of the other two monthly models. The difference between the absolute error produced by the two models was three protests. This equated to the MAE and RMAE of these two models differing by 0.226 and 0.029 respectively. The most visible difference between the two models was seen in the error. The lagged regression model overestimated the total number of protests, while the combined model underestimated the total number of protests. Consequently, neither of these two models outperformed the other.

*Table 29: The one-step forecasting procedure on monthly protest models.*

| | Error | Absolute Error | MAE | RMAE |
|---|---|---|---|---|
| Lagged Regression Model | -27 | 195 | 16.275 | 4.034 |
| VAR (1) Model | 60 | 293 | 24.436 | 4.943 |
| Combined Models | 17 | 198 | 16.501 | 4.063 |

The one-step forecasting procedure on the quarterly models was performed in the same way as the monthly models. An additional model, combined quarterly model, was created by taking the average forecast for the quarterly lagged regression and VAR (1) models. A summary of the quarterly one-step forecasting procedure performed on the lagged regression model, the VAR (1) model and the combined model, is shown in Table 30.

The monthly VAR (1) model had the smallest error, an overestimate of 10 protests. Both the other models underestimated the number of protests. The error produced by the lagged regression model was 223, while the combination of the two models had an error of 112.

In contrast to what was seen in the monthly validation, the VAR (1) model produced the smallest absolute error (150) and thereby the smallest RMAE (6.116). While the lagged regression model had the largest absolute error (233) and RMAE (7.717). The values produced by the combined models were also not as close to the other two models, as was seen in the validation of the monthly models.

The quarterly VAR (1) model outperformed the lagged regression and combined model in every way. It had the lowest error, absolute error, MAE and RMAE. The combined model was the second-best model, while the quarterly lagged regression model was the weakest of the three models.

*Table 30: One-step forecasting procedure on quarterly protest models.*

| | Error | Absolute Error | MAE | RMAE |
|---|---|---|---|---|
| Lagged Regression Model | 233 | 238 | 59.501 | 7.717 |
| VAR (1) Model | -10 | 150 | 37.404 | 6.116 |
| Combined Models | 112 | 177 | 44.252 | 6.652 |

### 5.4.2 The n-step forecasting procedure

The univariate time series and linear regression models were not validated using the n-step forecasting procedure. The reason is described in Section 5.4.1. The maximum forecast length based on the monthly and quarterly lagged regression models are two months and two quarters, respectively. Both these forecasting periods are too short for the n-step forecasting procedure to be used. The VAR (1) model creates forecasts for each of the variables that are included in the model. This independence allows the model to forecast multiple steps into the future.

The n-step forecasting procedure was performed on the monthly and quarterly VAR (1) models, as they were the only models that had long term forecasting abilities. A summary of the results from the n-step forecasting procedure performed on the monthly and quarterly VAR (1) model, is shown in Table 31. Both models underestimated the number of protests during the 12-month period. The error produced by the monthly VAR (1) model was 356, while the quarterly VAR (1) model had an error of 113. The RMAE for the monthly VAR (1) model was lower than the RMAE of the quarterly VAR (1) model, this is despite the monthly VAR (1) model producing a lager error.

*Table 31: The n-step forecasting procedure of the monthly and quarterly VAR (1) models.*

| | Error | Absolute Error | MAE | RMAE |
|---|---|---|---|---|
| Monthly VAR (1) Model | 356 | 462 | 38.470 | 6.203 |
| Quarterly VAR (1) Model | 113 | 228 | 57.083 | 7.555 |

The n-step forecasting procedure entails forecasting multiple periods in advance to determine the model’s efficiency. When multiple forecasts are made using this model, it is imperative to consider seasonal abnormalities within the data, to ensure that the model’s efficiency is not misconstrued. The monthly seasonal trend of protests, described in section 5.3.2.1, indicated that the number of protests in December are abnormally low in comparison to any other month. The reason behind the change in behaviour is probably as a result of South Africa’s summer holiday season.

The VAR (1) model uses the number of protests at time t to estimate the number of protests at time t+1. This forecasted value is used to forecast t+2, which is then used for t+3 and so forth. The VAR (1) model is not a seasonal model and an adjustment for abnormal seasonal behaviour, as happens in December, is not made. The abnormally low value for this end point may cause the forecasts to be considerably different to what it would have been if the data ended one period earlier or later. This also affects the MAE and RMAE value for the model. Thus, the model may be deemed to be less efficient.

The n-step forecasting procedure on the monthly VAR model, shown in Table 31, was performed on data ending in December (forecasts started in January 2016). Additional n-step forecasting procedure was performed on the monthly VAR (1) model to illustrate how the efficiency changes when different starting points are used. Firstly, the n-step forecasting procedure was performed with forecasts starting on three different months, November 2015, December 2015 and January 2016, and all forecasts ended 12 months after the initial forecast was made. The corresponding error, absolute error, MAE and RMAE for each of the three starting points are shown in Table 32. A second set of the n-step forecasting procedure was performed. In this set five different starting months are considered (November 2015, December 2015, January 2016, February 2016 and March 2016). This time all the forecasts ended in December 2016. The corresponding error, MAE and RMAE for each of these five starting points are shown in Table 33.

*Table 32: The n-step forecasting procedure the monthly VAR (1) model using three different starting points with the forecasts ending 12 months later.*

| First Forecast | Last Forecast | The n-Step Forecasting Period | Error | Absolute Error | MAE | RMAE |
|---|---|---|---|---|---|---|
| Nov 2015 | Oct 2016 | 12 months | 157.349 | 336.897 | 28.075 | 5.299 |
| Dec 2015 | Nov 2016 | 12 months | 199.832 | 364.8949 | 30.408 | 5.514 |
| **Jan 2016** | **Dec 2016** | **12 months** | **356.024** | **461.6367** | **38.470** | **6.202** |

*Table 33: The n-step forecasting procedure the monthly VAR (1) model using five different starting points with the forecasts ending December 2016.*

| First Forecast | Last Forecast | The n-Step Forecasting Period | Error | MAE | RMAE |
|---|---|---|---|---|---|
| Nov 2015 | Dec 2016 | 14 months | 77.701 | 29.753 | 5.455 |
| Dec 2015 | Dec 2016 | 13 months | 141.928 | 32.523 | 5.703 |
| **Jan 2016** | **Dec 2016** | **12 months** | **356.024** | **38.470** | **6.202** |
| Feb 2016 | Dec 2016 | 11 months | 237.955 | 33.890 | 5.821 |
| Mar 2016 | Dec 2016 | 10 months | 91.957 | 27.451 | 5.239 |

Irrespective of the starting point or the n-step forecasting procedure period used, the error, MAE and RMAE was higher when the first forecast started in January 2016 (data ended in December) than for other starting points. This does not make the model inappropriate, but it is not advised to use December as the point in time to make multiple forecast from.

### 5.4.3 Summary of the one-step and n-step forecasting procedure

For the one-step forecasting procedure, the monthly lagged regression model and combined model outperformed the monthly VAR (1) model. The combined model underestimated the number of protests, while the lagged regression model overestimated the number of protests. Very few other differences were found between the lagged regression and combined models. This could point to a beneficial relationship between the monthly lagged regression and VAR (1) models, possibly mitigating a portion of the model risk.

In the short-term, the quarterly VAR (1) model outperformed the quarterly lagged regression and combined model in every way. The combined model was the second-best model, once more indicating that the two modelling procedures capture different aspects of social unrest.

The n-step forecasting procedure could only be performed on the monthly and quarterly VAR (1) models. As would be expected, the short-term performance of both the monthly and quarterly VAR (1) models were better than their long-term performance. An explanation for why this is expected is given in Appendix I.

In terms of the MAE and RMAE, the monthly VAR (1) model outperformed the quarterly VAR (1) model. The error produced by the monthly model with the n-step forecasting procedure from December 2015 was quite large, partly due to the abnormally low seasonal trend within the monthly protest data. When the monthly VAR (1) model was forecasted for the two months prior and two months after December, the model performance improved. This is a limitation within the VAR (1) model, as it does not adjust to abnormal seasonal trends within the data. It is therefore important to consider abnormal seasonal trends within the data and it is advisable that December not be used as a starting point for multiple forecasts.

In the insurance world, forecasting multiple steps in advance is required. Although the lagged regression model performed well in the short term, the long-term forecasting ability of the model is a limitation. The long-term forecasting ability of the lagged regression model can be improved by modelling the economic variables that it makes use of. The lagged regression model makes use of numerous variables, with each variable that needs to be modelled, addition model risk is introduced, which increases the uncertainty within the forecast. In addition, this process is highly time consuming.

The VAR (1) and lagged regression models both have limitations. The VAR (1) model is not suited for abnormal seasonal trends within the data, while the lagged regression model has problems with its forecasting ability. The VAR (1) models' limitation can be mitigated by considering the seasonal trends when choosing a point from which to forecast from. It is however, not that easy to improve the long-term forecasting abilities of the lagged regression model. This makes the VAR (1) model better at modelling social unrest in South Africa.

## 5.5 Modelling subcategories of protests

In Section 5.3, various models are described for the incidence of social unrest and the impact that data frequency has on these relationships were discussed. This was done in isolation, without investigating the interactions present between various subcategories of protests. These models are very important as they shed light on protest trends within South Africa, but they do not enhance the understanding of how different economic and socio-economic variables play a role in different aspects of protests.

In this part, two main subcategories of protests are briefly evaluated. The first is the use of violence and property damage, while the second subcategory considers the reason for the protest. The three main reasons for social unrest events in South Africa is education, labour and municipal-related protests and each one of these are examined individually.

Summarised versions of these subcategories are discussed in this section. Additional information on the correlations and models can be found in the Appendices. The Spearman's rank correlations are shown in Appendix D. The univariate time series, regression and VAR (1) models are depicted in Appendices J, K and L, respectively, while, a summary of the SAS code that was used throughout this study is displayed in Appendix M.

### 5.5.1 Incidence of violence and property damage

Most of the insurable risks associated with protests lie in violent protests and protests where property damage is caused. From an insurance point of view, it is essential to understand the individual dynamics of these two subcategories and how they compare to all the protests. Changes in the characteristics of the seasonal trends, Spearman's rank correlation, regression models and the VAR (1) models are described below.

The day with the highest incidence of protests (Wednesday) and month with the lowest incidence of protests (December) were identical for all these three categories. Variations in all the other seasonal trends were however observed. Protests with property damage had a daily minimum on Saturdays, while the other two categories had their minimums on Sundays. Violent protest's monthly maximum occurred in February, whereas March had the highest incidence for both protests with property damage and all the protests. Furthermore, in contrast to all the protests, no quarterly seasonal trends where identified for either violent protests or protests with property damage.

Unlike the variations observed in the seasonal trends, the general directionality of the relationships identified by Spearman's rank correlation were consistent to those seen for all social unrest events, regardless of the subcategory examined. Slight changes were, however, observed in the strength and presence of statistically significant relationships.

The monthly regression models also failed to identify significant changes in the interactions between all the protests and the two subcategories of protests. The variables that were significant predictors in the monthly regression model for all protests, government revenue and surplus, also had predictive power in the monthly regression models of both subcategories. The regression model for all monthly protests produced a $R^2$ value of 0.6279. The $R^2$ values for the regression models of the two subcategories were only marginally lower than this. The associated $R^2$ values for these two models were 0.5662 for violent protests and 0.4932 for protests with property damage.

In contrast to the monthly regression models, the quarterly regression models found that the variables that had predictive power, differed from one another. The regression model for all protests had a $R^2$ of 0.6886 and made use of the income growth, internet indicator, FPI and CPI. The regression model for violent protests produced a higher $R^2$ value (0.7024) and made use of government surplus, income growth, CPI, inland petrol prices and an internet indicator variable. While the model for protests with property damage only made use of government expenditure and had a $R^2$ of only 0.0782.

Like the quarterly regression models, changes were observed in the VAR (1) models. The VAR (1) model for all protests had a $R^2$ of 0.7655 and made use of the FPI, inland petrol prices and internet indicator variable. The monthly VAR (1) model for violent protests had a $R^2$ of 0.7492 and made use of inland petrol prices, paraffin prices and monthly indicator variables. While the monthly VAR (1) model for protests with property damage produced a $R^2$ value of 0.6618 and included FPI and CPI. It was not only the variables that were different. A quarterly VAR (1) model was created for all protests. It produced a $R^2$ of 0.8659 and made use the government surplus, income growth, petrol prices and internet usage. However, quarterly VAR (1) models could not be created for either violent protests or protests with property damage.

There are variations within the natural trends of all protests, violent protests and protests with property damage. Spearman's rank correlation however, identified very few changes in the economic and socio-economic relationships, irrespective of the subcategory considered. Nevertheless, the variables that were statistically significant predictors, varied greatly from one another. This shows the effect that economic and socio-economic variables have on social unrest, changes as the dynamics of social unrest events is altered. Regardless of the variations in the interactions, VAR (1) models with high explanatory power were created for the number of violent protests and protests with property damage. These models are shown in Appendix L.

### 5.5.2 Reason for protest

The three main reasons for protests in South Africa, from 1997 till 2016 were education, labour and municipal services. Similar analyses that were performed on all protests were also done for each of these three reasons, to determine whether there were any changes in the interactions. The similarities and differences in the characteristics of the seasonal trends, Spearman's rank correlation, regression models and VAR (1) models of each of these three subcategories are described below.

Major discrepancies were observed in the seasonal trend between all protests and the three subcategories. Where daily, monthly and quarterly seasonal trends were observed in all protests, municipal service-related protest data contained daily and monthly seasonal trends. While only a daily seasonal trend was present within education-related protest data. Finally, no seasonal trends were observed in the labour-related protest data.

There was also very little consensus in the high and low points within these daily and monthly seasonal trends. Where all protests and municipal service-related protests had daily highs on Wednesday's, education-related protests peaked on Mondays. Education and municipal services related protests slumped on Saturday's, while all protests slumped a day later. The monthly peaks of all protests and municipal service-related protests also varied, with highs occurring in March and May, respectively. The only consensus was observed in the low point in the monthly trends, with both slumping in December.

In contrast to what was seen in seasonal trends, Spearman's rank correlation of these three subcategories rendered similar results to what was seen in all the protests. Changes were observed in the strength of some of the relationships, however the directionality of the relationships were generally consistent, irrespective of the data considered.

The monthly regression models for labour and education related protests made use of the same variables that were used in the regression model for all protests, namely government revenue and surplus. Unfortunately, these two models had less explanatory value than the model for all protests ($R^2$ = 0.6279), with the education related protest model producing a $R^2$ of 0.2826, while the labour-related model had a $R^2$ of 0.2352. The regression model for municipal-related protests varied from the three other models. It made use of inland petrol prices, CPI and an internet indicator variable and produced a $R^2$ of 0.4175.

Unlike the monthly regression models, the quarterly regression models of the three subcategories varied greatly from those that were seen in the model for all protests. Education-related protests incorporated the most variables, using government surplus and revenue, income growth, CPI and inland petrol prices ($R^2$ = 0.5114). The model for municipal service-related protests made use of government revenue and an internet indicator variable ($R^2$ = 0.5373), while labour-related protests only considered the internet indicator variable ($R^2$ = 0.3199). Although these models rendered higher $R^2$ values than their monthly counterparts, the regression model for all protests, which used the income growth, internet indicator, FPI and CPI, once again produced a highest $R^2$ value ($R^2$ = 0.6886).

The largest divergence between the three subcategories and all protests were visible in the VAR (1) models. The models of both education and labour related protests were unable to control the autocorrelation within the monthly and quarterly data. As a result, the VAR (1) modelling technique is inappropriate for these two subcategories. On the other hand, the VAR (1) models created for the monthly and quarterly incidence of municipal-related protests had no such problems.

The VAR (1) model for the monthly incidence of municipal-related protests made use of four main factors. Two of these, FPI and an internet indicator variable, corresponded to what was seen in the VAR (1) model for all protests. It also included monthly indicator variables and was concluded with the diesel price, which replaced the inland petrol price seen in the model for all protests. This model had a $R^2$ value of 0.563, which was lower than the 0.7655 seen for all protests. The quarterly VAR (1) model for municipal-related protests was completely different from what was seen for the quarterly model for all protests which had a $R^2$ of 0.8659. This model only made use of one variable, an internet indicator variable. As was the case with all the models within this section, the $R^2$ value produced by this model ($R^2$ = 0.5955) was lower than what was produced by the model for all protests ($R^2$ = 0.8659).

The interactions between social unrest and economic and socio-economic variables are complicated. This was once again highlighted by the intricate differences within the characteristics of all protests and the three main reasons for protests. Spearman's rank correlation was the only measures that showed that there was little difference between relationships of the four groups. Discrepancies were however, identified in all the other characteristics that were analysed. There was almost no consensus in seasonal trends of the four groups. The variables that were significant predictors varied based on the subcategory analysed. Furthermore, the VAR (1) modelling technique that best described the number of protests and did well to explain municipal service-related protests, was deemed inappropriate for two of the three reasons for protests. It is important to realise that the reason for the protest is a major factor that affects the interactions observed. For this reason, more research is required in this field to allow for the protest behaviour in South Africa to be understood.

# Chapter 6: Conclusion and Recommendations

Social unrest in South Africa has not been studied in great detail and in recent years this has given rise to various challenges. It has become more important to gain insight into the variables that play a role in South African protests. These insights can allow for better insurance models to be developed and even mitigate some protests through the adjustment of external variables.

Due to the limited research performed in this field, a database containing the protest behaviour of South Africa needed to be set up before any additional analysis could be performed. An additional database containing the South African economic and socio-economic variables that could influence social unrest was also set up. These two databases formed the foundation for all the results obtained in this study.

Spearman's rank correlation and trendline analysis were used to gain insight into the relationships between economic and socio-economic variables and protest behaviour in South Africa. Both national and provincial data as well as monthly, quarterly and annual data were used in this analysis. The results showed some similarities to published studies, but also showed some interesting behaviours that are more specific to South Africa.

The same databases were used to create four different types of predictive models (univariate time series models, linear regression, lagged regression, VAR (1)) over different time intervals (daily, monthly, quarterly and annual). Models with a good fit were built using the monthly and quarterly data, this was not possible with the daily and annual data. Significant Spearman's rank correlations between social unrest events and economic and socio-economic variables may not always translate to the variables being significant predictors of social unrest. This was clearly seen in the models that had the best predictive power over different time intervals.

Of the monthly models, the lagged regression model produced the highest $R^2$ value, equal to 0.8138. The VAR (1) model's $R^2$ value was marginally lower, equal to 0.7655, while the regression model had a $R^2$ of 0.6279. Government surplus, revenue and expenditure, CPI, internet usage, the inland petrol price and FPI were identified as significant predictors, in at least one of these three models.

The results from the quarterly models varied from the monthly models. Here the VAR (1) model had the highest $R^2$ value, equal to 0.8659. Whereas the regression and lagged regression models yielded $R^2$ values of 0.6886 and 0.7692, respectively. Six of the variables that had predictive power in the monthly models were also significant predictors in the quarterly models. These included government surplus and expenditure, CPI, internet usage, the inland petrol price and FPI. In addition to these variables, income growth was also identified as a significant predictor.

Validation of the models found that the short-term forecasting abilities of the monthly lagged regression model outperformed the monthly VAR (1) model. The converse was seen for the quarterly models, where the VAR (1) model outperformed the lagged regression model. Limitations exist within the long-term forecasting abilities of both the monthly and quarterly lagged regression models. The VAR (1) models did not have this problem. It was found that the monthly VAR (1) model's long-term forecasting abilities outperformed those of the quarterly VAR (1) model. In the insurance world, a lot of focus is placed on forecasting multiple steps in the future. Therefore, the VAR (1) model is deemed to be more suited for modelling South African social unrest.

To further enhance the understanding of South African protests, five subcategories of protests were briefly evaluated, namely violent protests, protests with property damage, education, labour and municipal service-related protests. Spearman's rank correlation of these five subcategories rendered similar results to what was seen when all the protests were considered. A few differences were observed in the strength of some of the relationships, however the directionality of these relationships was generally consistent, irrespective of the subcategory considered. The modelling of the different subcategories identified limitations that were not present when all the data were considered. The models of certain time intervals of the various subcategories could not deliver any usable results.

Further studies are required to find modelling techniques that are able to explain the subcategories present within social unrest. Other modelling techniques like the Threshold autoregressive model as well as switching Markov chains may be examined, to evaluate their effectiveness. Additional research is required to determine whether there are different characteristics between once off and ongoing protests. Furthermore, studies can focus on the number of protests per capita and on the incidence of protests after 2012.

To date, researchers have tried to better understand factors that cause social unrest by analysing straight forward relationships that may exist between social unrest and various economic and socio-economic variables. The relationships that were identified by Spearman's rank correlation, trendline analysis and the various models show that these relationships are anything but straight forward. They are not static in nature and may be influenced by the type of data, changes in data frequency, the reasons behind protests as well as the incidence of violence and property damage. These complexities may explain some of the inconsistent relationships that were observed within the literature.

With the Spearman's rank correlations and the different models that have been generated in this study it is clear that the protest behaviour in South Africa is multifaceted. It is influenced by several variables ranging from government spending to the availability of internet access. The insights gained here allowed for many of the uncertainty surrounding protests in South Africa to clarified.

Protests in South Africa has become a way for people to have their voices heard. The insight gained from this study forms a strong basis for understanding and assessing the risks associated with social unrest in South Africa. By making social unrest more predictable, the way protests are approached, as well as the risks that are associated with protests can be changed and improved. This may be beneficial to the South African insurance market and the country as a whole.

# Reference list


Airmic Technical (2013). Terrorism Insurance Review. London: Airmic Technical, pp.2-40.

Albertus, M., Brambor, T. and Ceneviva, R. (2016). Land Inequality and Rural Unrest: Theory and Evidence from Brazil. Journal of Conflict Resolution, 62(3), pp.557-596.

Alesina, A. and Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, 40(6), pp.1203-1228.

Alesina, A., Özler, S., Roubini, N. and Swagel, P. (1996). Political instability and economic growth. *Journal of Economic Growth*, 1(2), pp.189-211.

Alexander, P. (2010). Rebellion of the poor: South Africa's service delivery protests – a preliminary analysis. *Review of African Political Economy*, 37(123), pp.25-40.

Alexander, P. and Pfaffe, P. (2014). Social Relationships to the Means and Ends of Protest in South Africa's Ongoing Rebellion of the Poor: The Balfour Insurrections. *Social Movement Studies*, 13(2), pp.204-221.

An, L. (2012). Modelling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229, pp.25-36.

Andersen, T. (2005). A Strategic Risk Management Framework for Multinational Enterprise. *SSRN Electronic Journal*.

Armborst, A. (2010). Modelling Terrorism and Political Violence. *International Relations*, 24(4), pp.414-432.

Arpan, L., Baker, K., Lee, Y., Jung, T., Lorusso, L. and Smith, J. (2006). News Coverage of Social Protests and the Effects of Photographs and Prior Attitudes. *Mass Communication and Society*, 9(1), pp.1-20.

Asagba, R. (2008). The Logotherapist View of Xenophobia and Violence in South Africa. *IFE PsychologIA*, 16(3).

Bellemare, M. (2014). Rising Food Prices, Food Price Volatility, and Social Unrest. *American Journal of Agricultural Economics*, 97(1), pp.1-21.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Supplement 3), pp.7280-7287.

Buhaug, H. and Urdal, H. (2013). An urbanization bomb? Population growth and social disorder in cities. *Global Environmental Change*, 23(1), pp.1-10.

Burrows, M. and Harris, J. (2009). Revisiting the Future: Geopolitical Effects of the Financial Crisis. *The Washington Quarterly*, 32(2), pp.27-38.

Caruso, R. and Schneider, F. (2011). The socio-economic determinants of terrorism and political violence in Western Europe (1994–2007). *European Journal of Political Economy*, 27, pp. S37-S49.

Casilli, A. and Tubaro, P. (2012). Social Media Censorship in Times of Political Unrest - A Social Simulation Experiment with the UK Riots. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 115(1), pp.5-20.

Cesari, J. (2018). *Ethnicity, Islam, and les banlieues: Confusing the Issues*. [online] Riotsfrance.ssrc.org. Available at: http://riotsfrance.ssrc.org/Cesari/ [Accessed 25 Apr. 2018].

Chakravarty, A. (2015). *Time for a Global Legal Scheme for Terrorism Risk Insurance*. [online] Boston University School of Law: International Law Journal. Available at: https://www.bu.edu/ilj/2015/01/07/time-for-a-global-legal-scheme-for-terrorism-risk-insurance/ [Accessed 30 Mar. 2016].

Collins English Dictionary. (2018). In: *Collins English dictionary*, 1st ed. [online] Available at: https://www.collinsdictionary.com/dictionary/english [Accessed 16 Mar. 2018]

Cushing, B. and Poot, J. (2004). Crossing boundaries and borders: Regional science advances in migration modelling. *Papers in Regional Science*, 83(1), pp.317-338.

Dabrowski, J. and de Villiers, J. (2015a). A unified model for context-based behavioural modelling and classification. *Expert Systems with Applications*, 42(19), pp.6738-6757.

Dabrowski, J. and de Villiers, J. (2015b). Maritime piracy situation modelling with dynamic Bayesian networks. *Information Fusion*, 23, pp.116-130.

Daly, G. (1996). Migrants and gate keepers. *Cities*, 13(1), pp.11-23.

Daras, N. and Mazis, J. (2014). Systemic geopolitical modeling. Part 1: prediction of geopolitical events. *GeoJournal*, 80(5), pp.653-678.

Daras, N. and Mazis, J. (2015). Systemic geopolitical modeling. Part 2: subjectivity in prediction of geopolitical events. *GeoJournal*, 82(1), pp.81-108.

Dewey, T., Kaden, J., Marks, M., Zhu, B. and Matsushima, S. (2012). *The Impact of Social Media on Social Unrest in the Arab Spring*. Stanford: Stanford University, pp.2-76.

Dickey, D. and Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366), pp.427-431.

DiPasquale, D. and Glaeser, E. (1998). The Los Angeles Riot and the Economics of Urban Unrest. *Journal of Urban Economics*, 43(1), pp.52-78.

du Toit, A. (1993). *Understanding South African Political Violence*. UNRISD Discussion Paper 43. Geneva: UNRISD, pp.3-39.

Durbin, J. and Watson, G. (1950). Testing for Serial Correlation in Least Squares Regression. I. *Biometrika*, 37(3-4), pp.409-428.

Durbin, J. and Watson, G. (1951). Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*, 38(1/2), pp.159-177.

Elbadawi, I. and Sambanis, N. (2000). Why are there so many civil wars in Africa? Understanding and preventing violent conflict. *Journal of African Economics*, 9(3), pp.244-269.

Engle, E. (2008). Agent Models of Political Interactions. *SSRN Electronic Journal*.

Fearon, J. and Laitin, D. (2003). Ethnicity, Insurgency, and Civil War. American Political Science Review, 97(01), pp.75-90.

Fjelde, H. and Østby, G. (2014). Socioeconomic Inequality and Communal Conflict: A Disaggregated Analysis of Sub-Saharan Africa, 1990–2008. *International Interactions*, 40(5), pp.737-762.

Fox, S. and Bell, A. (2016). Urban geography and protest mobilization in Africa. *Political Geography*, 53, pp.54-64.

Goldin, I. and Vogel, T. (2010). Global Governance and Systemic Risk in the 21st Century: Lessons from the Financial Crisis. *Global Policy*, 1(1), pp.4-15.

Hamilton, J. (1994). *Time series analysis*. 1st ed. Princeton, New Jersey: Princeton University, pp.1-474.

Hendrix, C. and Haggard, S. (2015). Global food prices, regime type, and urban unrest in the developing world. *Journal of Peace Research*, 52(2), pp.143-157.

Howard, P. and Hussain, M. (2013). Democracy's fourth wave: digital media and the Arab Spring. *Choice Reviews Online*, 51(04), pp.51-2322-51-2322.

IFTRIP (2017). World Terrorism Insurance Pools and Schemes. In: *International Forum for Terrorism Risk (Re)Insurance Pools*. Paris.

ISS Crime Hub (2018a). *National crime stats*. [online] ISS Africa. Available at: https://issafrica.org/crimehub/facts-and-figures/national-crime [Accessed 1 May 2018].

ISS Crime Hub (2018b). *Province crime stats*. [online] ISS Africa. Available at: https://issafrica.org/crimehub/facts-and-figures/provincial-crime [Accessed 1 May 2018].

ISS Crime Hub (ISS) (2017). *Public protest and violence map*. [online] Issafrica.org. Available at: https://issafrica.org/crimehub/maps/public-violence [Accessed 5 Aug. 2017].

Jensen, N. and Young, D. (2008). A Violent Future? Political Risk Insurance Markets and Violence Forecasts. *Journal of Conflict Resolution*, 52(4), pp.527-547.

Khmelko, I. and Pereguda, Y. (2014). An Anatomy of Mass Protests: The Orange Revolution and Euromaydan Compared. *Communist and Post-Communist Studies*, 47(2), pp.227-236.

Klomp, J. and de Haan, J. (2013). Popular protest and political budget cycles: A panel data analysis. *Economics Letters*, 120(3), pp.516-520.

Kokott, K. (2005). The Impact of 9/11 on the South African Anti-Terrorism Legislation and the Constitutionality thereof. Master of Law. University of Western Cape.

Kuhn, R. (2012). On the Role of Human Development in the Arab Spring. *Population and Development Review*, 38(4), pp.649-683.

Le Billon, P. (2004). The Geopolitical economy of ‘resource wars’. *Geopolitics*, 9(1), pp.1-28.

Leistedt, S. (2013). Behavioural aspects of terrorism. *Forensic Science International*, 228(1-3), pp.21-27.

Levinsohn, A. (2002). New geopolitics spotlights political risk management. *Strategic Finance*, 83(7), pp.38-43.

Linke, A., Schutte, S. and Buhaug, H. (2015). Population Attitudes and the Spread of Political Violence in Sub-Saharan Africa. *International Studies Review*, 17(1), pp.26-45.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), pp.297-303.

Lodge, T. and Mottiar, S. (2016). Protest in South Africa: motives and meanings. *Democratization*, 23(5), pp.819-837.

Lord Carlile of Berriew Q.C. (2007). *The definition of Terrorism*. London: Crown Copyright, pp.2-52.

Mabasa, P. (2010). Political risk insurance in Africa – The Sasria Solution. In: *African Reinsurance Forum*. Nairobi: African Reinsurance Forum, pp.1-22.

Major, J. (2002). Advanced Techniques for Modeling Terrorism Risk. *The Journal of Risk Finance*, 4(1), pp.15-24.

Malmberg, B., Andersson, E. and Östh, J. (2013). Segregation and Urban Unrest in Sweden. *Urban Geography*, 34(7), pp.1031-1046.

McCreight, R. (2013). Scenario development: using geopolitical wargames and strategic simulations. *Environment Systems & Decisions*, 33(1), pp.21-32.

Mei, S., Zarrabi, N., Lees, M. and Sloot, P. (2015). Complex agent networks: An emerging approach for modeling complex systems. *Applied Soft Computing*, 37, pp.311-321.

Mountz, A. and Hiemstra, N. (2014). Chaos and Crisis: Dissecting the Spatiotemporal Logics of Contemporary Migrations and State Practices. *Annals of the Association of American Geographers*, 104(2), pp.382-390.

Nesvetailova, A. and Palan, R. (2008). A Very North Atlantic Credit Crunch: Geopolitical Implications of The Global Liquidity Crisis. *Journal of International Affairs*, 62(1), pp.165-185.

Nizamani, S., Memon, N. and Galam, S. (2014). From public outrage to the burst of public violence: An epidemic-like model. *Physica A: Statistical Mechanics and its Applications*, 416, pp.620-630.

Østby, G. (2008). Polarization, Horizontal Inequalities and Violent Civil Conflict. *Journal of Peace Research*, 45(2), pp.143-162.

Østby, G., Urdal, H., Tadjoeddin, M., Murshed, S. and Strand, H. (2011). Population Pressure, Horizontal Inequality and Political Violence: A Disaggregated Study of Indonesian Provinces, 1990–2003. *Journal of Development Studies*, 47(3), pp.377-398.

Oxford Dictionary. (2018). In: *Oxford dictionary*, 1st ed. [online] Available at: https://en.oxforddictionaries.com/ [Accessed 16 Mar. 2018]

Papadopoulos, A. and Fratsea, L. (2013). Between Insecurity and Integration: Sub-Saharan African Migrants in Crisis-Stricken Greece. *Regions Magazine*, 289(1), pp.5-9.

Park, A., Tsang, H., Sun, M. and Glässer, U. (2012). An agent-based model and computational framework for counterterrorism and public safety based on swarm intelligence. *Security Informatics*, 1(1), p.23.

Parvin, M. (1973). Economic Determinants of Political Unrest. *Journal of Conflict Resolution*, 17(2), pp.271-296.

Ponticelli, J. and Voth, H. (2011). *Austerity and Anarchy: Budget Cuts and Social Unrest in Europe, 1919-2009*. London: Centre for Economic Policy Research, pp.3-30.

Protection of Constitutional Democracy Against Terrorist and Related Activities Act 33 of 2004 (RSA)

Raleigh, C. (2014). Political hierarchies and landscapes of conflict across Africa. *Political Geography*, 42, pp.92-103.

Raleigh, C. and Hegre, H. (2009). Population size, concentration, and civil war. A geographically disaggregated analysis. *Political Geography*, 28(4), pp.224-238.

Raleigh, C., Linke, A., Hegre H., and Joakim. K., (2010). Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research*, 47(5), p.p.651-660.

Ramakrishnan, N., Korkmaz, G., Kuhlman, C., Marathe, A., Zhao, L., Hua, T., Chen, F., Lu, C., Huang, B., Srinivasan, A., Trinh, K., Butler, P., Getoor, L., Katz, G., Doyle, A., Ackermann, C., Zavorin, I., Ford, J., Summers, K., Fayed, Y., Arredondo, J., Gupta, D., Muthiah, S., Mares, D., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J. and Vullikanti, A. (2014). 'Beating the news' with EMBERS. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*.

Rânceanu, D. and Marghescu, R. (2015). Smuggling of IS-refugees. In: *Strategies XXI: Strategic changes in security and international relations*. Bucharest: Carol I National Defence University, pp.102-113.

Renn, O., Jovanovic, A. and Schröter, R. (2011). A Theoretical Approach towards Understanding Social Unrest*. Stutgart: European Virtual Institute for Integrated Risk Management,* pp.3-66.

Salazar-Moreno, R., López-Cruz, I. and Sánchez Cruz, A. (2019). Dynamic energy balance model in a greenhouse with tomato cultivation: simulation, calibration and evaluation. *Revista Chapingo Serie Horticultura*, 25(1), pp.45-60.

Seymour, A., Murray, M., Sigmon, J., Hook, M., Edmunds, C., Gagboury, M. and Coleman, G. (2002). *National Victim Assistance Academy Textbook*. [Washington, D.C.]: Office for Victims of Crimes.

Stillwell, J., Daras, K., Bell, M. and Lomax, N. (2014). The IMAGE Studio: A Tool for Internal Migration Analysis and Modelling. *Applied Spatial Analysis and Policy*, 7(1), pp.5-23.

Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), pp.437-450.

Terrorism Risk Insurance Act of 2002 (USA) s102

Tuathail, G. (1999). Understanding critical geopolitics: Geopolitics and risk society. *Journal of Strategic Studies*, 22(2-3), pp.107-124.

Tufekci, Z. and Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), pp.363-379.

Vally, N., (2009). National trends around protest action: mapping protest action in South Africa, Study undertaken on behalf of the CSR and the Freedom of Expression Institute, presented to CSR and Development Studies Seminar.

Viscusi, W. (2009). Valuing risks of death from terrorism and natural disasters. *Journal of Risk and Uncertainty*, 38(3), pp.191-213.

Wackerly, D., Mendenhall, W. and Scheaffer, R. (2008). Mathematical Statistics with Applications. 7th ed. United Kingdom: Thomson Brooks/Cole, pp.568-586.

Wallace, C., Pennell, M. and Robertson, N. (2014). Terrorism Risk Insurance in Australia. In: *Terrorism Risk Insurance in Australia*. Sydney: Australian Reinsurance Pool Corporation, pp.2-24.

Weinberg, J. and Bakker, R. (2015). Let them eat cake: Food prices, domestic policy and social unrest. *Conflict Management and Peace Science*, 32(3), pp.309-326.

Wolfsfeld, G., Segev, E. and Sheafer, T. (2013). Social Media and the Arab Spring. *The International Journal of Press/Politics*, 18(2), pp.115-137.

Yazici, B. and Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, 77(2), pp. 175-183.

# Reference list: Data sources

Freedom House. (2017). *Freedom of The Press 1980 – 2017 Public data.* [Microsoft Excel spreadsheet]. Freedom House. Available at: https://freedomhouse.org/report-types/freedom-press [Accessed 10 June 2017].

Food and Agriculture Organisation of the United Nations. (2017). *Suite of Food Security Indicators.* [online]. Food and Agriculture Organisation of the United Nations. Available at: http://www.fao.org/faostat/en/#data/FS [Accessed 1 June 2017].

Food and Agriculture Organisation of the United Nations. (2018). *Food price indices data.* [Microsoft Excel spreadsheet]. Food and Agriculture Organisation of the United Nations. Available at: http://www.fao.org/worldfoodsituation/foodpricesindex/en/ [Accessed 17 July 2018].

Heritage Foundation. (2017). *2017 Index of Economic Freedom*. [online] Heritage Foundation. Available at: https://www.heritage.org/index/explore?view=by-region-country-year [Accessed 10 June 2017].

Republic of South Africa's Department of Energy. (1997). *Basic fuel price 1997*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_1997.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (1998). *Basic fuel price 1998*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_1998.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (1999a). *Basic fuel price 1999*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_1999.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (1999b). *History of petrol prices*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/history_petrol_price.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (1999c). *History of paraffin*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/history_paraffin_price.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (1999d). *Wholesale prices of diesel*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/history_diesel_price.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2000a). *Basic fuel price 2000*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2000.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2000b). *South African fuel prices history 2000*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2000.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2001a). *Basic fuel price 2001*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2001.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2001b). *South African fuel prices history 2001*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2001.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2002a). *Basic fuel price 2002*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2002.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2002b). *South African fuel prices history 2002*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2002.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2003a). *Basic fuel price 2003*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2003.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2003b). *South African fuel prices history 2003*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2003.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2004a). *Basic fuel price 2004*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2004.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2004b). *South African fuel prices history 2004*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2004.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2005a). *Basic fuel price 2005*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2005.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2005b). *South African fuel prices history 2005*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2005.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2006a). *Basic fuel price 2006*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2006.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2006b). *South African fuel prices history 2006*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2006.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2007a). *Basic fuel price 2007*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2007.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2007b). *South African fuel prices history 2007*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2007.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2008a). *Basic fuel price 2008*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2008.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2008b). *South African fuel prices history 2008*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2008.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2009a). *Basic fuel price 2009*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2009.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2009b). *South African fuel prices history 2009*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2009.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2010a). *Basic fuel price 2010*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Achieves_Petrol/petroleum_2010.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2010b). *South African fuel prices history 2010*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/historyprice2010.html [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2011a). *Basic fuel price 2011*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Dec2011/BasicFuelPrice.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2011b). *South African fuel prices history 2011*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Dec2011/FuelPriceHistory.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2012a). *Basic fuel price 2012*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2012/Basic_Fuel_Price.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2012b). *South African fuel prices history 2012*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2012/Fuel_Price_History.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2013a). *Basic fuel price 2013*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Dec-2013/Basic-Fuel-Price.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2013b). *South African fuel prices history 2013*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/Dec-2013/Fuel-Price-History.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2014a). *Basic fuel price 2014*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2014/Basic-Fuel-Price.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2014b). *South African fuel prices history 2014*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2014/Fuel-Price-History.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2015a). *Basic fuel price 2015*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2015/Basic-Fuel-Price.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2015b). *South African fuel prices history 2015*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2015/Fuel-Price-History.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2016a). *Basic fuel price 2016*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2016/Basic-Fuel-Price.pdf [Accessed 11 Sep. 2018].

Republic of South Africa's Department of Energy. (2016b). *South African fuel prices history 2016*. [online] Republic of South Africa's Department of Energy. Available at: http://www.energy.gov.za/files/esources/petroleum/December2016/Fuel-Price-History.pdf [Accessed 11 Sep. 2018].

South African Reserve Bank. (2017a). *Money and Banking*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

South African Reserve Bank. (2017b). *Capital Markets*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

South African Reserve Bank. (2017c). *Public Finance*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

South African Reserve Bank. (2017d). *Balance of Payments*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

South African Reserve Bank. (2017e). *National Accounts*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

South African Reserve Bank. (2017f). *Business Cycles*. [Microsoft Excel spreadsheet]. South African Reserve Bank. Available at: https://www.resbank.co.za/Publications/Detail-Item-View/Pages/Publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=7720 [Accessed 15 June 2017].

Statistics South Africa. (2013). In-depth analysis of the *General Household Survey data 2002-2013*. [online] Statistics South Africa. Available at: https://www.gov.za/sites/www.gov.za/files/generalhouseholdsurvey-ict-03-18-052013_reduced.pdf [Accessed 2 June 2017].

Statistics South Africa. (2014). *General Household Survey 2014*. [online] Statistics South Africa. Available at: http://www.statssa.gov.za/publications/P0318/P03182014.pdf [Accessed 2 June 2017].

Statistics South Africa. (2015). *General Household Survey 2015*. [online] Statistics South Africa. Available at: https://www.statssa.gov.za/publications/P0318/P03182015.pdf [Accessed 2 June 2017].

Statistics South Africa. (2016). *General Household Survey 2016*. [online] Statistics South Africa. Available at: https://www.statssa.gov.za/publications/P0318/P03182016.pdf [Accessed 2 June 2017].

Statistics South Africa. (2018). *CPI history*. [online] Statistics South Africa. Available at: http://www.statssa.gov.za/publications/P0141/CPIHistory.pdf? [Accessed 17 July 2018].

Transparency International. (1995). *1995 TI Corruption Index*. [online] Transparency International. Available at: https://www.transparency.org/files/content/tool/1995_CPI_EN.pdf [Accessed 10 June 2017].

Transparency International. (1996). *TI Corruption Perception Index 1996.* [online] Transparency International. Available at: https://www.transparency.org/files/content/tool/1996_CPI_EN.pdf [Accessed 10 June 2017].

Transparency International. (1997). *1997 Corruption Perception Index.* [online] Transparency International. Available at: https://www.transparency.org/files/content/tool/1997_CPI_EN.pdf [Accessed 10 June 2017].

Transparency International. (1998).*1998 Corruption Perceptions Index.* [online] Transparency International. Available at: https://www.transparency.org/news/pressrelease/1998_corruption_perceptions_index [Accessed 10 June 2017].

Transparency International. (1999). *Corruption Perceptions Index 1999.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_1999/0 [Accessed 10 June 2017].

Transparency International. (2000). *Corruption Perceptions Index 2000.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2000/0 [Accessed 10 June 2017].

Transparency International. (2001). *Corruption Perceptions Index 2001.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2001/0 [Accessed 10 June 2017].

Transparency International. (2002). *Corruption Perceptions Index 2002.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2002/0 [Accessed 10 June 2017].

Transparency International. (2003). *Corruption Perceptions Index 2003.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2003/0 [Accessed 10 June 2017].

Transparency International. (2004). *Corruption Perceptions Index 2004.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2004/0 [Accessed 10 June 2017].

Transparency International. (2005). *Corruption Perceptions Index 2005.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2005/0 [Accessed 10 June 2017].

Transparency International. (2006). *Corruption Perceptions Index 2006.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2006/0 [Accessed 10 June 2017].

Transparency International. (2007). *Corruption Perceptions Index 2007.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2007/0 [Accessed 10 June 2017].

Transparency International. (2008). *Corruption Perceptions Index 2008.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2008/0 [Accessed 10 June 2017].

Transparency International. (2009). *Corruption Perceptions Index 2009.* [online] Transparency International. Available at: https://www.transparency.org/research/cpi/cpi_2009/0 [Accessed 10 June 2017].

Transparency International. (2010). *Corruption Perceptions Index 2010.* [online] Transparency International. Available at: https://www.transparency.org/cpi2010/results [Accessed 10 June 2017].

Transparency International. (2011). *Corruption Perceptions Index 2011.* [online] Transparency International. Available at: https://www.transparency.org/cpi2011/results [Accessed 10 June 2017].

Transparency International. (2012). *Corruption Perceptions Index 2012.* [online] Transparency International. Available at: https://www.transparency.org/cpi2012/results [Accessed 10 June 2017].

Transparency International. (2013). *Corruption Perceptions Index 2013.* [online] Transparency International. Available at: https://www.transparency.org/cpi2013/results [Accessed 10 June 2017].

Transparency International. (2014). *Corruption Perceptions Index 2014.* [online] Transparency International. Available at: https://www.transparency.org/cpi2014/results [Accessed 10 June 2017].

Transparency International. (2015). *Corruption Perceptions Index 2015.* [online] Transparency International. Available at: https://www.transparency.org/cpi2015#results-table [Accessed 10 June 2017].

Transparency International. (2016). *Corruption Perceptions Index 2016.* [online] Transparency International. Available at: https://www.transparency.org/news/feature/corruption_perceptions_index_2016 [Accessed 10 June 2017].

United Nations Development Program. (2017). *Human Development Reports: Human development Index*. [online] United Nations Development Program. Available at: http://hdr.undp.org/en/indicators/137506 [Accessed 12 June 2017].

World Bank. (2017). *World Development Indicators: South Africa*. [Microsoft Excel spreadsheet]. The World Bank. Available at: https://data.worldbank.org/country/south-africa?view=chart [Accessed 1 June 2017].

World Bank. (2018). *World Development Indicators: South Africa*. [Microsoft Excel spreadsheet]. The World Bank. Available at: https://data.worldbank.org/country/south-africa?view=chart [Accessed 2 July 2018].

# Appendix A: Data capture protocol

A data capture protocol was written up to ensure that the social unrest data were consistently extracted from the ACLED database. This protocol had four main components. The first contained definitions for the categories that political violence events were sorted into. The second component listed 30 possible reasons for protests. This was followed with a list of assumptions that would be used if there was any uncertainty. The data capture protocol is completed by describing the reasonability checks that were performed after all the data were extracted. Each one of these components is described below.

## A.1 Political violence categories

Thirteen categories were used to extract all the social unrest events from the ACLED database. The categories include full protests, political assassinations, terrorist attacks, personal, racial and tribal disputes, Gang violence, robbery and trespassing, murder, prison and detention centre events, ambiguous entries, double entries, and finally xenophobic attacks. The definitions of each of the subcategories, based on what was described in the literature, are given below:

1. Protest

It is defined as a group (more than one person) of people who gather in public to show support or disapproval for a specified reason. Usually done by holding up posters while singing and dancing, otherwise known as toyi-toyiing in South Africa.

2. Political assassinations or clashes

These are attacks where political figures are targeted for political reasons (i.e. politically motivated murders or assassinations). It is also where members of political parties have a tit for tat feud where an attack from the one triggers a revenge attack (ie members from one political party throw eggs at the house of one of the rival political party's members house which then causes the rival political party to react). These events aren't protests as the reason behind these events are politically driven and not in protest of anything.

3. Lone wolf / terrorist attacks

These are events that occur in isolation with an individual or extremely small group that makes use of extreme measures, like planting a bomb to show displeasure. The motives behind these events are not always clear but are usually driven by ideological reasons.

4. Personal, racial and tribal disputes

The definition for this category is quite wide. It includes any non-political group or individual that orchestrates an attack on another non-political group or individual; the main aim of these types of events are that they intend to do harm to but intend to do so with less force than would be used in a terrorist attack. These includes sexist attacks, homophobic attacks, personal rivalries, communal conflict, ethnic violence and tribal conflicts.

5. Gang violence and mob/jungle justice

These are events where an individual or community takes the law into their own hands and convict an individual(s) without a trial, usually to death, for crimes that they are thought of having committed; in some cases, innocent people are incorrectly "convicted". The methods used as punishment are grotesque ranging from beating, stoning, necklacing, lynching and/or setting the person alight. There are parallels between this category and xenophobia as communities sometimes also make use of these methods in xenophobic attacks. In broad terms, these attacks may be viewed as a community protesting against crime. However, the motives behind these events are very different; protests are used to show displeasure with something whereas mob justice is used to entice fear in a community. This fundamental difference may create confounding in the statistical analysis and for this reason it was decided to exclude these events from the definition of a full protest.

6. Robbery and trespassing

These are where political figures, their homes, their cars and/or the properties of political parties are either robbed or individuals are found trespassing on these properties.

7. Murder

These are murders of politically figures and top union officials. Unlike political assassinations these are a result personal disputes and aren't politically motivated.

8. Prison and detention centre events

These are any riots or clashes that occur in a prison or detention centre. The rationale behind this is that these unrest events are usually fuelled by internal factors and may skew the results.

9. Ambiguous entries

These are events where there are either contradicting accounts of what occurred, or where there is insufficient information and cannot be categorised in one of the other categories. It also includes police brutality, calls to protest and any protest where there was only one person protesting; as the definition of a protest excludes one-man protests.

10. Double entries

Events in the same city, on the same day and for the same reason, while having identical accounts of what occurred.

11. Xenophobic attacks

Any random attack on a foreigner that occurred just because the individual(s) was not born in South Africa. The root cause of xenophobia is the fear of strangers or foreigners this is also fundamentally different to the cause of protests. Therefore, like gang violence and mob justice these events should be analysed separately to increase the reliability of any statistical analysis performed on the data. It is important to note that protests against xenophobia are classified as protests and not as xenophobic attacks.

## A.2 Reason for Protest

Thirty different categories for the reason for protest where created whereby social unrest events could be classified. The first 29 categories were based on the categories that were used by the ISS in their classification of the reasons for protests (ISS Crime Hub, 2017). An additional category, Racism and discrimination, was added based on South Africa unique history. The thirty categories are as follows:

1. Business practice
2. By law enforcement
3. Corruption
4. Crime/Anti-crime
5. Democracy
6. Education
7. Elections
8. Electricity
9. Environmental
10. Foreigners/ xenophobia
11. Healthcare
12. Housing
13. International causes
14. Jobs
15. Labour strike
16. Land
17. Mismanagement
18. Mob justice
19. Municipal service
20. National cause
21. Political party politics
22. Roads
23. Sanitation
24. Transport
25. Unspecified
26. Various
27. Water
28. Vigilantism
29. Other
30. Racism and discrimination

## A.3 List of Assumptions

A list of assumptions pertaining to arrests, injuries, fatalities, violent protests, legality of the protest, property damage, the size of protest, police intervention and university events are described below.

**Arrests:** If someone was arrested, then assume that the police intervened and that there was some sort of intervention weapon used. If it says nothing about arrests assume that no one was arrested.

**Injuries:** If it doesn’t say anything about injuries assume that no one was injured.

**Fatalities:** If it doesn’t mention fatalities assume that there were none.

**Violent protest:** Unless stated otherwise assume that a protest is violent if there were any fatalities, police had to intervene (i.e. use rubber bullets or tear gas), violent crimes were committed (including arson, burning tyres, throwing stones and/or riot) or if there was property damage. If nothing in this regard is mentioned, then assume that it was peaceful.

**Legal or illegal protest:** Unless stated otherwise assume it was an illegal protest if there was police intervention or if violence was used (by law a protest becomes illegal the moment that violence is used). Otherwise, assume that the protest was legal.

**Property damage: P**roperty damage includes anything other than life insurance and healthcare insurance products that can currently be insured in the South African market. In general terms this ranges from infrastructure, property, vehicles, equipment, containers, loss of production to stock. However, there is one exception to this. Burning tyres in isolation is not considered as “property damage” although tyres are technically insurable in South Africa; the tyres on a car are insured as part of the car. This is because we assume that majority of the tyres that are burnt have been thrown out due to wear and tear before being used in protests; in this case, these tyres are no longer insurable. This assumption is also logical as it is more likely that protesters will burn the entire car in the heat of the moment than to take the time to find bricks to balance the car and then remove the tyres from the nearby cars before being able to torch them.

**Police intervention:** Assume police intervention if someone was arrested. If it says nothing, then assume that there was no intervention.

**University events:** All events that relate to education that occurred at any tertiary institution (universities, training college, Technikons) in South Africa was recorded as a university event, unless a different reason was provided.

## A.4 Training session, data capture and consolidation process

During the training session, the assumptions and categories were explained to each student where after the students received the data set with their assigned data entries. Several data entries were also done with the students in this session. This ensured that they knew what was expected from them and gave them the opportunity to ask questions. Thereafter each student was given their respective data set, with each of their assigned data entries. As a result of the protests at the University of Pretoria in 2016, the access to campus was severely limited making it impossible to get all the students to work in the same venue. As a result, a different plan had to be made to ensure consistency. A WhatsApp group was created for this specific reason. It allowed the students to ask questions when they were unsure, and it allowed them to compare the answers to their portion of the data thus ensuring that they were interpreting the data in the same way as the other students.

The students were given a week to complete their portion of the data and they were also required to email the data set at the end of each day so that their progress could be tracked, and their data could be checked so that any misunderstanding in the assumptions could be identified and rectified. After all the students completed their assigned portion, the completed data were combined to create a complete data set. An initial set of consistencies checks were performed on all nine the students' data and appropriate adjustments were made where necessary. I then read through each of the data entries description and consistently categorised the type of event so that the full protests could be analysed individually. Thereafter another series of reasonability and consistency checks were performed on the data.

## A.5 Reasonability checks

To ensure that there is consistency between the nine undergraduate students' data entries a series of checks were performed. This process focused on the following variables:

- political violence events,
- reason for protest,
- property damage,
- the number of fatalities and injuries and
- the completeness of entries

Once spot checks were completed, the areas of concern were identified, and attention were given to these areas. The main areas of concern were related to sorting the political violence events, categorising the reason for the protest and the definition of property damage. To rectify this, all the descriptions were reread to consistently distinguish between the various categories of political unrest. This was done in chronological order as it was seen that some of the reasons for protests were mentioned prior to the student's data assignments, this resulted in the incorrect categorisation. All the property damage was sorted to ensure that no incidents of burning tyres was recorded as property damage nor as arson. The other consistency checks ensured the following logical analogies, according to the list of assumptions, were made:

- All protest with property damage were recorded as being violent.
- All the protests that were recorded as violent had at least one crime recorded.
- All the protests marked as having police intervention have an intervention method recorded.
- All protests where arrests were made are recorded as having police interventions.
- All protests where there where fatalities, injuries, violence or police intervention were recorded to ensure that the legality of the protest was categorised correctly.
- Special attention was also given to whether or not arrests were made at peaceful protests.

# Appendix B: Bottom-up approach that could be used to model and price political violence

This approach combines three of the modelling methods that have been discussed in modelling techniques, namely ABM, dynamic Bayesian networks and game theory. Firstly, ABM is used to model the behaviour of different agents in different scenarios. This model can be used to simulate different environments, so the agents can be modelled into a specific city or country of interest. This allows the model to give a better description of the expected value of damage that may result, as actual values can be assigned to each building and each agent.

Hypothetically it is also possible to model the entire world by combining smaller models that can each represent a specific country or region. However, this will require huge amounts of computing power to ensure that it is done accurately. Dynamic Bayesian networks can then be used to evaluate the simulated data of each agent's behaviour to determine which agents are behaving suspiciously. The next step is to determine whether the model did correctly determine which agents acted suspiciously, which can be used to estimate the damage of each situation. A series of ABM and Dynamic Bayesian networks will be required so that the expected loss under different scenarios can be tested. Multiple simulations will also be required for each of the scenarios, so that a realistic value for the expected loss under each scenario can be determined. This may be done by either determining a distribution of the expected loss, or by determining the average loss that occurred under the different simulations. The exact definition of the expected loss will depend on the type of insurance that is being considered. For a life insurance company, the expected number of deaths will be used to determine the amount that is expected to be paid out after the attack. Similarly, a health insurance company will base its cost on the number of individuals that have been injured, whereas a general insurer will base its liability on the value of the property that was damaged.

After the expected losses for each is calculated, the various scenarios have to be combined. Game theory can be used to determine the likely pay-outs that may result from insuring the simulated environment from the simulated risk event. A brief overview of this approach is shown in Figure 25. This approach may also be useful for modelling and pricing lawlessness, corruption and migration patterns.

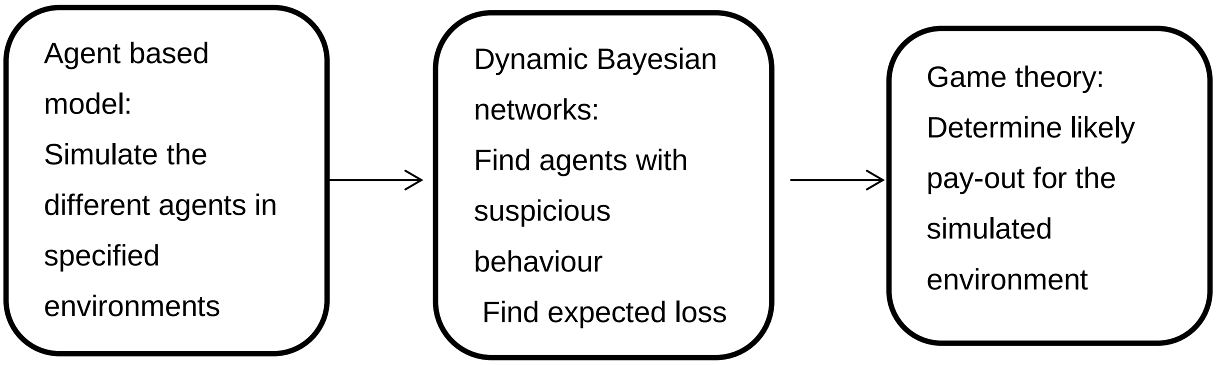



*Figure 25: A brief overview of the bottom-up approach described above.*

The first step is to create an ABM that mimics the real world. This is done by modelling the movements of each of the agents. This can be derived from each one's behaviour. The behaviour of the police agent is dependent on the region's strategies and protocols that are put into in place to protect the public. These strategies and protocols will influence or dictate the routes that should be patrolled. The terrorist's behaviour is influenced by their cause. This may incorporate likely targets, the timeframe involved, the method of attack deemed appropriate and the police agent's behaviour. Finally, the civilian's behaviour is centred around work and family life.

The movements of the police agent and civilian can be defined according to real world data. A hypothetical terrorist's strategy, which forms the base of the terrorist's movements, has to be tailor made based on the location. Once the movement for all three the agents have been defined, the agents can be combined in the same environment, where they can interact with one another.

A Dynamic Bayesian Network can then be used to compare the movements of each agent to identify any suspicious behaviour. This requires a definition for suspicious behaviour. In a train station, the civilian is likely to get onto and off the same train at around the same time as they go to and come from work. The police may have patrol routes that may be repetitive. Knowing the patterns of the civilian and the police, suspicious behaviour can then be defined as being any other form of train movements or schedule. As time passes the model can then categorise the agents as either a civilian, member of the police and likely terrorists.

Before the cost of the simulation can be determined, the cost of the intended attack and success rate has to be identified. The cost of the "intended damage" can be determined by the target that is specified in the terrorist's strategy. The success rate of the attack is dependent on the Dynamic Bayesian Network's ability to correctly predicted which of the are terrorists. This is done by comparing the agents who were identified as behaving suspiciously to those who were actually terrorists.

If the terrorist was not identified, then it can be assumed that a complete attack was launched and was successful, and that the intended damage resulted. If the terrorist was correctly identified, then there are two possible cases that can occur. The first being that the terrorist was captured before an attack was launched and no damage occurred. The second being that the terrorist attacked, but that it may not have been as big as initially planned. The damage caused could be less than what was initially intended. The probability of a successful attack and the simulated damage of each attack can then be used to determine the expected loss for a given scenario. A series of different terrorist and police strategies can then be modelled and the expected loss for each of these can be determined.

The overall cost to the insurance company can then be determined by setting up a game that accounts for the expected loss under each of the police and terrorist strategies that were modelled. Thus, giving an insurance company an idea of the maximum loss, which could result from a terrorist event. A summary of this process is depicted in Figure 26.

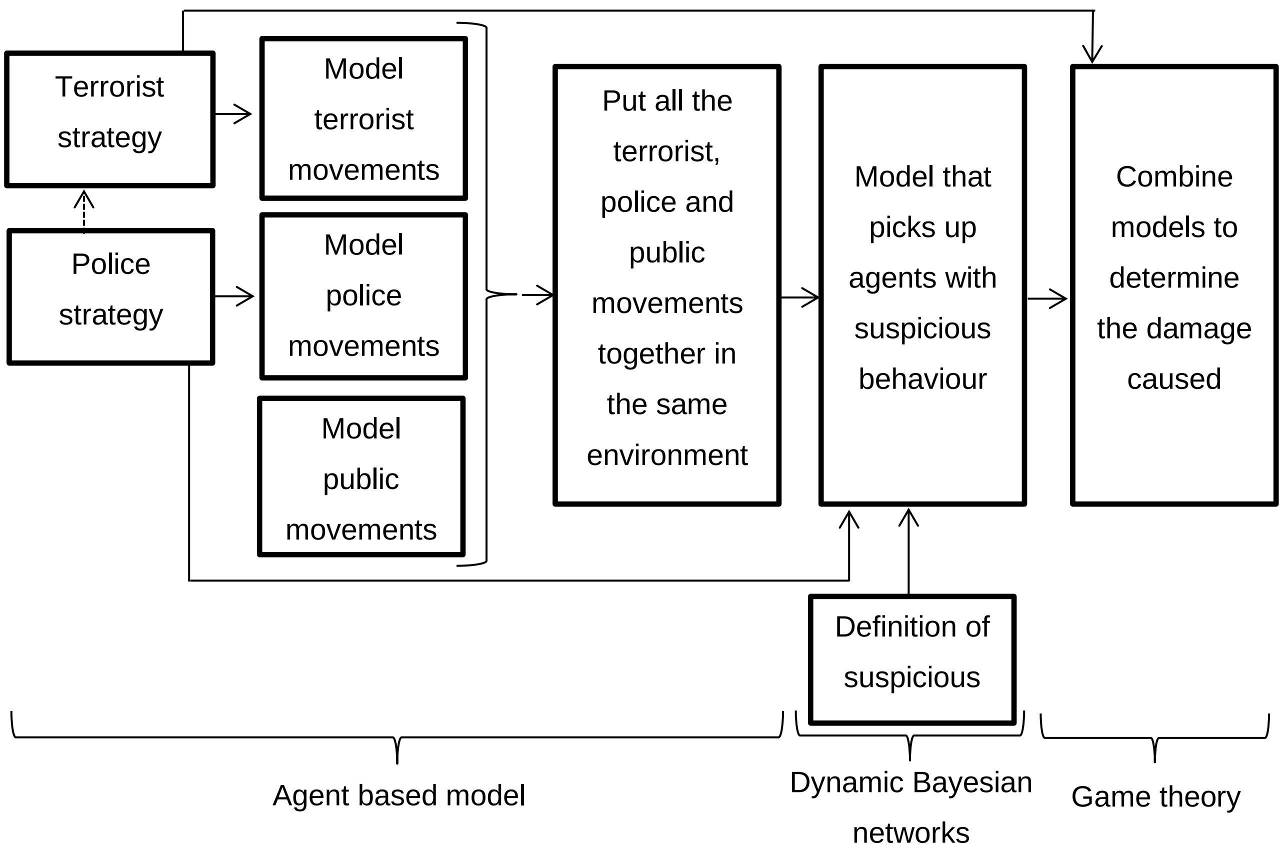



*Figure 26: A summary of the bottom-up approach used to model and price terrorism.*

The example that was just described makes use of terrorism rather than social unrest as it is conceptually the easiest of these two scenarios. Unlike terrorism which can be an individual act, social unrest, is a group action. Therefore, to create a realistic model, additional agents must be included. This makes the social unrest explanation more complicated than the explanation for terrorism.

# Appendix C: Basic analysis of the database

In the results section, the distribution of all protests, protests with violence and property damage as well as reasons for protests were briefly discussed. Additional analysis on these and other topics is described in this appendix.

The three biggest reasons for protests was Education (16.9%) narrowly followed by Labour strikes (16.2%) and then Municipal services (12.9%), as described in Section 5.1. The three biggest reasons for protests per year is shown in Table 34. Labour strikes appeared 19 out of 20 times, showing that this is a continuous problem in South Africa. The reason which appeared second most often was international causes (12 times) followed by municipal services (11 times), crime and anti-crime (10 times) and only then it was education (nine times). This indicates that education protests and municipal services protests are usually clustered together, unlike Labour which is a constant reason for protests.

Violence was recorded in 46.9% of all protests. The highest recorded rate of violence occurred in 2005, where 54.35% of protests were violent. While the lowest rate of violence, 19.77%, occurred in 2003. The annual proportion of violent protests is displayed in Figure 27.

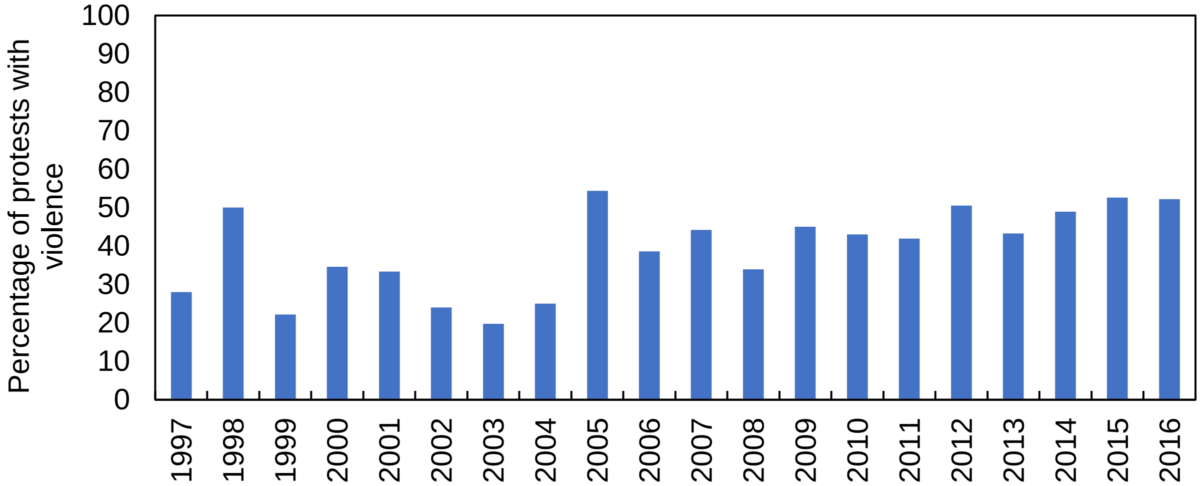



*Figure 27: The percentage of annual South African protests where violence was recorded.*

Property damage was recorded in only 14.4% of the protests. The highest recorded value was 21.48%, recorded in 2015. While the lowest incidence of property damage, 3.49%, occurred in 2003. The graph showing the yearly portion of protests with property damage is depicted in Figure 28.

*Table 34: The three biggest reasons of protests in South Africa for each of the 20 years from 1997-2016.*

| | **Reason for Protest** | | |
|---|---|---|---|
| | **Biggest Reason** | **Second Biggest** | **Third Biggest** |
| 1997 | Other | Crime/Anti-Crime | International Causes<br>Unspecified |
| 1998 | Labour Strike | International Causes | Education |
| 1999 | Labour Strike | International Causes | Unspecified<br>Other |
| 2000 | International Causes | Labour Strike<br>Crime/Anti-Crime | Other |
| 2001 | International Causes | Labour Strike | Crime/Anti-Crime |
| 2002 | International Causes | Labour Strike | National Cause |
| 2003 | International Causes | Labour Strike | Crime/Anti-Crime |
| 2004 | International Causes | Labour Strike | Crime/Anti-Crime<br>Education |
| 2005 | Municipal Service | Housing | Labour Strike |
| 2006 | International Causes | Labour Strike | Municipal Service |
| 2007 | Municipal Service | Crime/Anti-Crime | Labour Strike |
| 2008 | International Causes<br>Crime/Anti-Crime | Education | Labour Strike |
| 2009 | Municipal Service | Labour Strike | International Causes |
| 2010 | Municipal Service | Education | Labour Strike |
| 2011 | Education<br>Municipal Service | Labour Strike<br>Crime/Anti-Crime | International Causes<br>Housing |
| 2012 | Labour Strike | Education<br>Municipal Service | Crime/Anti-Crime |
| 2013 | Labour Strike | Municipal Service | Crime/Anti-Crime |
| 2014 | Municipal Service | Labour Strike | Education |
| 2015 | Education | Labour Strike | Municipal Service |
| 2016 | Education | Labour Strike | Municipal Service |

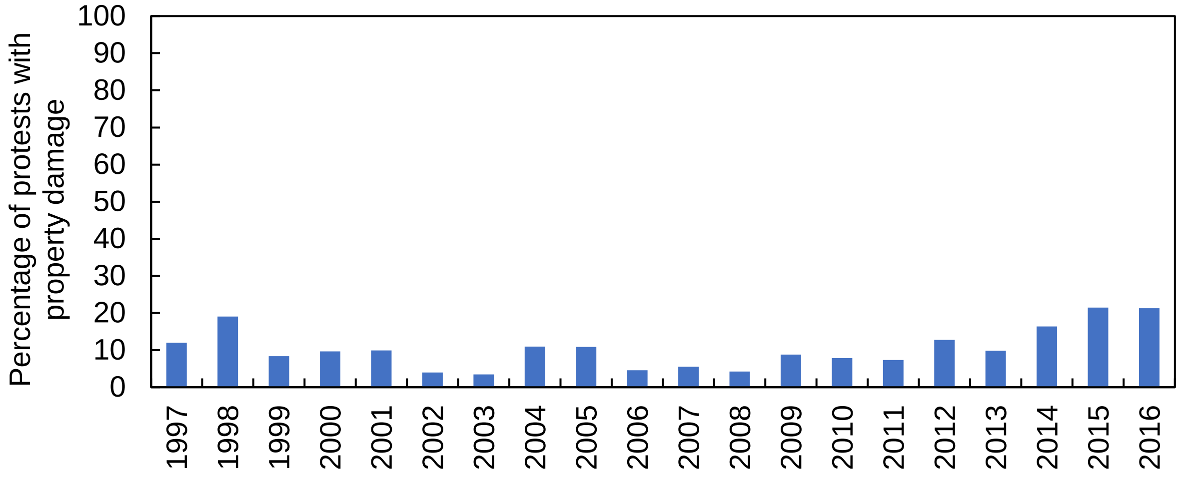



*Figure 28: The percentage of annual South African protests where property damage was recorded.*

A change in the annual distribution of property damage was observed when all the peaceful protests were removed. Here property damage occurred in 30.7% of all violent protests. This ranged between 11.76% and 43.75%, recorded in 2006 and in 2004, respectively. The graph showing the yearly portion of violent protests with property damage is depicted in Figure 29.

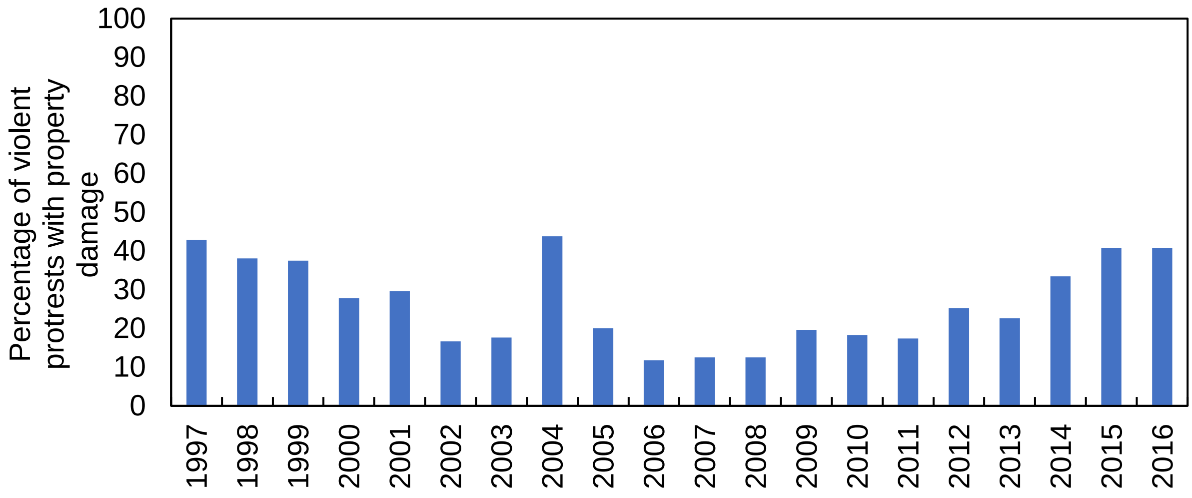



*Figure 29: The yearly percentage of violent protests in South Africa where property damage was recorded.*

Over 20 years there were 128 (1.8% of all protests) fatal protests with a total of 221 recorded fatalities. The highest number of fatalities recorded during a single protest was 34, which occurred in Marikana during a mine labour strike in 2012. Figure 30 shows the number of fatal protests and number of recorded fatalities per year from 1997 to 2016. This graph shows that there was a change in the incidence of fatal protests and fatalities. This change is apparent by increases in the occurrence of these events and is seen in the last six years of the data. The year with the most fatalities was 2012; which coincides with the fatal Marikana's protests. A similar increase was also observed in the number of fatal protests in that year.

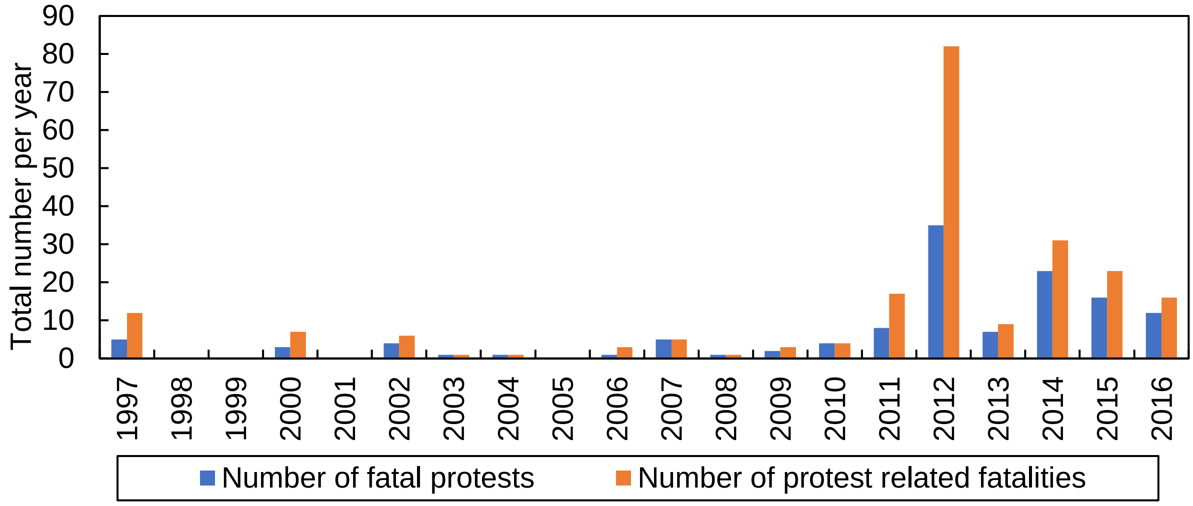



*Figure 30: The number of fatal protests and the resultant number of fatalities that was recorded in South Africa each year.*

The distribution of the number of fatalities recorded at a fatal protest is exponentially decreasing and can be seen in Figure 31. Of the fatal protests 68% resulted in only one fatality being recorded, 21.8% resulted in two fatalities, while a further 10.1% resulted in either three, four or five fatalities. Only 0.8% of fatal protests resulted in more than five fatalities being recorded.

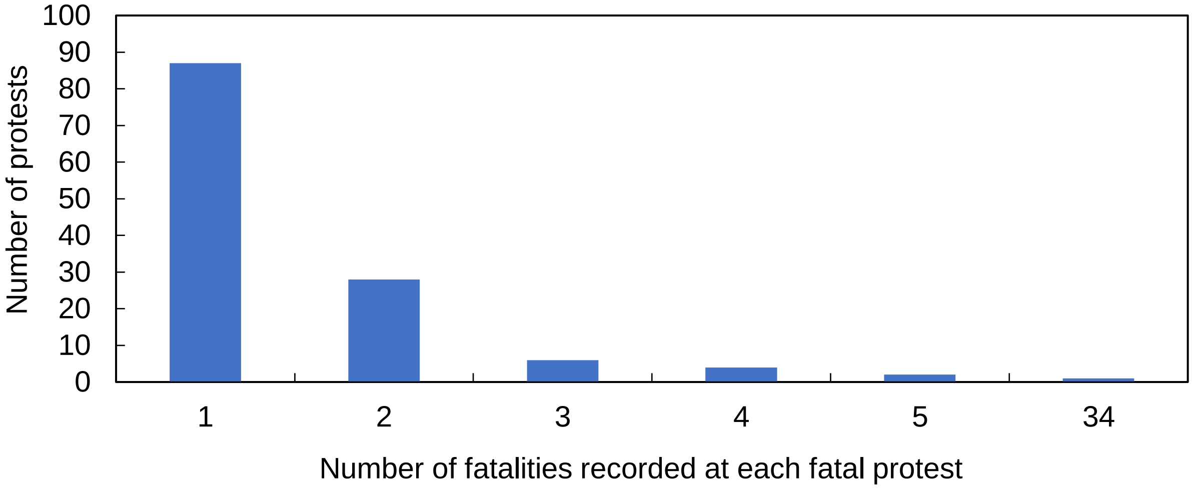



*Figure 31: The number of fatalities that occurred in each of the 128 fatal protests in South Africa.*

# Appendix D: Spearman’s rank correlation for different subcategories of social unrest

The Spearman’s rank correlation coefficients for all protests, violent protests, protests with property damage, education, labour and municipal related protests are displayed in this appendix. It starts by describing the relationships that are present within the monthly data, followed by the quarterly and annual data. Due to the number of annual variables, this section is divided into economic, demographic, country-specific, socio-economic, international and other factors have been grouped together. Unfortunately, daily economic and socioeconomic data in South Africa is scarce and no correlation analyses were performed using the daily protest data.

## D.1 Spearman’s rank correlation: Monthly variables

The Spearman’s rank correlation coefficients for the monthly social unrest subcategories (collective term referring to all protests, violent protests, protests with property damage, education-, labour- and municipal-related protests) are depicted in Table 35. Irrespective of the social unrest category considered, none of the economic and socio-economic variables displayed any significant changes in the relationships observed.

*Table 35: Monthly Spearman's rank correlation coefficients between the subcategories of social unrest and the economic and socioeconomic variables for South Africa.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| FPI | 0.69391* | 0.67219* | 0.52011* | 0.5191* | 0.45892* | 0.63946* |
| CPI Headline Index | 0.85077* | 0.81604* | 0.70144* | 0.6683* | 0.62466* | 0.78447* |
| Government Revenue (R Mil) | 0.7569* | 0.72176* | 0.61024* | 0.55686* | 0.56417* | 0.69962* |
| Government Expenditure (R Mil) | 0.84545* | 0.81925* | 0.69914* | 0.6708* | 0.61032* | 0.77172* |
| Government Surplus (R Mil) | -0.36224* | -0.36752* | -0.28725* | -0.40562* | -0.23323* | -0.3191* |
| Basic Diesel Price | 0.78225* | 0.73836* | 0.59067* | 0.54353* | 0.54594* | 0.7053* |
| Exchange Rate (Rands/Us$) | 0.67321* | 0.61703* | 0.55171* | 0.51896* | 0.47775* | 0.54407* |
| Wholesale Price for 93 Petrol (Inland) | 0.83794* | 0.80093* | 0.68119* | 0.6363* | 0.61474* | 0.77087* |
| Wholesale Price for Paraffin (Inland) | 0.79682* | 0.75013* | 0.61329* | 0.56738* | 0.5723* | 0.72849* |

* significant at a level of 0.01   ** significant at a level of 0.05   *** significant at a level of 0.1

## D.2 Spearman's rank correlation: Quarterly variables

The Spearman's rank correlation coefficients for the quarterly social unrest subcategories and economic variables are depicted in Table 36. Apart from unemployment, there were no significant changes in the relationships observed, irrespective of the subcategory of social unrest considered. It was found that the official unemployment rate was dependent on the social unrest subcategory considered.

*Table 36: Quarterly Spearman's rank correlation coefficients of the social unrest subcategories and economic and socioeconomic variables for South Africa.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Quarterly FPI | 0.70955* | 0.7263* | 0.58798* | 0.58749* | 0.5565* | 0.71723* |
| Quarterly CPI Headline Index | 0.88832* | 0.87132* | 0.75366* | 0.76179* | 0.73791* | 0.85869* |
| Quarterly Government Revenue | 0.87478* | 0.85672* | 0.73822* | 0.74266* | 0.72185* | 0.85286* |
| Quarterly Government Expenditure | 0.89296* | 0.88081* | 0.75311* | 0.77447* | 0.74937* | 0.86197* |
| Quarterly Government Surplus | -0.69222* | -0.68665* | -0.66184* | -0.6555* | -0.6205* | -0.60349* |
| Quarterly Income Growth | -0.40465* | -0.42958* | -0.42938* | -0.36694* | -0.32085* | -0.40287* |
| Official Unemployment (%) | 0.27536** | 0.2031*** | 0.28029** | 0.35888* | 0.1959*** | 0.09582 |
| Basic Diesel Price | 0.81934* | 0.78801* | 0.62938* | 0.62775* | 0.66933* | 0.80592* |
| Exchange Rate (Rands/Us$) | 0.69331* | 0.5953* | 0.53783* | 0.58331* | 0.56659* | 0.56462* |
| Wholesale Price for 93 Petrol (Inland) | 0.86695* | 0.85158* | 0.72724* | 0.71238* | 0.72662* | 0.84913* |
| Wholesale Price for Paraffin (Inland) | 0.83104* | 0.81* | 0.67276* | 0.66179* | 0.68733* | 0.83251* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

## D.3 Spearman's rank correlation: Annual variables

The Spearman's rank correlation coefficients for the annual data are divided into six different categories. Economic factors are described first. This is followed by the demographic, country-specific, socio-economic, international interactions and concluded with the other factors.

### D.3.1 Economic factors

The Spearman's rank correlation coefficients for the economic factors and social unrest subcategories are depicted in Table 37. Irrespective of the social unrest category considered, CPI did not show significant changes in the relationships present. The relationships observed for the GDP growth rate and GDP per capita growth rate, however, were dependent on the social unrest subcategory that was considered.

*Table 37: Annual Spearman's rank correlation coefficients between the economic factors and the social unrest subcategories in South Africa.*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| CPI | 20 | 0.92481* | 0.90218* | 0.83478* | 0.8509* | 0.83315* | 0.92621* |
| GDP Growth Rate | 20 | -0.38496*** | -0.44921** | -0.52842** | -0.52184** | -0.48136** | -0.25828 |
| GDP Per Capita Growth Rate | 20 | -0.36692 | -0.4304*** | -0.51712** | -0.50301** | -0.4565** | -0.23268 |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

### D.3.2 Demographic factors

The Spearman's rank correlation coefficients between the demographic factors and the social unrest subcategories are depicted in Table 38. Irrespective of the social unrest category considered, the population density, rural population, population size, population growth and the urban population did not show significant changes in the relationships present. The relationships observed for rural population growth and urban population growth, however, was dependent on the social unrest subcategory that was considered.

*Table 38: Annual Spearman's rank correlation coefficients between the demographic factors and the social unrest subcategories in South Africa.*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| Population Density | 20 | 0.9218* | 0.89993* | 0.84155* | 0.85618* | 0.83164* | 0.93072* |
| Rural Population (% Of Total Population) | 20 | -0.9218* | -0.89993* | -0.84155* | -0.85618* | -0.83164* | -0.93072* |
| Rural Population Growth | 20 | -0.41955*** | -0.39052*** | -0.30109 | -0.30271 | -0.26968 | -0.49247** |
| Population Size | 20 | 0.9218* | 0.89993* | 0.84155* | 0.85618* | 0.83164* | 0.93072* |
| Population Growth (Annual %) | 20 | -0.28872 | -0.23853 | -0.1385 | -0.15136 | -0.13785 | -0.32907 |
| Urban Population (% Of Total Population) | 20 | 0.9218* | 0.89993* | 0.84155* | 0.85618* | 0.83164* | 0.93072* |
| Urban Population Growth | 20 | -0.48271** | -0.4763** | -0.39368*** | -0.38253*** | -0.32618 | -0.56852* |

* significant at a level of 0.01   ** significant at a level of 0.05   *** significant at a level of 0.1

## D.3.3 Country-specific factors

The Spearman's rank correlation coefficients between country-specific factors and the social unrest subcategories are depicted in Table 39. Irrespective of the social unrest category considered, the government surplus, expenditure, revenue and subsidies did not show significant changes in the relationships present. The relationships observed for corruption perception index, however, was dependent on the social unrest subcategory that was considered.

*Table 39: Annual Spearman's rank correlation coefficients between the country-specific factors and the social unrest subcategories in South Africa*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| Government surplus (R mil) | 20 | -0.73985* | -0.76298* | -0.75725* | -0.73042* | -0.69228* | -0.86973* |
| Corruption Perception Index | 20 | -0.66894* | -0.57537* | -0.54346** | -0.54821** | -0.59834* | -0.66465* |
| Total Government Expenditure (R Mil) | 20 | 0.9218* | 0.89993* | 0.84155* | 0.85618* | 0.83164* | 0.93072* |
| Total Government Revenue (R Mil) | 20 | 0.91729* | 0.89466* | 0.83478* | 0.85316* | 0.82411* | 0.92319* |
| Government Subsidies(% of Expense) | 20 | 0.72632* | 0.71558* | 0.66692* | 0.6875* | 0.65236* | 0.76883* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

## D.3.4 Socioeconomic factors

The Spearman's rank correlation coefficients for the socio-economic factors and the annual social unrest subcategories are depicted in Table 40. Regardless of the social unrest category considered, the age dependency ratio, access to electricity, FPI and access to improved sanitation facilities did not show significant changes in the relationships present. The relationships observed for income growth, the Human Development index and unemployment, however, was dependent on the social unrest subcategory that was considered. The relationships observed for unemployment was not only dependent on the subcategory of social unrest, it was also dependent on the type of unemployment rate (male, female, total or youth rates) which was analysed.

*Table 40: Annual Spearman's rank correlation coefficients between the socio-economic factors and the social unrest subcategories in South Africa.*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| Age Dependency Ratio | 20 | -0.9218* | -0.89993* | -0.84155* | -0.85618* | -0.83164* | -0.93072* |
| Access to Electricity(% Of Population) | 20 | 0.89282* | 0.85585* | 0.82605* | 0.82448* | 0.84702* | 0.93032* |
| FPI | 20 | 0.72632* | 0.73062* | 0.74671* | 0.73569* | 0.64934* | 0.79593* |
| Female Unemployment Rate | 20 | -0.18346 | -0.23476 | -0.26797 | -0.22816 | -0.07081 | -0.31551 |
| Male Unemployment Rate | 20 | 0.46617** | 0.39503*** | 0.32819 | 0.44352*** | 0.53785** | 0.21235 |
| Total Unemployment Rate | 20 | 0.12185 | 0.08656 | 0.04443 | 0.12655 | 0.28033 | -0.11299 |
| Total Youth Unemployment Rate | 20 | 0.28722 | 0.18134 | 0.1385 | 0.27636 | 0.49868** | 0.12425 |
| Female Youth Unemployment Rate | 20 | 0.13835 | 0.04289 | 0.05495 | 0.17846 | 0.36008 | 0.04518 |
| Male Youth Unemployment Rate | 20 | 0.33083 | 0.25132 | 0.18291 | 0.31325 | 0.48663** | 0.13705 |
| Income Growth | 20 | -0.52275** | -0.50621** | -0.49473** | -0.43541*** | -0.42577*** | -0.59962* |
| Human Development Index | 19 | 0.50461** | 0.5762* | 0.64763* | 0.62989* | 0.52771** | 0.42725*** |
| Access to Improved Sanitation | 19 | 0.90877* | 0.88499* | 0.81687* | 0.83216* | 0.8211* | 0.93673* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

### D.3.5 International factors

The Spearman's rank correlation coefficients for the international factors and the social unrest subcategories are depicted in Table 41. Irrespective of the social unrest category considered, economic freedom index, access to the internet, access to cell phones, access to fixed telephones and the freedom of the press did not show significant changes in the relationships present. The relationships observed for government integrity, monetary freedom and the tax burden, however, were dependent on the social unrest subcategory that was considered.

*Table 41: Annual Spearman's rank correlation coefficients between the international factors and the social unrest subcategories in South Africa.*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| Economic Freedom Index | 20 | -0.66918* | -0.68738* | -0.73587* | -0.59593* | -0.59012* | -0.70034* |
| Government Integrity | 20 | -0.67647* | -0.64528* | -0.60665* | -0.54192** | -0.61013* | -0.77908* |
| Monetary Freedom | 20 | -0.62105* | -0.60948* | -0.74069* | -0.71386* | -0.5936* | -0.55045** |
| Tax Burden | 20 | 0.59337* | 0.54861** | 0.49943** | 0.49585** | 0.50321** | 0.72775* |
| Access to the Internet | 20 | 0.91429* | 0.87434* | 0.83779* | 0.87726* | 0.8339* | 0.91416* |
| Access to Cell Phones | 20 | 0.9203* | 0.90143* | 0.84306* | 0.85467* | 0.83465* | 0.93223* |
| Access to Fixed Telephone | 20 | -0.90977* | -0.88036* | -0.84005* | -0.8381* | -0.81507* | -0.94051* |
| Freedom of the Press | 20 | 0.78141* | 0.8119* | 0.8478* | 0.81289* | 0.7321* | 0.77349* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

### D.1.6 Other factors

The Spearman's rank correlation coefficients for the other factors and the social unrest subcategories are depicted in Table 42. Irrespective of the social unrest category considered, basic diesel price, exchange rate, wholesale petrol price, wholesale paraffin price and access to improved water sources did not show significant changes in the relationships present. The relationships observed for trade freedom, however, was dependent on the social unrest subcategory that was considered.

*Table 42: Annual Spearman's rank correlation coefficients between the other factors and the social unrest subcategories in South Africa.*

| | N= | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|---|
| Trade Freedom | 20 | 0.58304* | 0.61193* | 0.54574** | 0.43205*** | 0.44209*** | 0.72173* |
| Basic Diesel Price | 20 | 0.89474* | 0.8623* | 0.77832* | 0.80723* | 0.81733* | 0.9119* |
| Exchange Rate (Rands/Us$) | 20 | 0.76992* | 0.67494* | 0.53368* | 0.66491* | 0.7774* | 0.5994* |
| Wholesale 93 Petrol Price (Inland) | 20 | 0.90376* | 0.87058* | 0.80693* | 0.82229* | 0.81582* | 0.92244* |
| Wholesale Price for Paraffin (Inland) | 20 | 0.90226* | 0.89767* | 0.82198* | 0.82229* | 0.80075* | 0.93901* |
| Access to Improved Water Sources | 19 | 0.90877* | 0.88499* | 0.81687* | 0.83216* | 0.8211* | 0.93673* |

* significant at a level of 0.01 ** significant at a level of 0.05 *** significant at a level of 0.1

# Appendix E: An example of the SAS output for a linear and lagged regression model

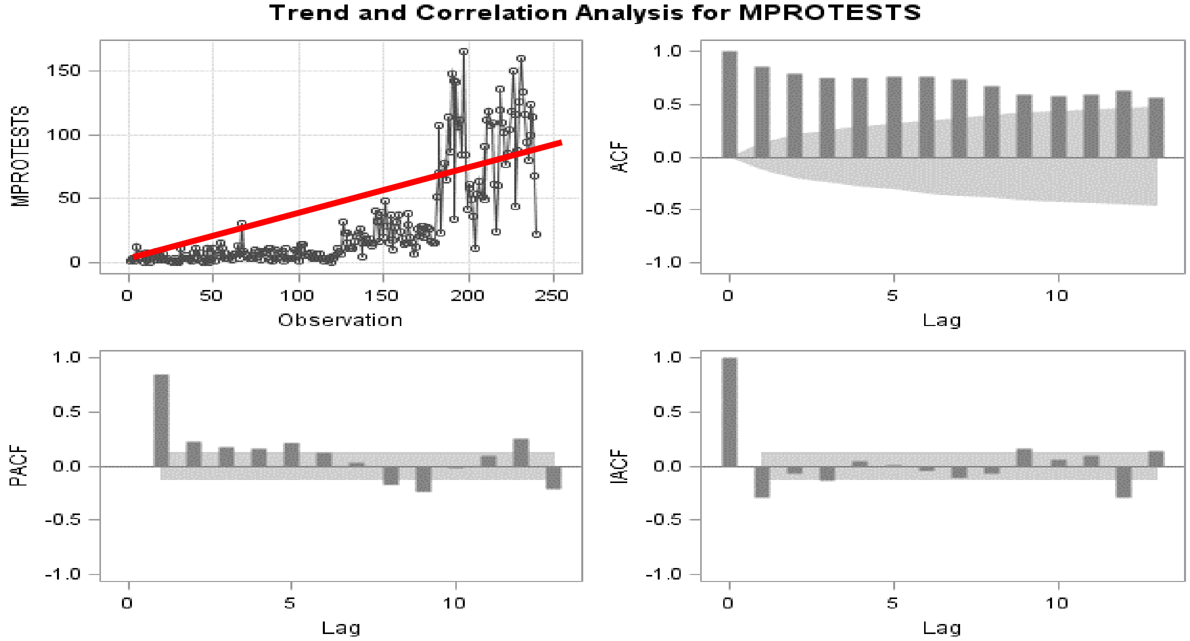
An example of the statistical analysis that is performed to create a linear regression as well as a lagged regression model is given below. These two models are performed in a similar manner, with lagged regression only requiring a few additional steps. For this reason, the linear regression model is described first, where after the additional process for lagged regression is shown. Portions of the raw SAS output is displayed in the black boxes. The red lines, blocks and circles have been added to highlight the details within the SAS output that are being discussed.

## E.1 Linear regression model

To prevent spurious regression arising the regression models will only make use of stationary variables. Therefore, the first step is to test for stationarity in all the variables by using the Augmented Dicky Fuller unit root test. This test always produces statistical analysis for three different cases. The case that is used is determined by the time plot. If the data moves around a zero mean, the first case is used. If the data moves around a single mean, that is not zero, the second case is used. If a different trend is observed within the observations, then the third case is used.

The Augmented Dicky Fuller unit root tests the null hypothesis $H_0$: $\rho=1$ vs the alternative $H_A$: $\rho<1$. If the p-value for Rho and Tau is smaller than 0.05, then the null hypothesis is rejected, and it is concluded that the series is either stationary (if the series is classified as case 1 or 2) or trend stationary (if the series falls into case 3). If the null hypothesis is not rejected, then the series is non-stationary. In this case the first difference is taken, and the test is repeated to see whether the unit root has been removed.

The time plot of the number of monthly protests (MPROTESTS) has an upward trend, as shown with the red line on the following page. Therefore, the third case of the Augmented Dicky Fuller test is used. The Pr < Rho and Pr < Tau values for this case are outlined with red blocks on the next page. The p-values for lag 0, 1, 2 and 3 are all smaller than 0.05, thus $H_0$ is easily rejected at a 5% significance level. Therefore, it is concluded that the number of monthly protests is trend stationary, with no unit roots.



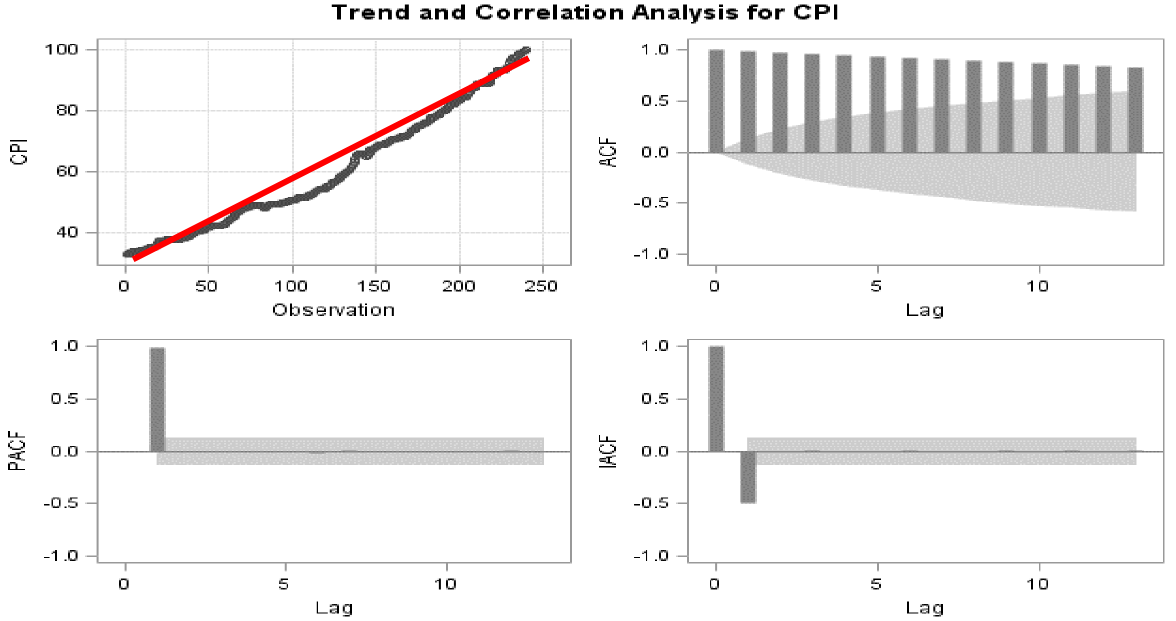

**Augmented Dickey-Fuller Unit Root Tests**

| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| **Zero Mean** | **0** | -22.5973 | 0.0006 | -3.43 | 0.0007 | | |
| | **1** | -13.3373 | 0.0105 | -2.55 | 0.0109 | | |
| | **2** | -8.3809 | 0.0441 | -1.96 | 0.0485 | | |
| | **3** | -4.9718 | 0.1245 | -1.43 | 0.1424 | | |
| | **4** | -2.3019 | 0.2969 | -0.87 | 0.3409 | | |
| | **5** | -1.1998 | 0.4393 | -0.53 | 0.4842 | | |
| **Single Mean** | **0** | -36.1142 | 0.0015 | -4.42 | 0.0004 | 9.76 | 0.0010 |
| | **1** | -22.7148 | 0.0053 | -3.37 | 0.0133 | 5.69 | 0.0197 |
| | **2** | -15.2008 | 0.0355 | -2.71 | 0.0748 | 3.69 | 0.1312 |
| | **3** | -9.8679 | 0.1355 | -2.15 | 0.2259 | 2.37 | 0.4661 |
| | **4** | -5.4328 | 0.3920 | -1.55 | 0.5091 | 1.33 | 0.7324 |
| | **5** | -3.6706 | 0.5744 | -1.24 | 0.6601 | 0.98 | 0.8208 |
| **Trend** | **0** | -80.4893 | 0.0006 | -6.80 | <.0001 | 23.14 | 0.0010 |
| | **1** | -60.1576 | 0.0006 | -5.35 | <.0001 | 14.31 | 0.0010 |
| | **2** | -44.9939 | 0.0006 | -4.44 | 0.0023 | 9.86 | 0.0010 |
| | **3** | -31.7213 | 0.0045 | -3.69 | 0.0251 | 6.80 | 0.0373 |
| | **4** | -19.5806 | 0.0700 | -2.96 | 0.1451 | 4.41 | 0.2959 |
| | **5** | -14.3680 | 0.1988 | -2.56 | 0.2994 | 3.32 | 0.5148 |

Acronyms in SAS output:

| | |
|---|---|
| PACF | Partial Autocorrelation Function |
| ACF | Autocorrelation Function |
| IACF | Inverse Autocorrelation Function |

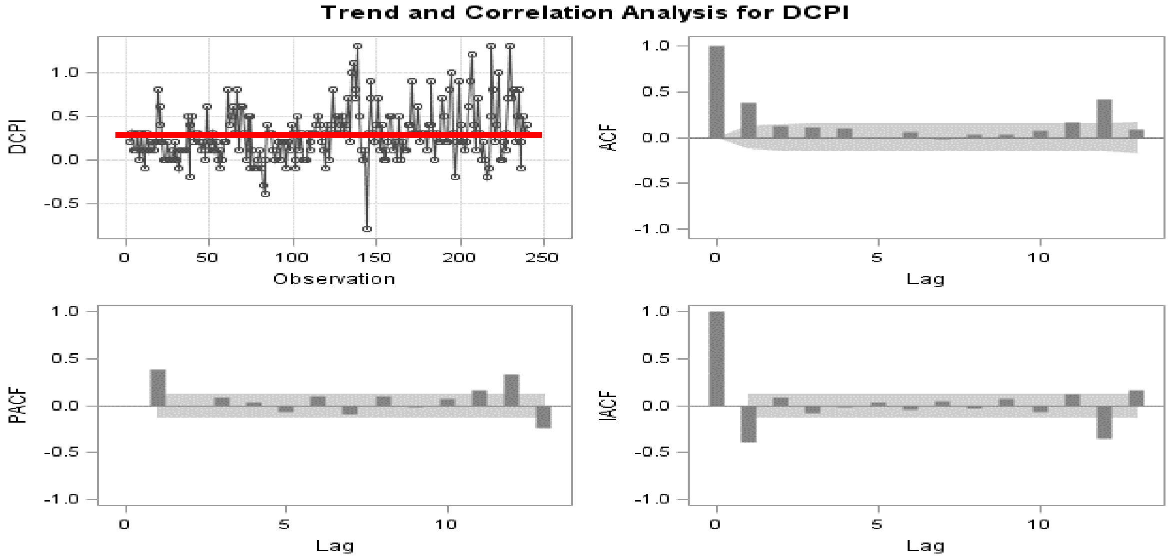
The time plot of CPI displayed below also has an upward trend. In this case $H_0$ cannot be rejected as the corresponding p-values aren't smaller than 0.05. It is therefore concluded that the series is non-stationary and has a unit root.




**Augmented Dickey-Fuller Unit Root Tests**

| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| **Zero Mean** | **0** | 1.1088 | 0.9278 | 15.45 | 0.9999 | | |
| | **1** | 1.1029 | 0.9270 | 7.72 | 0.9999 | | |
| | **2** | 1.0981 | 0.9263 | 7.21 | 0.9999 | | |
| | **3** | 1.0921 | 0.9255 | 6.17 | 0.9999 | | |
| | **4** | 1.0887 | 0.9250 | 5.70 | 0.9999 | | |
| | **5** | 1.0856 | 0.9246 | 5.94 | 0.9999 | | |
| **Single Mean** | **0** | 1.0225 | 0.9890 | 4.35 | 0.9999 | 119.07 | 0.0010 |
| | **1** | 0.9954 | 0.9886 | 2.89 | 0.9999 | 29.77 | 0.0010 |
| | **2** | 0.9996 | 0.9886 | 2.97 | 0.9999 | 25.97 | 0.0010 |
| | **3** | 1.0046 | 0.9887 | 2.77 | 0.9999 | 19.00 | 0.0010 |
| | **4** | 0.9885 | 0.9884 | 2.67 | 0.9999 | 16.22 | 0.0010 |
| | **5** | 1.0013 | 0.9887 | 2.96 | 0.9999 | 17.61 | 0.0010 |
| **Trend** | **0** | -0.4581 | 0.9933 | -0.31 | 0.9899 | 9.99 | 0.0010 |
| | **1** | -1.5672 | 0.9792 | -0.75 | 0.9673 | 4.96 | 0.1870 |
| | **2** | -1.4101 | 0.9821 | -0.70 | 0.9711 | 5.18 | 0.1417 |
| | **3** | -1.7751 | 0.9750 | -0.83 | 0.9604 | 4.74 | 0.2293 |
| | **4** | -1.7187 | 0.9762 | -0.80 | 0.9638 | 4.39 | 0.3000 |
| | **5** | -1.1800 | 0.9856 | -0.60 | 0.9776 | 5.03 | 0.1711 |

The presence of a unit root prompts taking the first difference of CPI (DCPI) and retesting the new series. This time plot moves around a single mean, prompting the use of the second case of the Augmented Dicky-Fuller unit root test. For this series $H_0$ is easily rejected at a 5% significance level as the corresponding p-values are smaller than 0.05. It is therefore concluded that DCPI is stationary with no unit roots.




**Augmented Dickey-Fuller Unit Root Tests**

| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| **Zero Mean** | **0** | -79.9712 | <.0001 | -6.90 | <.0001 | | |
| | **1** | -57.2201 | <.0001 | -5.31 | <.0001 | | |
| | **2** | -34.5926 | <.0001 | -3.96 | <.0001 | | |
| | **3** | -24.8335 | 0.0003 | -3.28 | 0.0011 | | |
| | **4** | -22.9260 | 0.0006 | -3.09 | 0.0022 | | |
| | **5** | -14.5788 | 0.0074 | -2.47 | 0.0135 | | |
| **Single Mean** | **0** | -148.585 | 0.0001 | -10.35 | <.0001 | 53.53 | 0.0010 |
| | **1** | -151.371 | 0.0001 | -8.64 | <.0001 | 37.36 | 0.0010 |
| | **2** | -120.491 | 0.0001 | -6.86 | <.0001 | 23.54 | 0.0010 |
| | **3** | -109.399 | 0.0001 | -6.05 | <.0001 | 18.28 | 0.0010 |
| | **4** | -141.032 | 0.0001 | -6.01 | <.0001 | 18.09 | 0.0010 |
| | **5** | -94.0990 | 0.0014 | -5.02 | <.0001 | 12.60 | 0.0010 |
| **Trend** | **0** | -161.069 | 0.0001 | -10.96 | <.0001 | 60.11 | 0.0010 |
| | **1** | -177.897 | 0.0001 | -9.35 | <.0001 | 43.72 | 0.0010 |
| | **2** | -155.843 | 0.0001 | -7.58 | <.0001 | 28.74 | 0.0010 |
| | **3** | -157.030 | 0.0001 | -6.77 | <.0001 | 22.92 | 0.0010 |
| | **4** | -263.547 | 0.0001 | -6.88 | <.0001 | 23.64 | 0.0010 |
| | **5** | -185.901 | 0.0001 | -5.88 | <.0001 | 17.26 | 0.0010 |

After stationarity of all variables have been tested, all the non-stationary variables are removed, and a regression model can be built. The regression model is built using forward stepwise selection. The model starts by testing all of the variables and then includes the most significant variable. Hereafter variables are retested, and an additional variable is included with each iteration. Simultaneously, the parameter estimates of each of the variables included in the model is tested to ensure that all the parameters are significantly different from zero. If this is not the case, then the variable is removed from the model. The process will continue until all the parameters that are significantly different from zero, at a significance level of 5%, are included in the model.

A summary of this process is described in the summary of stepwise selection. The regression model that best describes the number of monthly protests only incorporates government surplus (SUR), government revenue (REV). The forward stepwise selection process included the month and internet usage. These variables were however removed from the model as the corresponding Pr > F, shown with the red block, was larger than 0.05. This indicated that the parameters weren't statistically significant in explaining the number of monthly protests.

**Summary of Stepwise Selection**

| Step | Variable Entered | Variable Removed | Number Vars in | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
|---|---|---|---|---|---|---|---|---|
| **1** | Month | | 1 | 0.5575 | 0.5575 | 45.7664 | 298.60 | <.0001 |
| **2** | SUR | | 2 | 0.0241 | 0.5816 | 32.4507 | 13.62 | 0.0003 |
| **3** | REV | | 3 | 0.0470 | 0.6286 | 4.6349 | 29.74 | <.0001 |
| **4** | | Month | 2 | 0.0007 | 0.6279 | 3.1094 | 0.47 | 0.4922 |
| **5** | INTER1IND | | 3 | 0.0035 | 0.6314 | 2.9025 | 2.22 | 0.1378 |
| **6** | | INTER1IND | 2 | 0.0035 | 0.6279 | 3.1094 | 2.22 | 0.1378 |

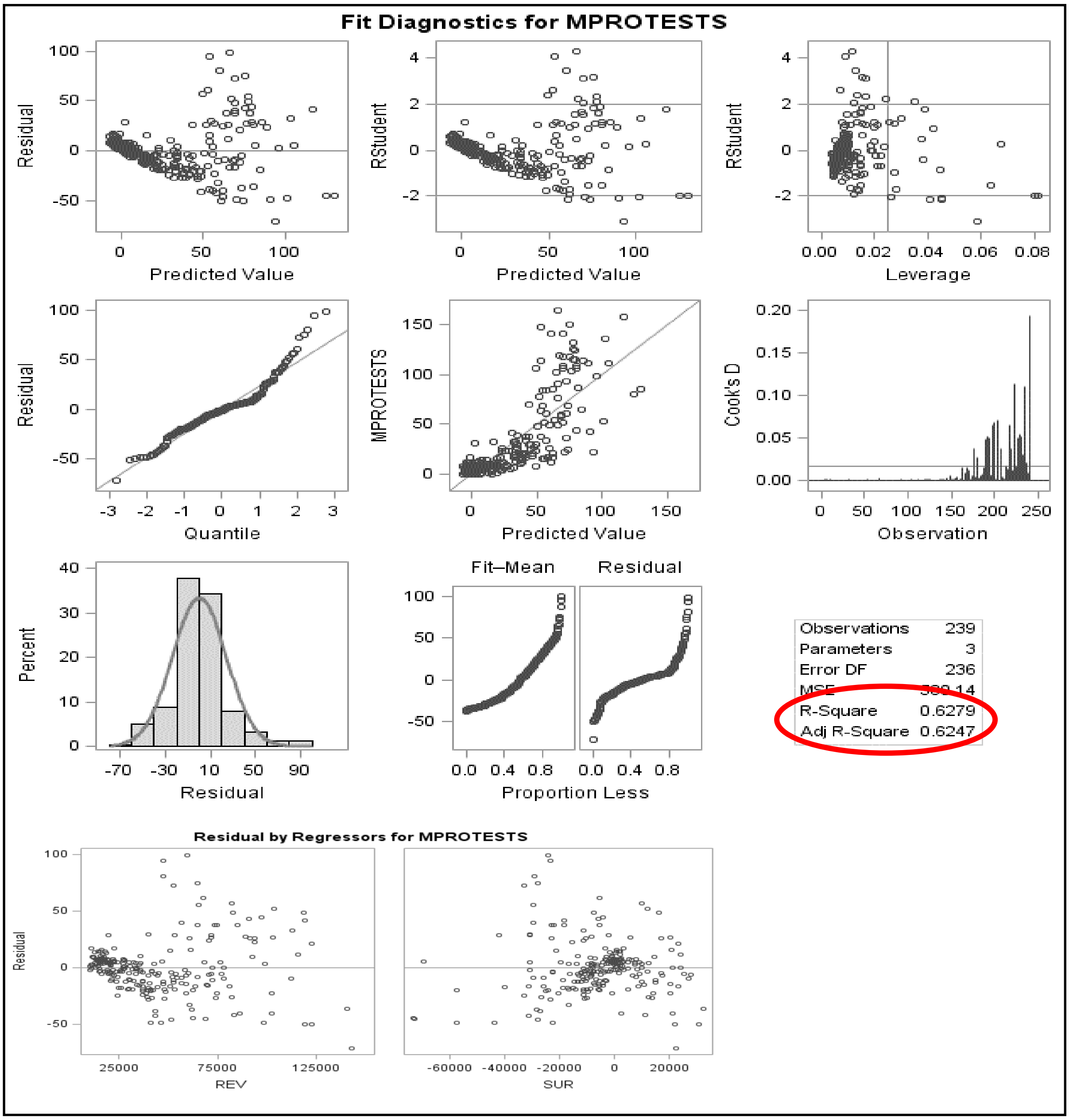
The final model is significant as the Pr > F value show in the Analysis of Variance, indicated with the red oval, is smaller than 0.05. The parameter estimates of this model are indicated with a red block on the next page.

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 2 | 231016 | 115508 | 199.10 | <.0001 |
| **Error** | 236 | 136913 | 580.13841 | | |
| **Corrected Total** | 238 | 367929 | | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| **Intercept** | -19.38466 | 2.99875 | 24242 | 41.79 | <.0001 |
| **REV** | 0.00096962 | 0.00005465 | 182599 | 314.75 | <.0001 |
| **SUR** | -0.00114 | 0.00009434 | 84296 | 145.30 | <.0001 |

Further diagnostics on the regression model are given in the fit diagnostics section that follows. This includes the R-squared, adjusted R-squared, scatter plots of the residuals and predicted values as well as tests for normality among others.

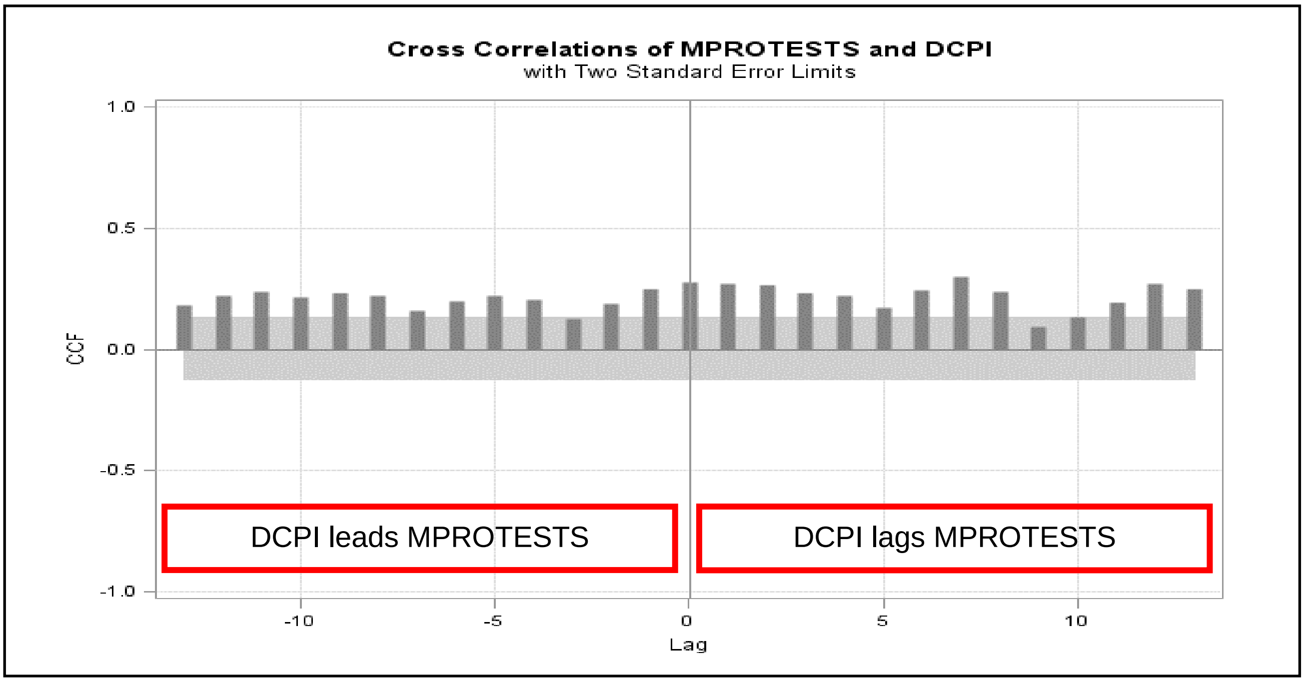


## E.2 Lagged regression model

The regression model that has just been discussed only makes use of the immediate relationship that is present between variables. This is not the only relationship that may exist between time series variables, as delayed relationship may exist. Lagged regression makes use of these relationships when building a model. To incorporate this, an additional step, is performed when building this type of model. The entire process is described below.

The first step to build a lagged regression model is identical to the regression model. To prevent the emergence of co-integration in the model, Augmented Dicky-Fuller tests are performed on all the variables and all the non-stationary variables are removed. Before the forward stepwise selection is performed, cross correlations are done. This is done to identify the presence of delayed relationships.

An example of the cross correlation between the number of protests and the first difference of CPI is shown below. The correlation at a lag is significant if the bar plots outside the 95% confidence interval, which is indicated by the light blue shaded area on the graph. In this case there is a significant relationship between these two-time series as lags -13, -12, -11, -10, -9, -8, -7, -6, -5, -4, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 12 and 13 all plot outside the 95% confidence interval.  Further analysis of the cross correlations also shows the direction of the relationships that exist. If a cross correlation is identified for any lag between -1 and -13, it shows that a change in the DCPI precedes a change in the MPROTESTS. Whereas cross correlations present for any lag between 1 and 13 indicates that changes in DCPI occur as a result of changes in MPROTESTS. In this case we are interested in variables that lead to MPROTESTS.



After the cross correlations between the stationary economic variables and MPROTESTS have been analysed, the appropriate variables are lagged. Thereafter forward stepwise selection, identical to what was done with the regression model, are performed with the lagged variables.

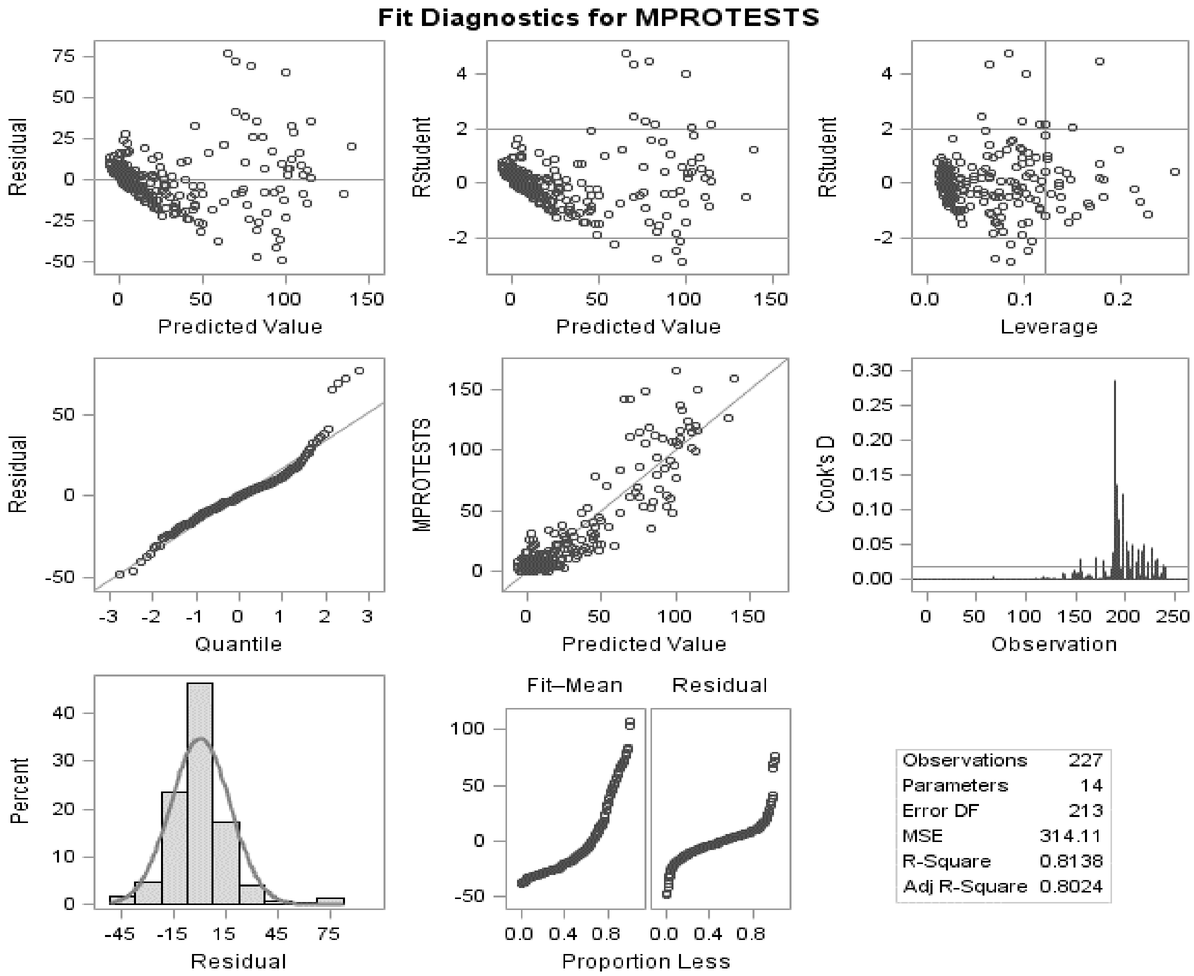
The lagged regression model that best describes the MPROTESTS government surplus (SUR), government revenue (REV), the first difference of the government expenditure (EXP), internet usage (INTER1IND), the first difference of the FPI and DCPI. The statistical output pertaining to this model is displayed until the end of this appendix and is analysed in the same manner that was described above, in the regression model and will therefore not be repeated.

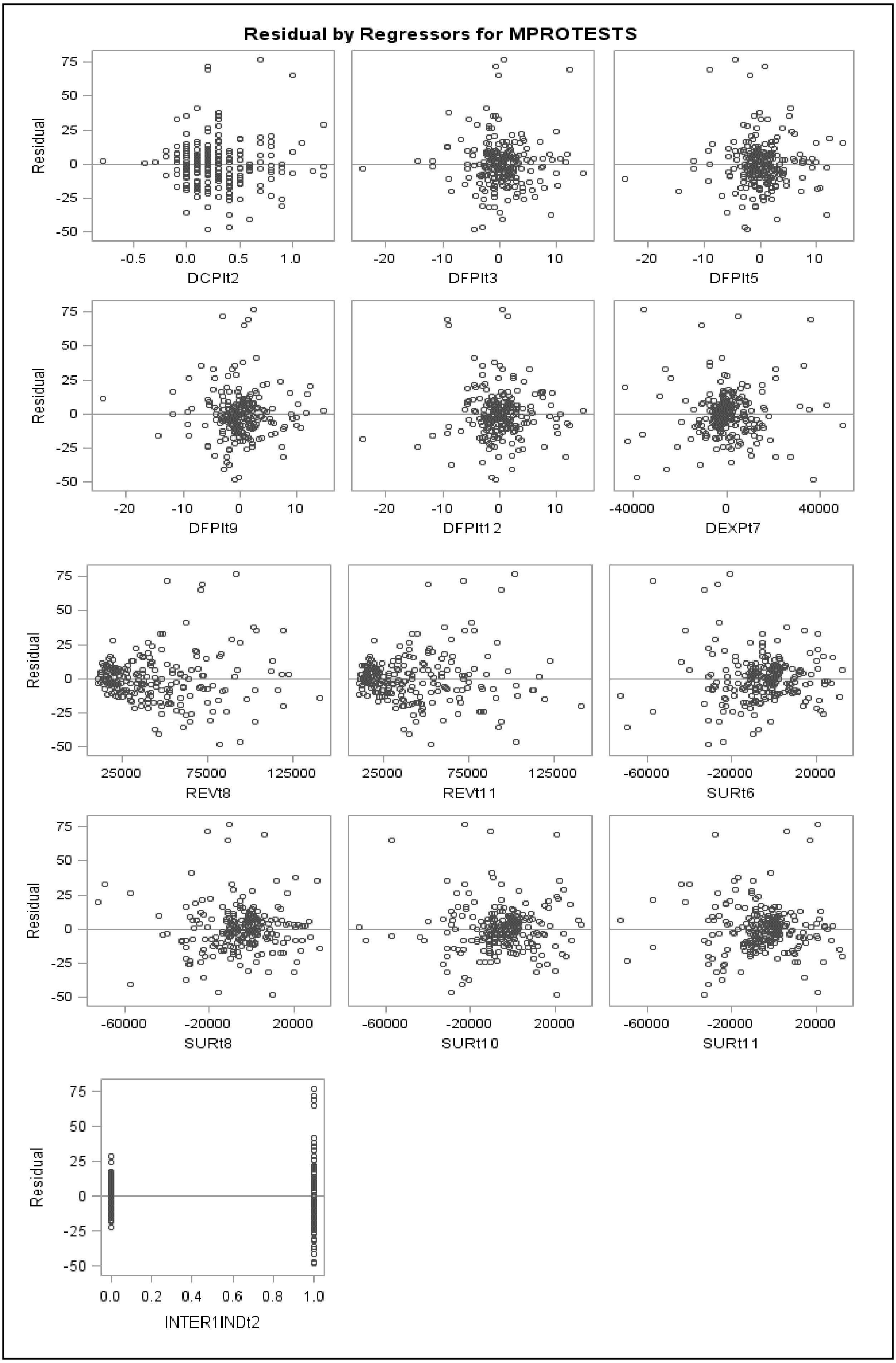
**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 13 | 292393 | 22492 | 71.60 | <.0001 |
| **Error** | 213 | 66906 | 314.11439 | | |
| **Corrected Total** | 226 | 359299 | | | |

| Summary of Stepwise Selection | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In Model** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | REVt8 | | 1 | 0.5387 | 0.5387 | 297.330 | 262.78 | <.0001 |
| **2** | SURt8 | | 2 | 0.1252 | 0.6639 | 158.113 | 83.44 | <.0001 |
| **3** | DEXPt5 | | 3 | 0.0231 | 0.6870 | 134.078 | 16.44 | <.0001 |
| **4** | SURt11 | | 4 | 0.0193 | 0.7063 | 114.272 | 14.61 | 0.0002 |
| **5** | DFPIt9 | | 5 | 0.0172 | 0.7235 | 96.9256 | 13.71 | 0.0003 |
| **6** | DEXPt9 | | 6 | 0.0128 | 0.7363 | 84.4999 | 10.67 | 0.0013 |
| **7** | DFPIt3 | | 7 | 0.0102 | 0.7465 | 74.9791 | 8.82 | 0.0033 |
| **8** | DFPIt12 | | 8 | 0.0121 | 0.7586 | 63.2989 | 10.95 | 0.0011 |
| **9** | REVt2 | | 9 | 0.0106 | 0.7692 | 53.3093 | 9.99 | 0.0018 |
| **10** | DRPETROLt12 | | 10 | 0.0056 | 0.7748 | 48.9900 | 5.37 | 0.0214 |
| **11** | INTER1INDt2 | | 11 | 0.0047 | 0.7795 | 45.6868 | 4.58 | 0.0334 |
| **12** | SURt2 | | 12 | 0.0045 | 0.7840 | 42.6447 | 4.43 | 0.0365 |
| **13** | DCPIt9 | | 13 | 0.0040 | 0.7881 | 40.0782 | 4.07 | 0.0449 |
| **14** | DFPIt5 | | 14 | 0.0048 | 0.7928 | 36.6872 | 4.89 | 0.0281 |
| **15** | | SURt2 | 13 | 0.0038 | 0.7891 | 38.9376 | 3.86 | 0.0509 |
| **16** | | DRPETROLt12 | 12 | 0.0031 | 0.7860 | 40.4534 | 3.15 | 0.0775 |
| **17** | SURt12 | | 13 | 0.0050 | 0.7909 | 36.8195 | 5.09 | 0.0251 |
| **18** | DEXPt10 | | 14 | 0.0037 | 0.7946 | 34.6665 | 3.80 | 0.0526 |
| **19** | | DEXPt9 | 13 | 0.0033 | 0.7914 | 36.3494 | 3.37 | 0.0678 |
| **20** | SURt10 | | 14 | 0.0068 | 0.7981 | 30.7008 | 7.12 | 0.0082 |
| **21** | | DCPIt9 | 13 | 0.0023 | 0.7958 | 31.3249 | 2.44 | 0.1195 |
| **22** | | REVt2 | 12 | 0.0037 | 0.7921 | 33.4916 | 3.85 | 0.0509 |
| **23** | SURt6 | | 13 | 0.0100 | 0.8021 | 24.2429 | 10.73 | 0.0012 |
| **24** | | DEXPt5 | 12 | 0.0007 | 0.8014 | 22.9995 | 0.72 | 0.3965 |
| **25** | | SURt12 | 11 | 0.0005 | 0.8009 | 21.5910 | 0.57 | 0.4531 |
| **26** | DEXPt7 | | 12 | 0.0057 | 0.8066 | 17.1544 | 6.31 | 0.0127 |
| **27** | | DEXPt10 | 11 | 0.0033 | 0.8033 | 18.8455 | 3.62 | 0.0584 |
| **28** | DCPIt2 | | 12 | 0.0064 | 0.8098 | 13.5813 | 7.24 | 0.0077 |
| **29** | REVt11 | | 13 | 0.0040 | 0.8138 | 11.0552 | 4.59 | 0.0333 |
| **30** | SURt3 | | 14 | 0.0026 | 0.8164 | 10.1475 | 2.98 | 0.0860 |
| **31** | | SURt3 | 13 | 0.0026 | 0.8138 | 11.0552 | 2.98 | 0.0860 |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| **Intercept** | -26.80081 | 2.94819 | 25958 | 82.64 | <.0001 |
| **DCPIt2** | 10.64259 | 4.04476 | 2174.69161 | 6.92 | 0.0091 |
| **DFPIt3** | -0.78167 | 0.30574 | 2053.20426 | 6.54 | 0.0113 |
| **DFPIt5** | -0.76009 | 0.30198 | 1990.08387 | 6.34 | 0.0126 |
| **DFPIt9** | -0.70761 | 0.28899 | 1883.30166 | 6.00 | 0.0152 |
| **DFPIt12** | -1.11516 | 0.29227 | 4572.90054 | 14.56 | 0.0002 |
| **DEXPt7** | 0.00059278 | 0.00015692 | 4482.21010 | 14.27 | 0.0002 |
| **REVt8** | 0.00157 | 0.00019224 | 20990 | 66.82 | <.0001 |
| **REVt11** | -0.00041032 | 0.00019153 | 1441.63153 | 4.59 | 0.0333 |
| **SURt6** | -0.00069460 | 0.00009649 | 16277 | 51.82 | <.0001 |
| **SURt8** | -0.00112 | 0.00018085 | 12075 | 38.44 | <.0001 |
| **SURt10** | -0.00068894 | 0.00009876 | 15285 | 48.66 | <.0001 |
| **SURt11** | -0.00051457 | 0.00017657 | 2667.80060 | 8.49 | 0.0039 |
| **INTER1INDt2** | -27.46936 | 5.47106 | 7918.47027 | 25.21 | <.0001 |





Residual by Regressors for MPROTESTS
Residual
DCPIt2
DFPIt3
DFPIt5
DFPIt9
DFPIt12
DEXPt7
REVt8
REVt11
SURt6
SURt8
SURt10
SURt11
INTER1INDt2

# Appendix F: Example and explanation of SAS output for a VAR (1) model

An example and interpretation of the SAS output for a VAR (1) model is described below. This example considers both endogenous (affected by the other variables in the model) and exogenous (unaffected by the other variables in the model) variables. There are three endogenous variables, namely the number of monthly protests (MPROTESTS), FPI and inland petrol prices (RPETROL). While there is only one exogenous variable, an internet usage indicator (INTER1IND).

The normal VAR (1) model does not incorporate the Granger-Causality Wald test. This was incorporated to help to determine whether the model improves by adding a variable(s) from the previous time period. If the p-value for the test is smaller than 0.05, then the addition is advantageous, and these variables are described as Granger-causal.

In this example, the effects of FPI, RPETROL and INTER1IND on MPROTESTS are analysed. The p-value for the Granger-Causality Wald test shown below, is smaller than 0.0001. This indicates that changes in the FPI, RPETROL and INTER1IND (all the variables in group 2) lead to a change in MPROTESTS (variables in group 1).

**Granger-Causality Wald Test**

| Test | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| **1** | 3 | 42.11 | <.0001 |

**Test 1: Group 1 Variables:** MPROTESTS
**Group 2 Variables:** FPI RPETROL INTER1IND

The output for a standard VAR (1) model is shown in the rest of the appendix. It is important to note that the model creates an equation for each of the endogenous variables. As exogenous variables are unaffected by the other variables in the model, an equation for these variables are not produced. The output below shows all the models created for the three endogenous models; the explanation will however focus on MPROTEST.

The schematic representation of parameter estimates is used in combination with the model parameter estimates to determine whether the parameters incorporated in the model are significant or not. In the schematic representation a "+" or "–" indicates significant parameters, while "." is indicative of parameters that are not significant. The model parameter estimates work with a t test. Here a p-value ≤ 0.05 indicates that a parameter is significantly different from 0.

The schematic representation, shown below, indicates that the majority of the parameters are significant. The model parameter estimates confirm this. With the exception of the constant variable, all the parameters relating to MPROTESTS (in the red box) had p-values smaller than 0.05.

**Schematic Representation of Parameter Estimates**

| Variable/Lag | C | XL0 | AR1 |
|---|---|---|---|
| **MPROTESTS** | . | + | +-+ |
| **FPI** | . | + | .+- |
| **RPETROL** | . | . | +++ |

**+ is > 2*std error, - is < -2*std error, . is between, * is N/A**

**Model Parameter Estimates**

| Equation | Parameter | Estimate | Standard Error | t Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| **MPROTESTS** | **CONST1** | 2.95222 | 5.01414 | 0.59 | 0.5566 | 1 |
| | **XL0_1_1** | 10.51751 | 4.91779 | 2.14 | 0.0335 | INTER1IND(t) |
| | **AR1_1_1** | 0.55293 | 0.05505 | 10.04 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_1_2** | -0.14726 | 0.05217 | -2.82 | 0.0052 | FPI(t-1) |
| | **AR1_1_3** | 0.03992 | 0.00890 | 4.49 | 0.0001 | RPETROL(t-1) |
| **FPI** | **CONST2** | 1.39876 | 1.11806 | 1.25 | 0.2122 | 1 |
| | **XL0_2_1** | 2.27810 | 1.09658 | 2.08 | 0.0388 | INTER1IND(t) |
| | **AR1_2_1** | 0.00248 | 0.01227 | 0.20 | 0.8398 | MPROTESTS(t-1) |
| | **AR1_2_2** | 1.00710 | 0.01163 | 86.57 | 0.0001 | FPI(t-1) |
| | **AR1_2_3** | -0.00460 | 0.00198 | -2.32 | 0.0213 | RPETROL(t-1) |
| **RPETROL** | **CONST3** | -13.51028 | 9.59705 | -1.41 | 0.1605 | 1 |
| | **XL0_3_1** | -0.72406 | 9.41264 | -0.08 | 0.9387 | INTER1IND(t) |
| | **AR1_3_1** | 0.49517 | 0.10536 | 4.70 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_3_2** | 0.45714 | 0.09986 | 4.58 | 0.0001 | FPI(t-1) |
| | **AR1_3_3** | 0.90868 | 0.01703 | 53.34 | 0.0001 | RPETROL(t-1) |

The schematic representation of cross correlations of residuals show how well the model describes the data. As was seen with the schematic representation of parameter estimates, significance is indicated by “+” or “–“, while insignificance is shown with “.”.

If the model fits well, then it accounts for most of the trends that were present in the initial data. As a result, the residuals from lag 1 onwards will be uncorrelated with one another and the schematic representation, will be filled with “.”.

The schematic representation of cross correlations of residuals for MPROTESTS is shown in the red block below. Significant cross correlations are only observed at lag 4, 6, 7 and 12. As such, the model removed most of the trends within the data and is a good fit.

**Schematic Representation of Cross Correlations of Residuals**

| Variable/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MPROTESTS | +.- | ... | ... | ... | ..- | ... | +.- | +.- | ... | -.. | ... | ... | +.. |
| FPI | .+. | .++ | .++ | .+. | .+. | .+. | ... | ... | ... | ... | .-- | ... | ... |
| RPETROL | -.+ | ..+ | .+. | ... | ... | ... | ... | ... | ..- | ... | ..+ | ..+ | ... |

**+ is > 2*std error, - is < -2*std error, . is between**

The R squared value for the models can be seen in the Univariate Model ANOVA diagnostics, shown below. This shows that 76.55% of the variation in MPROTESTS at time t can be explained using INTER1IND at t, FPI at t-1, RPETROL at t-1 and MPROTESTS at time t-1.

**Univariate Model ANOVA Diagnostics**

| Variable | R-Square | Standard Deviation | F Value | Pr > F |
|---|---|---|---|---|
| MPROTESTS | 0.7655 | 19.20301 | 190.94 | <.0001 |
| FPI | 0.9920 | 4.28193 | 7250.50 | <.0001 |
| RPETROL | 0.9907 | 36.75453 | 6212.79 | <.0001 |

The univariate model white noise diagnostic, show below, gives the Durbin-Watson test statistic. If this value is close to 2 then it concluded that the autocorrelation at lag 1 is close to zero. This was seen in the Durbin-Watson test statistic for MPROTESTS. Here a value equal to 2.04911, which is sufficiently close to 2, was obtained.

**Univariate Model White Noise Diagnostics**

| Variable | Durbin Watson |
|---|---|
| MPROTESTS | 2.04911 |
| FPI | 0.91873 |
| RPETROL | 1.46194 |

The roots of AR characteristic polynomial, shown below, helps to determine whether the VAR (1) model is stationary or not. If the modulus is smaller than one, then the time series is stationary. The modulus for this model, highlighted by the red block, is equal to 0.9812, 0.9812 and 0.5068. All of these values are smaller than one, as such the time series is stationary.

**Roots of AR Characteristic Polynomial**

| Index | Real | Imaginary | Modulus | Radian | Degree |
|---|---|---|---|---|---|
| **1** | 0.98093 | 0.02314 | 0.9812 | 0.0236 | 1.3515 |
| **2** | 0.98093 | -0.02314 | 0.9812 | -0.0236 | -1.3515 |
| **3** | 0.50684 | 0.00000 | 0.5068 | 0.0000 | 0.0000 |

# Appendix G: Example and explanation of the validation process for the VAR (1) model

The VAR (1) model creates forecasts for each of the variables that are included in the model. This independence of projected values from actual observations allows the model to forecast n-steps into the future. This allows the model to be validated using both one-step and n-step forecasting procedure. Both these procedures were performed on the monthly and quarterly VAR (1) models. The examples and explanation for both techniques are done for the monthly VAR (1) model. The n-step forecasting procedure of the VAR (1) model is described first. This is followed by the explanation for the one-step forecasting procedure for the VAR (1) model.

## G.1 The n-step forecasting procedure: VAR (1) model

A VAR (1) model for the number of monthly protests (MPROTESTS) was built using only 19 years' worth of data, the 20th year's data were removed from the data set. This model was built in the same manner that was explained in Appendix F. The best VAR (1) model, shown below, incorporated internet usage, FPI and the inland petrol price.

**Model Parameter Estimates (19 years data)**

| Equation | Parameter | Estimate | Standard Error | t Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| **MPROTESTS** | **CONST1** | 0.57107 | 4.96362 | 0.12 | 0.9085 | 1 |
| | **XL0_1_1** | 9.77822 | 4.75473 | 2.06 | 0.0409 | INTER1IND(t) |
| | **AR1_1_1** | 0.53125 | 0.05680 | 9.35 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_1_2** | -0.11895 | 0.05364 | -2.22 | 0.0276 | FPI(t-1) |
| | **AR1_1_3** | 0.03811 | 0.00898 | 4.25 | 0.0001 | RPETROL(t-1) |
| **FPI** | **CONST2** | 0.89572 | 1.15483 | 0.78 | 0.4388 | 1 |
| | **XL0_2_1** | 2.26118 | 1.10623 | 2.04 | 0.0421 | INTER1IND(t) |
| | **AR1_2_1** | -0.00668 | 0.01322 | -0.51 | 0.6138 | MPROTESTS(t-1) |
| | **AR1_2_2** | 1.01387 | 0.01248 | 81.25 | 0.0001 | FPI(t-1) |
| | **AR1_2_3** | -0.00512 | 0.00209 | -2.45 | 0.0150 | RPETROL(t-1) |
| **RPETROL** | **CONST3** | -14.79281 | 9.76399 | -1.52 | 0.1312 | 1 |
| | **XL0_3_1** | -0.81883 | 9.35308 | -0.09 | 0.9303 | INTER1IND(t) |
| | **AR1_3_1** | 0.44740 | 0.11174 | 4.00 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_3_2** | 0.46432 | 0.10551 | 4.40 | 0.0001 | FPI(t-1) |
| | **AR1_3_3** | 0.91043 | 0.01766 | 51.56 | 0.0001 | RPETROL(t-1) |

A years' worth of forecasts, for MPROTESTS, are created with the use of this VAR (1) model. The 12 forecasted values, standard errors and 95% confidence intervals for the 2016 forecasts are shown below.

| Forecasts | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **229** | 61.55478 | 18.49977 | 25.29588 | 97.81367 |
| | **230** | 69.76638 | 20.94260 | 28.71965 | 110.81311 |
| | **231** | 73.40091 | 21.73465 | 30.80178 | 116.00004 |
| | **232** | 74.79546 | 22.11984 | 31.44137 | 118.14954 |
| | **233** | 75.09616 | 22.38298 | 31.22632 | 118.96599 |
| | **234** | 74.86522 | 22.59915 | 30.57170 | 119.15873 |
| | **235** | 74.37853 | 22.79077 | 29.70944 | 119.04762 |
| | **236** | 73.77159 | 22.96581 | 28.75943 | 118.78375 |
| | **237** | 73.11106 | 23.12780 | 27.78142 | 118.44071 |
| | **238** | 72.42983 | 23.27871 | 26.80439 | 118.05527 |
| | **239** | 71.74422 | 23.41993 | 25.84200 | 117.64643 |
| | **240** | 71.06239 | 23.55252 | 24.90030 | 117.22448 |

The actual number of monthly protests are then compared to the forecasted values. The error (difference between the actual and forecasted number of protests) and the absolute error are calculated for each month. All four these values are shown below.

| | **Actual** | **Forecast** | **Error** | **\|Error\|** |
|---|---|---|---|---|
| **January 2016** | 87 | 61.55478 | 25.44522 | 25.44522 |
| **February 2016** | 126 | 69.76638 | 56.23362 | 56.23362 |
| **March 2016** | 159 | 73.40091 | 85.59909 | 85.59909 |
| **April 2016** | 133 | 74.79546 | 58.20454 | 58.20454 |
| **May 2016** | 116 | 75.09616 | 40.90384 | 40.90384 |
| **June 2016** | 94 | 74.86522 | 19.13478 | 19.13478 |
| **July 2016** | 80 | 74.37853 | 5.62147 | 5.62147 |
| **August 2016** | 99 | 73.77159 | 25.22841 | 25.22841 |
| **September 2016** | 124 | 73.11106 | 50.88894 | 50.88894 |
| **October 2016** | 114 | 72.42983 | 41.57017 | 41.57017 |
| **November 2016** | 68 | 71.74422 | -3.74422 | 3.74422 |
| **December 2016** | 22 | 71.06239 | -49.0624 | 49.06239 |
| **Total** | **1222** | **865.9765** | **356.0235** | **461.6367** |

The error over 12 months was 365 (positive error indicates model underestimated the true value, while a negative value shows overestimation). The absolute error during the 12 months was 462. The MAE and RMAE are both calculated using the absolute error. The RMAE is equal to the absolute error divided by the number of forecasts. For this model, the MAE was equal to 38.469, while the RMAE was equal to 6.202.

## G.2 One-step forecasting procedure: VAR (1) model

This technique is very similar to the forecasting procedure that was just described. A VAR (1) model for the number of monthly protests (MPROTESTS) is built using only 19 years' worth of data (228 observations), the 20th year's data once again removed from the data set. This model is the used to forecast the next month's protests (here only one month is forecasted on the model, previously 12 months where forecasted based on this model). The model and forecast are shown below.

**Model Parameter Estimates (19 years data)**

| Equation | Parameter | Estimate | Standard Error | t Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| **MPROTESTS** | **CONST1** | 0.57107 | 4.96362 | 0.12 | 0.9085 | 1 |
| | **XL0_1_1** | 9.77822 | 4.75473 | 2.06 | 0.0409 | INTER1IND(t) |
| | **AR1_1_1** | 0.53125 | 0.05680 | 9.35 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_1_2** | -0.11895 | 0.05364 | -2.22 | 0.0276 | FPI(t-1) |
| | **AR1_1_3** | 0.03811 | 0.00898 | 4.25 | 0.0001 | RPETROL(t-1) |
| **FPI** | **CONST2** | 0.89572 | 1.15483 | 0.78 | 0.4388 | 1 |
| | **XL0_2_1** | 2.26118 | 1.10623 | 2.04 | 0.0421 | INTER1IND(t) |
| | **AR1_2_1** | -0.00668 | 0.01322 | -0.51 | 0.6138 | MPROTESTS(t-1) |
| | **AR1_2_2** | 1.01387 | 0.01248 | 81.25 | 0.0001 | FPI(t-1) |
| | **AR1_2_3** | -0.00512 | 0.00209 | -2.45 | 0.0150 | RPETROL(t-1) |
| **RPETROL** | **CONST3** | -14.79281 | 9.76399 | -1.52 | 0.1312 | 1 |
| | **XL0_3_1** | -0.81883 | 9.35308 | -0.09 | 0.9303 | INTER1IND(t) |
| | **AR1_3_1** | 0.44740 | 0.11174 | 4.00 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_3_2** | 0.46432 | 0.10551 | 4.40 | 0.0001 | FPI(t-1) |
| | **AR1_3_3** | 0.91043 | 0.01766 | 51.56 | 0.0001 | RPETROL(t-1) |

**Forecast (January 2016)**

| Variable | Obs | Forecast | Standard Error | 95% Confidence Limits | |
|---|---|---|---|---|---|
| **MPROTESTS** | **229** | 61.55478 | 18.49977 | 25.29588 | 97.81367 |

One data point (actual value for January 2016) is added to the data set which now contains 229 observations. This data set is then used to recalibrate the VAR (1) model variables included remain the same, but the parameter estimates change as a result. This model is the used to forecast the next month's protests. The output of the new model's parameters and forecast is shown below.

**Model Parameter Estimates (19 years and 1 month's data)**

| Equation | Parameter | Estimate | Standard Error | t Value | Pr > \|t\| | Variable |
|---|---|---|---|---|---|---|
| **MPROTESTS** | **CONST1** | 1.23254 | 4.94795 | 0.25 | 0.8035 | 1 |
| | **XL0_1_1** | 10.23778 | 4.75089 | 2.15 | 0.0322 | INTER1IND(t) |
| | **AR1_1_1** | 0.52230 | 0.05651 | 9.24 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_1_2** | -0.13175 | 0.05288 | -2.49 | 0.0134 | FPI(t-1) |
| | **AR1_1_3** | 0.04017 | 0.00886 | 4.53 | 0.0001 | RPETROL(t-1) |
| **FPI** | **CONST2** | 0.82036 | 1.14766 | 0.71 | 0.4755 | 1 |
| | **XL0_2_1** | 2.20883 | 1.10195 | 2.00 | 0.0462 | INTER1IND(t) |
| | **AR1_2_1** | -0.00566 | 0.01311 | -0.43 | 0.6663 | MPROTESTS(t-1) |
| | **AR1_2_2** | 1.01533 | 0.01226 | 82.79 | 0.0001 | FPI(t-1) |
| | **AR1_2_3** | -0.00536 | 0.00206 | -2.61 | 0.0098 | RPETROL(t-1) |
| **RPETROL** | **CONST3** | -13.93520 | 9.71099 | -1.43 | 0.1527 | 1 |
| | **XL0_3_1** | -0.22301 | 9.32424 | -0.02 | 0.9809 | INTER1IND(t) |
| | **AR1_3_1** | 0.43580 | 0.11091 | 3.93 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_3_2** | 0.44772 | 0.10378 | 4.31 | 0.0001 | FPI(t-1) |
| | **AR1_3_3** | 0.91310 | 0.01739 | 52.51 | 0.0001 | RPETROL(t-1) |

**Forecast (February 2016)**

| Variable | Obs | Forecast | Standard Error | 95% Confidence Limits | |
|---|---|---|---|---|---|
| **MPROTESTS** | **230** | 85.80445 | 18.53397 | 49.47853 | 122.13037 |

Another data point (actual value for February 2016) is added to the data set which now contains 230 observations. The VAR (1) model is then recalibrated on the new data set and the next month's protests are forecasted. The model for February 2016 is shown below.

| Model Parameter Estimates (19 years and 2 month's data) | | | | | | |
|---|---|---|---|---|---|---|
| **Equation** | **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** | **Variable** |
| **MPROTESTS** | **CONST1** | 2.27301 | 4.96240 | 0.46 | 0.6474 | 1 |
| | **XL0_1_1** | 10.64148 | 4.78437 | 2.22 | 0.0271 | INTER1IND(t) |
| | **AR1_1_1** | 0.52377 | 0.05695 | 9.20 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_1_2** | -0.14811 | 0.05273 | -2.81 | 0.0054 | FPI(t-1) |
| | **AR1_1_3** | 0.04215 | 0.00888 | 4.75 | 0.0001 | RPETROL(t-1) |
| **FPI** | **CONST2** | 0.87342 | 1.14005 | 0.77 | 0.4444 | 1 |
| | **XL0_2_1** | 2.22942 | 1.09915 | 2.03 | 0.0437 | INTER1IND(t) |
| | **AR1_2_1** | -0.00559 | 0.01308 | -0.43 | 0.6699 | MPROTESTS(t-1) |
| | **AR1_2_2** | 1.01450 | 0.01211 | 83.75 | 0.0001 | FPI(t-1) |
| | **AR1_2_3** | -0.00526 | 0.00204 | -2.58 | 0.0106 | RPETROL(t-1) |
| **RPETROL** | **CONST3** | -13.40546 | 9.64849 | -1.39 | 0.1661 | 1 |
| | **XL0_3_1** | -0.01747 | 9.30234 | -0.00 | 0.9985 | INTER1IND(t) |
| | **AR1_3_1** | 0.43654 | 0.11073 | 3.94 | 0.0001 | MPROTESTS(t-1) |
| | **AR1_3_2** | 0.43939 | 0.10251 | 4.29 | 0.0001 | FPI(t-1) |
| | **AR1_3_3** | 0.91411 | 0.01727 | 52.94 | 0.0001 | RPETROL(t-1) |

| Forecast (March 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **231** | 107.95035 | 18.68022 | 71.33779 | 144.56291 |

This procedure of adding a data point, recalibrating the model and forecasting the next month is repeated for the remainder of the 12 months. Each of these forecasts are shown on the next page.

The 12 individual forecasts are then combined and compared to the actual number of protests in each month. The error, absolute error, MAE and RMAE are calculated in the same way that was shown in Appendix G.1.

| Forecast (April 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **232** | 125.37049 | 18.93115 | 88.26611 | 162.47486 |

| Forecast (May 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **233** | 115.10806 | 18.89544 | 78.07367 | 152.14245 |

| Forecast (June 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **234** | 105.70568 | 18.85387 | 68.75278 | 142.65857 |

| Forecast (July 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **235** | 94.32548 | 18.82781 | 57.42366 | 131.22731 |

| Forecast (August 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **236** | 86.92965 | 18.80967 | 50.06337 | 123.79593 |

| Forecast (September 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **237** | 92.78245 | 18.78520 | 55.96414 | 129.60076 |

| Forecast (October 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **238** | 105.70366 | 18.85445 | 68.74960 | 142.65771 |

| Forecast (November 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **239** | 102.14942 | 18.82138 | 65.26019 | 139.03864 |

| Forecast (December 2016) | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Obs** | **Forecast** | **Standard Error** | **95% Confidence Limits** | |
| **MPROTESTS** | **240** | 78.25430 | 18.91036 | 41.19067 | 115.31792 |

# Appendix H: Example and explanation of the forecasting process for the lagged regression model

The VAR (1) model creates forecasts for each of the variables that are included in the model. This independence of projected values from actual observations allows the model to forecast n step into the future. The lagged regression model does not have this freedom. Its forecasting ability is dependent on actual variables incorporated. Both the monthly and quarterly lagged regression models can only forecast two-steps into the future, eliminating the use of the n-step forecasting procedure to verify these two models. The example and explanation in this Appendix use monthly protest data.

The procedure is identical to the procedure that was used for the one-step forecasting procedure, explained in Appendix G. The last year's data are removed from the data set and a model is created based on this data set. One additional data point is added to the data set and the parameter estimates of the model are recalibrated. A shortened version of the parameter estimates for January and February 2016 are shown below.

**Parameter Estimates (January 2016)**

| Variable | DF | Parameter Estimate |
|---|---|---|
| Intercept | 1 | -27.021 |
| DCPIt2 | 1 | 7.55137 |
| DFPIt3 | 1 | -0.7019 |
| DFPIt5 | 1 | -0.7878 |
| DFPIt9 | 1 | -0.6676 |
| DFPIt12 | 1 | -1.1999 |
| DEXPt7 | 1 | 0.00061 |
| REVt8 | 1 | 0.00146 |
| REVt11 | 1 | -0.0002 |
| SURt6 | 1 | -0.0006 |
| SURt8 | 1 | -0.0011 |
| SURt10 | 1 | -0.0007 |
| SURt11 | 1 | -0.0005 |
| INT1t2 | 1 | -27.063 |

**Parameter Estimates (February 2016)**

| Variable | DF | Parameter Estimate |
|---|---|---|
| Intercept | 1 | -26.805 |
| DCPIt2 | 1 | 7.669 |
| DFPIt3 | 1 | -0.7295 |
| DFPIt5 | 1 | -0.7421 |
| DFPIt9 | 1 | -0.6718 |
| DFPIt12 | 1 | -1.1732 |
| DEXPt7 | 1 | 0.00061 |
| REVt8 | 1 | 0.00149 |
| REVt11 | 1 | -0.0003 |
| SURt6 | 1 | -0.0006 |
| SURt8 | 1 | -0.0011 |
| SURt10 | 1 | -0.0007 |
| SURt11 | 1 | -0.0005 |
| INT1t2 | 1 | -26.54 |

The method of forecasting with the lagged regression model is not as simple as the VAR (1) model. The forecast is equal to the sum of the intercept and the product of the parameter estimate of each variable and the corresponding variable value. The parameter estimates, actual data and forecast calculation is shown on the next two pages.

**Parameter Estimates ($\beta_1$, $\beta_2$, …, $\beta_{14}$ respectively)**

| | Intercept | DCPIt2 | DFPIt3 | DFPIt5 | DFPIt9 | DFPIt12 | DEXPt7 | REVt8 | REVt11 | SURt6 | SURt8 | SURt10 | SURt11 | INT1t2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan-16 | -27.0212 | 7.5514 | -0.7019 | -0.7878 | -0.6676 | -1.1999 | 0.0006 | 0.0015 | -0.0002 | -0.0006 | -0.0011 | -0.0007 | -0.0005 | -27.0635 |
| Feb-16 | -26.8047 | 7.6690 | -0.7295 | -0.7421 | -0.6718 | -1.1732 | 0.0006 | 0.0015 | -0.0003 | -0.0006 | -0.0011 | -0.0007 | -0.0005 | -26.5401 |
| Mar-16 | -26.4564 | 7.8265 | -0.7118 | -0.7533 | -0.6729 | -1.1513 | 0.0006 | 0.0015 | -0.0003 | -0.0006 | -0.0011 | -0.0007 | -0.0005 | -25.9408 |
| Apr-16 | -27.2683 | 8.0942 | -0.7045 | -0.7420 | -0.6927 | -1.1701 | 0.0006 | 0.0016 | -0.0003 | -0.0006 | -0.0011 | -0.0007 | -0.0005 | -27.6074 |
| May-16 | -27.9630 | 9.8123 | -0.7366 | -0.7630 | -0.6945 | -1.1956 | 0.0006 | 0.0016 | -0.0004 | -0.0006 | -0.0012 | -0.0006 | -0.0005 | -27.9165 |
| Jun-16 | -27.9737 | 9.8185 | -0.7367 | -0.7629 | -0.6960 | -1.1955 | 0.0006 | 0.0016 | -0.0004 | -0.0006 | -0.0012 | -0.0006 | -0.0005 | -27.9353 |
| Jul-16 | -28.0951 | 9.9112 | -0.7299 | -0.7717 | -0.6944 | -1.1970 | 0.0006 | 0.0016 | -0.0004 | -0.0006 | -0.0011 | -0.0006 | -0.0005 | -28.1033 |
| Aug-16 | -27.9365 | 9.9990 | -0.7356 | -0.7704 | -0.7021 | -1.1925 | 0.0006 | 0.0016 | -0.0004 | -0.0006 | -0.0011 | -0.0006 | -0.0005 | -27.8893 |
| Sep-16 | -27.6070 | 10.1540 | -0.7534 | -0.7664 | -0.6946 | -1.1546 | 0.0006 | 0.0015 | -0.0003 | -0.0006 | -0.0011 | -0.0006 | -0.0005 | -27.5469 |
| Oct-16 | -27.9068 | 10.3858 | -0.7200 | -0.7694 | -0.7027 | -1.1439 | 0.0006 | 0.0016 | -0.0003 | -0.0007 | -0.0011 | -0.0006 | -0.0005 | -28.1223 |
| Nov-16 | -27.9777 | 10.3121 | -0.7213 | -0.7650 | -0.7071 | -1.1421 | 0.0006 | 0.0016 | -0.0003 | -0.0007 | -0.0011 | -0.0006 | -0.0005 | -28.2108 |
| Dec-16 | -27.2926 | 10.7140 | -0.7419 | -0.7936 | -0.6979 | -1.1283 | 0.0006 | 0.0016 | -0.0004 | -0.0007 | -0.0011 | -0.0007 | -0.0005 | -27.5219 |

**Data**

| | DCPIt2 | DFPIt3 | DFPIt5 | DFPIt9 | DFPIt12 | DEXPt7 | REVt8 | REVt11 | SURt6 | SURt8 | SURt10 | SURt11 | INT1t2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan-16 | 0.1 | 2.9 | -9.2 | -3.1 | -6.9 | 8382 | 64856 | 103576 | -73040 | -18631 | -11520 | 12056 | 1 |
| Feb-16 | 0.3 | -3 | 0.3 | -1.2 | -3.1 | 49744 | 115862 | 112757 | -8588 | 23992 | -42095 | -11520 | 1 |
| Mar-16 | 0.7 | -1.8 | 2.9 | -2.3 | -4.3 | -43718 | 68574 | 53236 | -5804 | -73040 | -18631 | -42095 | 1 |
| Apr-16 | 1.3 | -4.1 | -3 | -0.7 | -3.1 | -1305 | 89306 | 64856 | -27940 | -8588 | 23992 | -18631 | 1 |
| May-16 | 0.7 | 0.4 | -1.8 | -9.2 | -1.2 | -3728 | 90786 | 115862 | -22114 | -5804 | -73040 | 23992 | 1 |
| Jun-16 | 0.8 | 1.1 | -4.1 | 0.3 | -2.3 | 4150 | 64923 | 68574 | 32433 | -27940 | -8588 | -73040 | 1 |
| Jul-16 | 0.2 | 2 | 0.4 | 2.9 | -0.7 | 11446 | 74898 | 89306 | -30986 | -22114 | -5804 | -8588 | 1 |
| Aug-16 | 0.5 | 3.9 | 1.1 | -3 | -9.2 | -9153 | 140892 | 90786 | 16240 | 32433 | -27940 | -5804 | 1 |
| Sep-16 | 0.8 | 7.2 | 2 | -1.8 | 0.3 | 3163 | 68319 | 64923 | -18283 | -30986 | -22114 | -27940 | 1 |
| Oct-16 | -0.1 | -1.4 | 3.9 | -4.1 | 2.9 | 35229 | 118707 | 74898 | -29855 | 16240 | 32433 | -22114 | 1 |
| Nov-16 | 0.2 | 4.1 | 7.2 | 0.4 | -3 | -42571 | 119414 | 140892 | -23183 | -18283 | -30986 | 32433 | 1 |
| Dec-16 | 0.5 | 4.3 | -1.4 | 1.1 | -1.8 | 1717 | 65271 | 68319 | 23342 | -29855 | 16240 | -30986 | 1 |

**Forecast Calculation**

$$Forecast_{January}\ (t) = \beta_{1:January} + \beta_{2:January}\, DCPI\ (t\text{-}2) + \beta_{3:January}\, DFPI\ (t\text{-}3) + \beta_{4:January}\, DFPI\ (t\text{-}5) + \beta_{5:January}\, DFPI\ (t\text{-}9) + \beta_{6:January}\, DFPI\ (t\text{-}12) + \beta_{7:January}\, DEXP\ (t\text{-}7) + \beta_{8:January}\, REV\ (t\text{-}8) + \beta_{9:January}\, REV\ (t\text{-}11) + \beta_{10:January}\, SUR\ (t\text{-}6) + \beta_{11:January}\, SUR\ (t\text{-}8) + \beta_{12:January}\, SUR\ (t\text{-}10) + \beta_{13:January}\, SUR\ (t\text{-}11) + \beta_{14:January}\, INT1\ (t\text{-}2)$$

| | Intercept | DCPI t2 | DFPI t3 | DFPI t5 | DFPI t9 | DFPI t12 | DEXP t7 | REV t8 | REV t11 | SUR t6 | SUR t8 | SUR t10 | SUR t11 | INT1 t2 | FORECAST |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan-16 | -27.021 | 0.755 | -2.035 | 7.247 | 2.070 | 8.279 | 5.087 | 94.690 | -25.732 | 46.237 | 19.749 | 7.651 | -6.225 | -27.063 | 103.687 |
| Feb-16 | -26.805 | 2.301 | 2.189 | -0.223 | 0.806 | 3.637 | 30.143 | 172.634 | -32.905 | 5.288 | -25.671 | 28.054 | 5.685 | -26.540 | 138.593 |
| Mar-16 | -26.456 | 5.479 | 1.281 | -2.185 | 1.548 | 4.950 | -25.078 | 101.490 | -15.663 | 3.586 | 77.422 | 12.195 | 20.863 | -25.941 | 133.492 |
| Apr-16 | -27.268 | 10.522 | 2.889 | 2.226 | 0.485 | 3.627 | -0.739 | 138.424 | -22.264 | 17.514 | 9.704 | -16.039 | 9.186 | -27.607 | 100.660 |
| May-16 | -27.963 | 6.869 | -0.295 | 1.373 | 6.389 | 1.435 | -2.126 | 144.350 | -43.412 | 14.069 | 6.675 | 46.843 | -11.429 | -27.917 | 114.862 |
| Jun-16 | -27.974 | 7.855 | -0.810 | 3.128 | -0.209 | 2.750 | 2.368 | 103.877 | -25.675 | -20.636 | 32.131 | 5.514 | 34.783 | -27.935 | 89.166 |
| Jul-16 | -28.095 | 1.982 | -1.460 | -0.309 | -2.014 | 0.838 | 6.400 | 118.339 | -31.421 | 19.681 | 24.989 | 3.716 | 4.283 | -28.103 | 88.826 |
| Aug-16 | -27.937 | 5.000 | -2.869 | -0.847 | 2.106 | 10.971 | -5.047 | 222.609 | -32.417 | -10.294 | -36.325 | 17.932 | 2.898 | -27.889 | 117.891 |
| Sep-16 | -27.607 | 8.123 | -5.425 | -1.533 | 1.250 | -0.346 | 1.742 | 104.528 | -21.019 | 11.782 | 34.394 | 14.091 | 14.354 | -27.547 | 106.789 |
| Oct-16 | -27.907 | -1.039 | 1.008 | -3.001 | 2.881 | -3.317 | 19.752 | 183.996 | -25.114 | 19.505 | -18.351 | -20.984 | 11.454 | -28.122 | 110.763 |
| Nov-16 | -27.978 | 2.062 | -2.957 | -5.508 | -0.283 | 3.426 | -24.182 | 186.286 | -48.371 | 15.151 | 20.843 | 20.001 | -16.614 | -28.211 | 93.666 |
| Dec-16 | -27.293 | 5.357 | -3.190 | 1.111 | -0.768 | 2.031 | 1.018 | 101.823 | -25.708 | -15.521 | 33.736 | -10.691 | 15.985 | -27.522 | 50.368 |

The 12 forecasts, in the red box, are then compared to the actual number of protests per month. The error, absolute error, MAE and RMAE are calculated in the same way that was shown in Appendix G.1.

# Appendix I: Testing the forecasting abilities of the protest models

An example and explanation of the one-step and n-step procedure for both the VAR (1) model and lagged regression model was described in Appendix G and H respectively. This Appendix gives additional information about the actual and forecasted number of protests, error, absolute error, MAE and RMAE produced by each of the models. The efficiency of the models are determined by comparing their MAE and RMAE with each other. The n-step forecasting procedure is described first, followed by the one-step forecasting procedure. The Appendix is concluded with a comparison between the one-step and n-step forecasting abilities of VAR (1) models.

## I.1 The n-step forecasting procedure

The n-step forecasting procedure was performed on the monthly and quarterly VAR (1) models. A summary for the monthly VAR (1) model is described first. This is followed by the quarterly VAR (1) model.

### I.1.1 Monthly VAR (1) model

The actual and forecasted number of protests, error and absolute error obtained for the monthly VAR (1) model using the n-step forecasting procedure are shown in Table 43. The VAR (1) model underestimated the number of protests, during the 12-month period, by 365. The absolute error during the period was 462. The MAE are RMAE for this model is equal to 38.469 and 6.202 respectively.

*Table 43: Validation of the monthly VAR (1) model using the n-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| January 2016 | 87 | 61.55478 | 25.44522 | 25.44522 |
| February 2016 | 126 | 69.76638 | 56.23362 | 56.23362 |
| March 2016 | 159 | 73.40091 | 85.59909 | 85.59909 |
| April 2016 | 133 | 74.79546 | 58.20454 | 58.20454 |
| May 2016 | 116 | 75.09616 | 40.90384 | 40.90384 |
| June 2016 | 94 | 74.86522 | 19.13478 | 19.13478 |
| July 2016 | 80 | 74.37853 | 5.62147 | 5.62147 |
| August 2016 | 99 | 73.77159 | 25.22841 | 25.22841 |
| September 2016 | 124 | 73.11106 | 50.88894 | 50.88894 |
| October 2016 | 114 | 72.42983 | 41.57017 | 41.57017 |
| November 2016 | 68 | 71.74422 | -3.74422 | 3.74422 |
| December 2016 | 22 | 71.06239 | -49.0624 | 49.06239 |
| **Total** | **1222** | **865.9765** | **356.0235** | **461.6367** |

### I.1.2 Quarterly VAR (1) model

The actual and forecasted number of protests, error and absolute error obtained for the quarterly VAR (1) model using the n-step forecasting procedure are shown in Table 44. The VAR (1) model underestimated the number of protests, during the four quarters, by 113. The absolute error during the period was 228. The MAE are RMAE for this model is equal to 57.082 and 7.555 respectively.

*Table 44: Validation of the quarterly VAR (1) model using the n-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| 1st Quarter 2016 | 372 | 302.023 | 69.97704 | 69.97704 |
| 2nd Quarter 2016 | 343 | 277.9244 | 65.07565 | 65.07565 |
| 3rd Quarter 2016 | 303 | 267.1852 | 35.81484 | 35.81484 |
| 4th Quarter 2016 | 204 | 261.4644 | -57.4644 | 57.46443 |
| **Total** | **1222** | **1108.597** | **113.403** | **228.332** |

## I.2 One-step forecasting procedure

The one-step forecasting procedure was performed on both the VAR (1) and lagged regression models. A summary for the monthly and quarterly VAR (1) models is described first. This is followed by the monthly and quarterly lagged regression models. This section is concluded with the combined VAR (1) and lagged regression models.

### I.2.1 VAR (1) model

The one-step forecasting procedure was performed on the monthly and quarterly VAR (1) models. A summary for the monthly VAR (1) model is described first. This is followed by the quarterly VAR (1) model.

#### I.2.1.1 Monthly VAR (1) model

The actual and forecasted number of protests, error and absolute error obtained for the monthly VAR (1) model using the one-step forecasting procedure are shown in Table 45. The VAR (1) model underestimated the number of protests, during the 12-month period, by 60. The absolute error during the period was 293. The MAE are RMAE for this model is equal to 24.436 and 4.943 respectively.

*Table 45: Validation of the monthly VAR (1) model using the one-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| January 2016 | 87 | 61.5548 | 25.44522 | 25.44522 |
| February 2016 | 126 | 85.8045 | 40.19555 | 40.19555 |
| March 2016 | 159 | 107.95 | 51.04965 | 51.04965 |
| April 2016 | 133 | 125.37 | 7.62951 | 7.62951 |
| May 2016 | 116 | 115.108 | 0.89194 | 0.89194 |
| June 2016 | 94 | 105.706 | -11.7057 | 11.70568 |
| July 2016 | 80 | 94.3255 | -14.3255 | 14.32548 |
| August 2016 | 99 | 86.9297 | 12.07035 | 12.07035 |
| September 2016 | 124 | 92.7825 | 31.21755 | 31.21755 |
| October 2016 | 114 | 105.704 | 8.29634 | 8.29634 |
| November 2016 | 68 | 102.149 | -34.1494 | 34.14942 |
| December 2016 | 22 | 78.2543 | -56.2543 | 56.2543 |
| **Total** | **1222** | **1161.639** | **60.36123** | **293.231** |

#### I.2.1.2 Quarterly VAR (1) model

The actual and forecasted number of protests, error and absolute error obtained for the quarterly VAR (1) model using the one-step forecasting procedure are shown in Table 46. The VAR (1) model overestimated the number of protests, during the four quarters, by 10. The absolute error during the period was 150. The MAE are RMAE for this model is equal to 37.404 and 6.116 respectively.

*Table 46: Validation of the quarterly VAR (1) model using the one-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| $1^{st}$ Quarter 2016 | 372 | 302.023 | 69.97704 | 69.97704 |
| $2^{nd}$ Quarter 2016 | 343 | 345.035 | -2.03467 | 2.03467 |
| $3^{rd}$ Quarter 2016 | 303 | 327.396 | -24.3964 | 24.3964 |
| $4^{th}$ Quarter 2016 | 204 | 257.207 | -53.2073 | 53.20727 |
| **Total** | **1222** | **1231.66** | **-9.6613** | **149.615** |

### I.2.2 Lagged regression model

The one-step forecasting procedure was performed on the monthly and quarterly VAR (1) models. A summary for the monthly lagged regression model is described first. This is followed by the quarterly lagged regression model.

#### I.2.2.1 Monthly lagged regression model

The actual and forecasted number of protests, error and absolute error obtained for the monthly lagged regression model using the one-step forecasting procedure are shown in Table 47. The lagged regression model underestimated the number of protests, during the 12-month period, by 27. The absolute error during the period was 195. The MAE are RMAE for this model is equal to 16.275 and 4.903423 respectively.

*Table 47: Validation of the monthly lagged regression model using the one-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| January 2016 | 87 | 103.68702 | -16.687 | 16.68702 |
| February 2016 | 126 | 138.59334 | -12.5933 | 12.59334 |
| March 2016 | 159 | 133.49209 | 25.50791 | 25.50791 |
| April 2016 | 133 | 100.66005 | 32.33995 | 32.33995 |
| May 2016 | 116 | 114.86158 | 1.138421 | 1.138421 |
| June 2016 | 94 | 89.16588 | 4.83412 | 4.83412 |
| July 2016 | 80 | 88.82622 | -8.82622 | 8.82622 |
| August 2016 | 99 | 117.89132 | -18.8913 | 18.89132 |
| September 2016 | 124 | 106.78924 | 17.21076 | 17.21076 |
| October 2016 | 114 | 110.76257 | 3.237428 | 3.237428 |
| November 2016 | 68 | 93.665821 | -25.6658 | 25.66582 |
| December 2016 | 22 | 50.367785 | -28.3678 | 28.36779 |
| **Total** | **1222** | **1248.7629** | **-26.7629** | **195.3001** |

#### I.2.2.2 Quarterly lagged regression model

The actual and forecasted number of protests, error and absolute error obtained for the quarterly lagged regression model using the one-step forecasting procedure are shown in Table 48. The lagged regression model underestimated the number of protests, during the 4 quarters, by 233. The absolute error during the period was 195. The MAE are RMAE for this model is equal to 59.501 and 7.714 respectively.

*Table 48: Validation of the quarterly lagged regression model using the one-step forecasting procedure.*

| | Actual | Forecast | Error | Absolute Error |
|---|---|---|---|---|
| 1st Quarter 2016 | 372 | 247.2784 | 124.7216 | 124.72158 |
| 2nd Quarter 2016 | 343 | 247.2023 | 95.7977 | 95.797691 |
| 3rd Quarter 2016 | 303 | 288.2323 | 14.7677 | 14.767715 |
| 4th Quarter 2016 | 204 | 206.715 | -2.715 | 2.7150162 |
| **Total** | **1222** | **989.428** | **232.572** | **238.002** |

### I.2.3 Combination of the VAR (1) and lagged regression models

The VAR (1) and lagged regression models were combined to determine whether the forecasts can be improved by combining both models. The models were combined by taking the average forecast for the VAR (1) and lagged regression models. A summary for the combined monthly models is described first, followed by the combined quarterly models.

#### I.2.3.1 Combination of the monthly VAR (1) and lagged regression models

The actual number of protests, average forecasted number of protests, error and absolute error obtained for the combined monthly VAR (1) and lagged regression models using the one-step forecasting procedure are shown in Table 49. The combined model underestimated the number of protests, during the 12-month period, by 17. The absolute error during the period was 198. The MAE are RMAE for this model is equal to 16.507 and 4.063 respectively.

*Table 49: Validation of the combined monthly VAR (1) and lagged regression model using the one-step forecasting procedure.*

| | **Actual** | **Average Forecast** | **Error** | **Absolute Error** |
|---|---|---|---|---|
| January 2016 | 87 | 82.6209 | 4.3791 | 4.3791 |
| February 2016 | 126 | 112.199 | 13.8011 | 13.8011 |
| March 2016 | 159 | 120.721 | 38.2788 | 38.2788 |
| April 2016 | 133 | 113.015 | 19.9847 | 19.9847 |
| May 2016 | 116 | 114.985 | 1.01518 | 1.01518 |
| June 2016 | 94 | 97.4358 | -3.4358 | 3.43578 |
| July 2016 | 80 | 91.5759 | -11.576 | 11.5759 |
| August 2016 | 99 | 102.41 | -3.4105 | 3.41048 |
| September 2016 | 124 | 99.7858 | 24.2142 | 24.2142 |
| October 2016 | 114 | 108.233 | 5.76688 | 5.76688 |
| November 2016 | 68 | 97.9076 | -29.908 | 29.9076 |
| December 2016 | 22 | 64.311 | -42.311 | 42.311 |
| **Total** | **1222** | **1205.2** | **16.7992** | **198.081** |

#### I.2.3.2 Combination of the quarterly VAR (1) and lagged regression models

The actual number of protests, average forecasted number of protests, error and absolute error obtained for the combined quarterly VAR (1) and lagged regression models using the one-step forecasting procedure are shown in Table 50. The combined model underestimated the number of protests, during the four quarters, by 111. The absolute error during the period was 177. The MAE are RMAE for this model is equal to 44.252 and 6.652 respectively.

*Table 50: Validation of the combined quarterly VAR (1) and lagged regression model using the one-step forecasting procedure.*

| | Actual | Average Forecast | Error | Absolute Error |
|---|---|---|---|---|
| 1st Quarter 2016 | 372 | 274.6507 | 97.34931 | 97.34931 |
| 2nd Quarter 2016 | 343 | 296.1185 | 46.88151 | 46.88151 |
| 3rd Quarter 2016 | 303 | 307.8143 | -4.81434 | 4.814343 |
| 4th Quarter 2016 | 204 | 231.9611 | -27.9611 | 27.96114 |
| **Total** | **1222** | **1110.545** | **111.4553** | **177.0063** |

## I.3 Comparison between the one-step and n-step validation procedure of VAR (1) models

The MAE and RMAE obtained using in-sample testing are lower than what were seen in the out of sample testing. This is as a result of the number of steps that are forecasted prior to the model being calibrated. The monthly out of sample testing technique, forecasts 12 steps without any model recalibration. In contrast, the monthly in-sample testing technique only forecasts one step before the model is recalibrated.

The trustworthiness of forecasts decreases with time (one step into the future is more trustworthy than three steps and three steps forward are more reliable than six). This is because the error between the actual and the forecasted value is the first step is carried over into the second forecasted value, which impacts the next and so forth. This error is amplified as the forecasts go further into the future. This error that is amplified and carried throughout the forecasts is responsible for the MAE and RMAE of out of sample testing being higher than the MAE and RMAE for the one-step procedure.

# Appendix J: Univariate time series models for subcategories of social unrest

There is variability in the proportion of annual protests that turn violent or result in property damage. Similar variability exists in the reasons behind protests. This variability may lead to changes in the proposed models and seasonal patterns in each of the social unrest subcategories (for all protests, violent protests, protests with property damage, education, labour and municipal related protests). To determine these changes univariate models were drawn from the daily monthly, quarterly and annual data for each of the social unrest subcategories. A summary of these models, for each timeframe, is given in this appendix. The summary includes the name of the model that best describes the data, the model's R squared value, presence of seasonality as well as model confirmation. The daily summary is described first followed by the monthly, quarterly and annual models.

## J.1 Univariate time series models for the daily protest data

Variation between the daily trends of the various subcategories of social unrest are present. This is visible in the collective time plot of the social unrest categories, depicted in Figure 32. At closer inspection the peaks and slumps of the different categories do not always co-inside with one another.

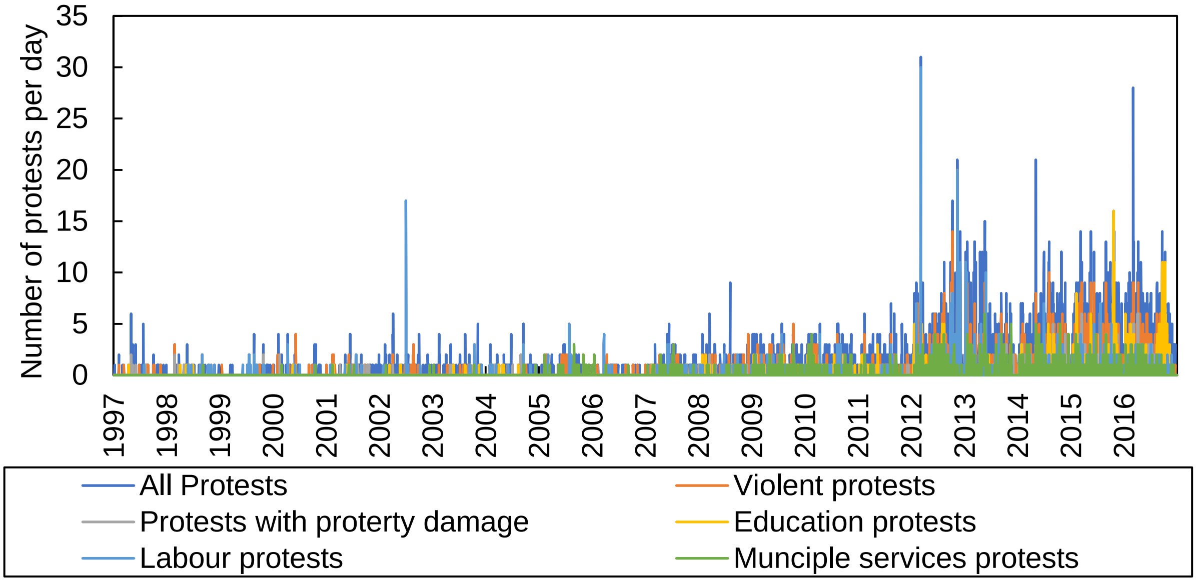



*Figure 32: The daily number of recorded protests for each of the social unrest subcategories.*

These differences are also apparent when looking at the univariate models proposed for each category, depicted in Table 51. The same type of model was proposed for five of the categories, the explanatory power of the models varied greatly. With these models producing

$R^2$ values between 0.1866 and 0.4995. Five of the six subcategories of social unrest displayed seasonal trends. Further variability was observed in these trends. Education-related protests peaked on Mondays, whereas the four remaining subcategories had peaks on Wednesdays. The lowest incidence of protests occurred on Saturdays and Sundays. These models produce high $R^2$ values and account for the seasonal trends that are present within the data. These models, however, do not meet all the model confirmation requirements. Thus, they may be inadequate in describing the data.

*Table 51: Daily univariate time series models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Proposed Model | Seasonal Exponential Smoothing | Seasonal Exponential Smoothing | Seasonal Exponential Smoothing | Seasonal Exponential Smoothing | ARIMA (2,0,0)(1,0,0) | Seasonal Exponential Smoothing |
| Model $R^2$ | 0.4995 | 0.4079 | 0.2741 | 0.3998 | 0.1656 | 0.1866 |
| Seasonality | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ |
| Highest Day | Wednesday | Wednesday | Wednesday | Monday | - | Wednesday |
| Lowest Day | Sunday | Sunday | Saturday | Saturday | - | Saturday |
| **Model Confirmation:** | | | | | | |
| Autocorrelation Not Significant | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Partial Autocorrelation Not Significant | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Residuals Stationary | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Residuals Uncorrelated | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Does the Model Adhere to Guidelines | No | No | No | No | No | No |

## J.2 Univariate time series models for the monthly protest data

There is variation between the monthly trends of the various subcategories of social unrest. This is visible in the collective time plot of the social unrest categories, depicted in Figure 33. Where the peaks and slumps of the different categories do not necessarily co-inside with one another.

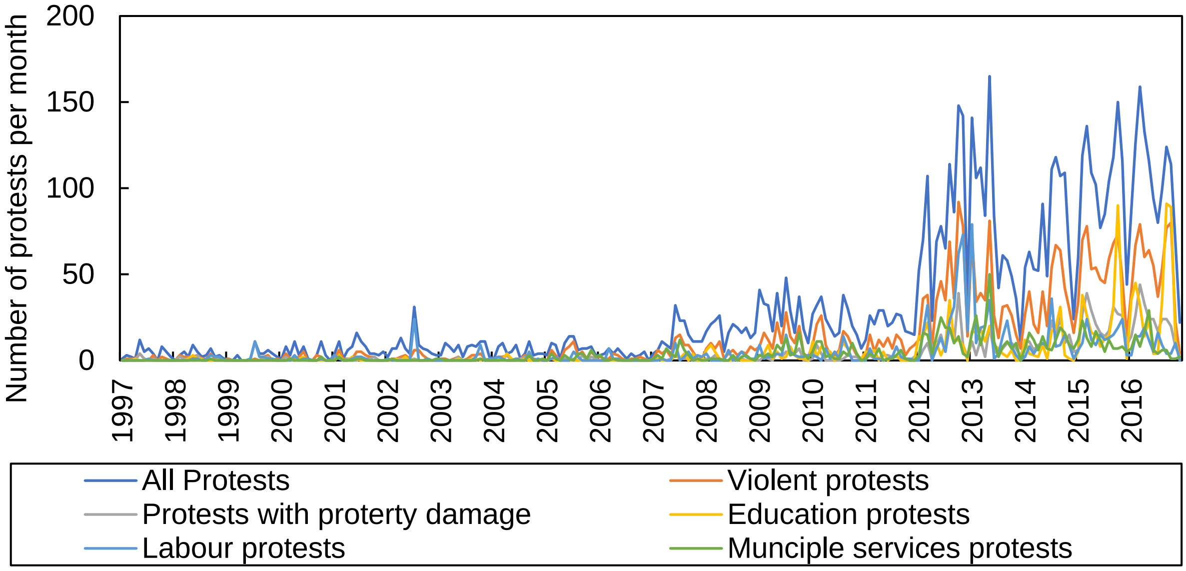



*Figure 33: Monthly number of recorded protests for each of the social unrest subcategories.*

Additional differences are apparent in the univariate time series models that are proposed for each category, depicted in Table 52. The data were described by three different types of models. Once again, the explanatory power of the models varied greatly and the $R^2$ values ranged between 0.3599 and 0.8100. Seasonality was observed in four of the six subcategories. The months with the highest incidence of protests varied and included February, March and May. While December consistently had the lowest incidence. These models produced high $R^2$ values and account for the seasonal trends that are present within the data. Once more, these models do not meet all the model confirmation requirements. Thus, they may be inadequate in describing the data.

*Table 52: Monthly univariate time series models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Proposed Model | Winters Model-Additive | Winters Model-Additive | Winters Model-Additive | ARIMA (2,1,2)(0,1,1) | Seasonal Exponential Smoothing | Seasonal Exponential Smoothing |
| Model $R^2$ | 0.8100 | 0.7399 | 0.6829 | 0.5955 | 0.3599 | 0.5140 |
| Seasonality | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ |
| Highest Month | March | February | March | - | - | May |
| Lowest Month | December | December | December | - | - | December |
| **Model Confirmation:** | | | | | | |
| Autocorrelation Not Significant | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Partial Autocorrelation Not Significant | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Residuals Stationary | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Residuals Uncorrelated | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ |
| Does the Model Adhere to Guidelines | No | No | No | No | No | No |

## J.3 Univariate time series models for the quarterly protest data

There is variation between the quarterly trends of the various subcategories of social unrest. This is visible in the collective time plot of the social unrest categories, depicted in Figure 34. Similarity exists between the peaks and slumps of the different categories; however, the amplitude of these events was not consistent with one another.

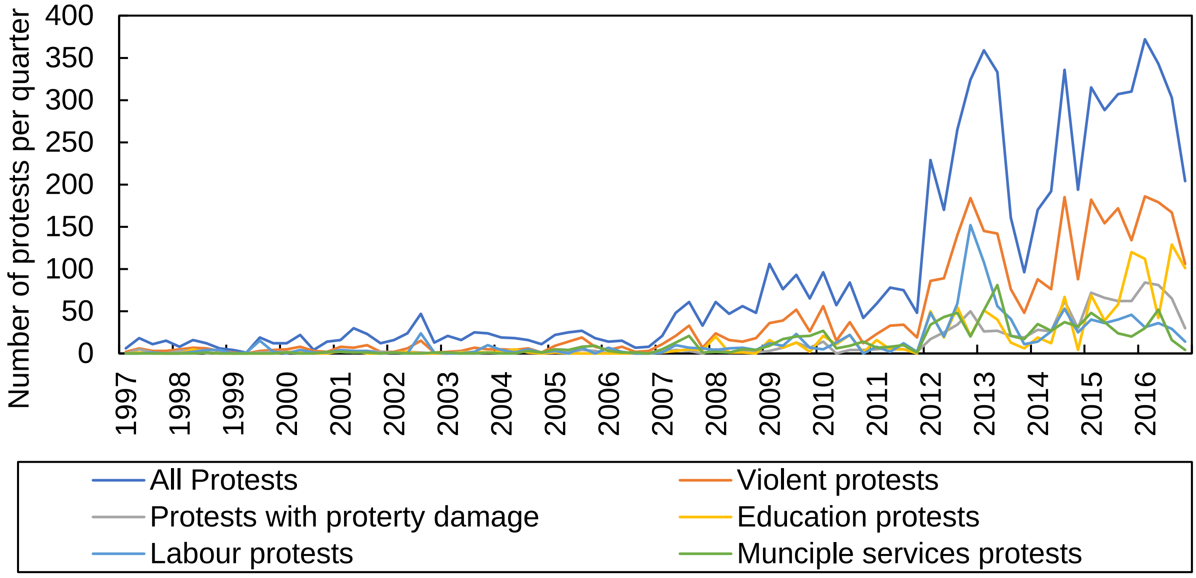



*Figure 34: The quarterly number of recorded protests for each of the social unrest subcategories.*

Additional differences are apparent in the univariate time series models that are proposed for each category, depicted in Table 53. The data were described by three different types of models. Once again, the explanatory power of the models varied greatly and the $R^2$ values ranged between 0.4767 and 0.8493. Seasonality was only observed in all the protests. The first quarter (January, February and March) had the highest incidence of protests, while the last quarter (October, November and December) had the lowest incidence.

The model proposed for protests with property damage not only produced a $R^2$ value of 0.8062, but it complied with all the model confirmation requirements. Thus, meaning that the model adequately describes the data. The other five models did not comply with all the model confirmation requirements and may be inadequate at describing the data.

*Table 53: Quarterly univariate time series models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Proposed Model | Winters Model-Additive | Linear Exponential Smoothing | Linear Exponential Smoothing | Linear Exponential Smoothing | Dampened Trend Exponential Smoothing | Dampened Trend Exponential Smoothing |
| Model $R^2$ | 0.8493 | 0.7967 | 0.8062 | 0.6411 | 0.4767 | 0.5376 |
| Seasonality | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Highest Quarter | 1st | - | - | - | - | - |
| Lowest Quarter | 4th | - | - | - | - | - |
| **Model Confirmation:** | | | | | | |
| Autocorrelation Not Significant | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ |
| Partial Autocorrelation Not Significant | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ |
| Residuals Stationary | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Residuals Uncorrelated | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ |
| Does the Model Adhere to Guidelines | No | No | Yes | No | No | No |

## J.4 Univariate time series models for the annual protest data

As was seen with the daily, monthly and quarterly data, there is variation between the annual trends of the various subcategories of social unrest. This is visible in the collective time plot of the social unrest categories, depicted in Figure 35. Similarity exists between the peaks and slumps of the different categories; however, the amplitude of these events did not co-inside with one another.

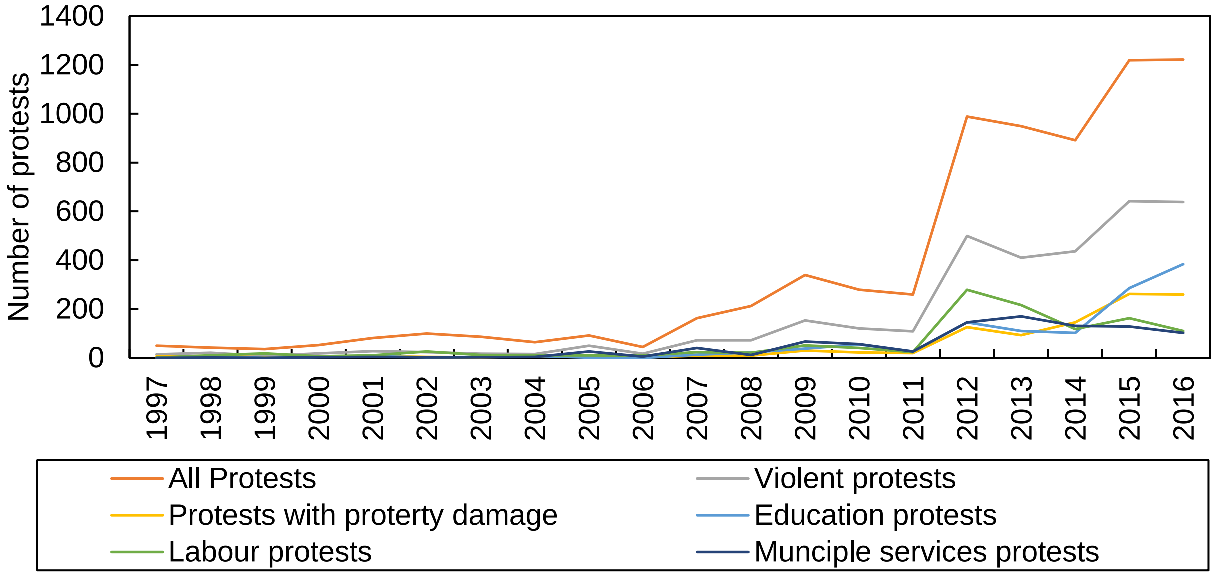



*Figure 35: The annual number of recorded protests for each of the social unrest subcategories.*

Additional differences are apparent in the univariate time series models that are proposed for each category, depicted in Table 54. The data were described by three different types of models. Once again, the explanatory power of the models varied greatly and the $R^2$ values ranged between 0.4830 and 0.8721. These models did well at meeting the model confirmation requirements except for stationarity of residuals, where the models failed consistently. As a result, these may be inadequate in describing the annual data for the social unrest subcategories.

*Table 54: Annual univariate time series models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Proposed Model | Exponential | Log Linear Trend | Linear Exponential Smoothing | Linear Exponential Smoothing | Linear Trend | Linear Trend |
| Model $R^2$ | 0.8721 | 0.8527 | 0.8642 | 0.7688 | 0.4830 | 0.6930 |
| **Model Confirmation:** | | | | | | |
| Autocorrelation Not Significant | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Partial Autocorrelation Not Significant | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Residuals Stationary | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Residuals Uncorrelated | ✓ | ✗ | - | - | ✓ | ✓ |
| Does the Model Adhere to Guidelines | Yes | No | No | No | No | No |

# Appendix K: Linear regression models for subcategories of social unrest

A summary of the monthly and quarterly regression models, described in the results, is given in this appendix. Models have been created for each of the social unrest subcategories (the number of protests, violent protests, protests with property, education-, labour- and municipal-related protests). This summary describes the model's $R^2$ value and the economic and socio-economic variables included into each of the models that have been fitted. The monthly models are described first followed by the quarterly models.

## K.1 Linear regression models for the monthly protest data

Regression models were fitted for the monthly data for each of the social unrest subcategories. All the economic and socio-economic variables, that have been included in each of the models, are significant at a 5% level. A summary of these models, including the model $R^2$ and the variables included, can be found in Table 55.

*Table 55: Monthly linear regression models for each of the social unrest subcategories, with all variables that are statistically significant at a 5% level.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Model $R^2$ | 0.6279 | 0.5662 | 0.4932 | 0.2829 | 0.2352 | 0.4175 |
| **Variables Included into Linear Regression Models:** | | | | | | |
| Change in FPI | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Change in CPI | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Change in Gov Revenue | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| Change in Gov Surplus | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| Internet Indicator | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Change Inland Petrol Price | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Change in Paraffin Prices | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Exchange Rate | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |

The monthly linear regression models produced $R^2$ values between 0.2352 and 0.6279. Government revenue and surplus had predictive power in five of the subcategories of social unrest. In the remaining subcategory, municipal service-related protests, the change in CPI, internet indicator variable and the change in inland petrol prices were all significant predictors. Changes in FPI, inland paraffin prices and the exchange rate did not have predictive power for any of the monthly regression models.

## K.2 Linear regression models for the quarterly protest data

Regression models were fitted for the quarterly data for each of the social unrest subcategories. All the economic and socio-economic variables, that have been included in each of the models, are significant at a 5% level. A summary of these models, including the model $R^2$ and the variables included, can be found in Table 56.

The quarterly linear regression models produced $R^2$ values between 0.0782 and 0.7024. With the internet indicator variable, income growth, changes in FPI, CPI and inland petrol prices, government revenue, surplus and expenditure all having predictive power in at least one of the subcategories of social unrest.

*Table 56: Quarterly linear regression models for each of the social unrest subcategories, with all variables that are statistically significant at a 5% level.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Model $R^2$ | 0.6886 | 0.7024 | 0.0782 | 0.5114 | 0.3199 | 0.5373 |
| **Variables Included into Linear Regression Models:** | | | | | | |
| Internet Indicator | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ |
| Government Surplus | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Income Growth | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Unemployment Rate | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Exchange Rate | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Change in FPI | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Change in CPI | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Change in Gov Revenue | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ |
| Change in Gov Expense | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ |
| Change Inland Petrol Price | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |

# Appendix L: VAR (1) models for subcategories of social unrest

A summary of the monthly and quarterly VAR (1) models, described in the results, is given in this appendix. Models have been created for each of the social unrest subcategories (number of protests, violent protests, protests with property, education-, labour- and municipal-related protests). This summary describes the model's $R^2$ value, variables incorporated as well as model diagnostics for each of the fitted models. The monthly models are described first followed by the quarterly models.

## L.1 VAR (1) models for monthly protest data

VAR (1) models were fitted for the monthly data for each of the social unrest subcategories. Unfortunately, the VAR (1) was unable to control the autocorrelation that was present in the education- and labour-related protest data. The VAR (1) model had no problems with the other monthly protest figures. A summary of these models can be found in Table 57.

*Table 57: Summary of the monthly VAR (1) models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Model $R^2$ | 0.7655 | 0.7492 | 0.6618 | - | - | 0.5630 |
| **Variables Used Within VAR (1) Model:** | | | | | | |
| FPI | ✓ | ✗ | ✓ | - | - | ✓ |
| CPI | ✗ | ✗ | ✓ | - | - | ✓ |
| Inland Petrol Price | ✓ | ✓ | ✗ | - | - | ✗ |
| Basic Diesel Price | ✗ | ✗ | ✗ | - | - | ✓ |
| Paraffin Price | ✗ | ✓ | ✗ | - | - | ✗ |
| Internet Indicator | ✓ | ✗ | ✗ | - | - | ✗ |
| Month Indicators | ✗ | ✓ | ✗ | - | - | ✓ |
| **Model Diagnostics:** | | | | | | |
| Granger-Causal | ✓ | ✓ | ✓ | - | - | ✓ |
| Durbin-Watson | 2.049 | 2.026 | 2.065 | - | - | 1.940 |
| Cross Correlation Plot Acceptable | ✓ | ✓ | ✓ | - | - | ✓ |
| Stationarity | ✓ | ✓ | ✓ | - | - | ✓ |

The monthly VAR (1) models produced $R^2$ values between 0.563 and 0.7655. The FPI, CPI, inland petrol, diesel and paraffin prices, internet indicator and monthly indicators were significant in at least one of the VAR (1) models.

## L.2 VAR (1) models for quarterly protest data

VAR (1) models were fitted for the quarterly data for each of the social unrest subcategories. Unfortunately, the VAR (1) was unable to control the autocorrelation that was present in the violent protests, protests with property damage, education- and labour-related protest data. The VAR (1) model had no problems with all protests and municipal service-related protests. A summary of these models can be found in Table 58.

The quarterly VAR (1) models produced $R^2$ values ranging between 0.5955 and 0.8659. Government surplus, income growth, inland petrol prices and the internet indicator were all significant for Al the protests. While for municipal services-related protests it was only the internet indicator variable that was significant.

*Table 58: Summary of quarterly VAR (1) models for each of the social unrest subcategories.*

| | All Protests | Violent Protests | Protests with Property Damage | Education Protests | Labour Protests | Municipal Service Protests |
|---|---|---|---|---|---|---|
| Model $R^2$ | 0.8659 | - | - | - | - | 0.5955 |
| **Variables Used Within VAR (1) Model:** | | | | | | |
| Government Surplus | ✓ | - | - | - | - | χ |
| Income Growth | ✓ | - | - | - | - | χ |
| Inland Petrol Prices | ✓ | - | - | - | - | χ |
| Internet Indicator | ✓ | - | - | - | - | ✓ |
| **Model Diagnostics:** | | | | | | |
| Granger-Causal | ✓ | - | - | - | - | ✓ |
| Durbin-Watson | 1.95 | - | - | - | - | 1.90 |
| Cross Correlation Plot Acceptable | ✓ | - | - | - | - | ✓ |
| Stationarity | ✓ | - | - | - | - | ✓ |

# Appendix M: SAS code for monthly protests

The basic SAS code of Spearman's correlation, regression, lagged regression and VAR (1) models are given below. The variable names within this basic code was changed to create monthly and quarterly models, and to model the various subcategories of social unrest.

## M.1 Correlations

```
Proc corr data = monthly spearman;
var MPROTESTS INTER1IND FPI CPI REV EXP SUR BPETROL BDIESEL BPARAFFIN
EXCHANGE RPETROL CPARAFFIN RPARAFFIN;
run;
```

## M.2 Regression

Before a regression model can be built, all the variables are evaluated to group all the variables between stationary and non-stationary. In an attempt to transform the non-stationary variables to stationary variables, adjustments are made. One of these adjustments that aid in this process is taking the difference in the variable from one time point to the next. Hereafter tests for stationarity are performed once more. Only after completing this, a regression model is built. The regression models performed in this study only makes use of stationary variables.

### M.2.1 Testing stationarity

```
proc arima data = monthly;
      identify var = MPROTESTS nlag=13 scan esacf minic p = (0:13) q =
(0:13) stationarity = (adf = (0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = INTER1IND nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = FPI nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = CPI nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = REV nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = EXP nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = SUR nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = BDIESEL nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = EXCHANGE nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = RPETROL nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = CPARAFFIN nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = RPARAFFIN nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
run;
```

### M.2.2 Taking the first difference of the non-stationary variables

```
DATA timeseries;
SET monthly;
DFPI = DIF1(FPI);
DCPI = DIF1(CPI);
DEXP = DIF1(EXP);
DBDIESEL = DIF1(BDIESEL);
DRPETROL = DIF1(RPETROL);
DCPARAFFIN = DIF1(CPARAFFIN);
DRPARAFFIN = DIF1(RPARAFFIN);
run;
```

### M.2.3 Testing stationarity once more

```
Proc arima data = timeseries;
      identify var = DFPI nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DCPI nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DREV nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DEXP nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DSUR nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DBDIESEL nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DRPETROL nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DCPARAFFIN nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
      identify var = DRPARAFFIN nlag = 13 stationarity = (adf =
(0,1,2,3,4,5,6,7,8,9,10,11,12));
run;
```

### M.2.4 Building a regression model using only stationary variables

```
ods graphics on;
Proc reg data = timeseries plots = (criteria sbc);
model MPROTESTS = MONTH INTER1IND DFPI DCPI REV SUR EXCHANGE DRPETROL
DCPARAFFIN DRPARAFFIN / SLSTAY= 0.05 selection = stepwise;
Title 'Stepwise regression of monthly protests with full fuel prices';
run;
```

## M.3 Lagged regression

The code for regression and lagged regression is very similar to one another. Stationarity is tested in exactly the same way. Since the data set is exactly the same, the results will not deviate and therefor does not have to be repeated in this scenario. Cross correlation, which is not done in normal regression is performed. This sheds light on underlying relationships that may exist between the variables being analysed. If there are significant relationships present, the data are transformed by adding lagged variables. All of these lagged variables are then included in the lagged regression model.

### M.3.1 Testing for the presence of cross correlations

```
Proc arima data = timeseries;
      identify var = MPROTESTS crosscorr = (DFPI) nlag = 13;
      identify var = MPROTESTS crosscorr = (DCPI) nlag = 13;
      identify var = MPROTESTS crosscorr = (REV) nlag = 13;
      identify var = MPROTESTS crosscorr = (SUR) nlag = 13;
      identify var = MPROTESTS crosscorr = (INTER1IND) nlag = 13;
      identify var = MPROTESTS crosscorr = (DRPETROL) nlag = 13;
run;
```

### M.3.2 Transforming the data into the correct format for lagged regression

```
DATA TIMESERIES13;
SET mprotests13;
      DFPI = DIF1(FPI);
      DCPI = DIF1(CPI);
      DEXP = DIF1(EXP);
      DBPETROL = DIF1(BPETROL);
      DBDIESEL = DIF1(BDIESEL);
      DBPARAFFIN = DIF1(BPARAFFIN);
      DRPETROL = DIF1(RPETROL);
      DCPARAFFIN = DIF1(CPARAFFIN);
      DRPARAFFIN = DIF1(RPARAFFIN);

      MPROTESTSt1 = LAG(MPROTESTS);
      MPROTESTSt2 = LAG(MPROTESTSt1);
      MPROTESTSt3 = LAG(MPROTESTSt2);
      MPROTESTSt4 = LAG(MPROTESTSt3);
      MPROTESTSt5 = LAG(MPROTESTSt4);
      MPROTESTSt6 = LAG(MPROTESTSt5);
      MPROTESTSt7 = LAG(MPROTESTSt6);
      MPROTESTSt8 = LAG(MPROTESTSt7);
      MPROTESTSt9 = LAG(MPROTESTSt8);
      MPROTESTSt10 = LAG(MPROTESTSt9);
      MPROTESTSt11 = LAG(MPROTESTSt10);
      MPROTESTSt12 = LAG(MPROTESTSt11);

      DCPIt1 = LAG(DCPI);
      DCPIt2 = LAG(DCPIt1);
      DCPIt3 = LAG(DCPIt2);
      DCPIt4 = LAG(DCPIt3);
      DCPIt5 = LAG(DCPIt4);
      DCPIt6 = LAG(DCPIt5);
      DCPIt7 = LAG(DCPIt6);
      DCPIt8 = LAG(DCPIt7);
      DCPIt9 = LAG(DCPIt8);
      DCPIt10 = LAG(DCPIt9);
      DCPIt11 = LAG(DCPIt10);
      DCPIt12 = LAG(DCPIt11);

      DFPIt1 = LAG(DFPI);
      DFPIt2 = LAG(DFPIt1);
      DFPIt3 = LAG(DFPIt2);
      DFPIt4 = LAG(DFPIt3);
      DFPIt5 = LAG(DFPIt4);
      DFPIt6 = LAG(DFPIt5);
      DFPIt7 = LAG(DFPIt6);
      DFPIt8 = LAG(DFPIt7);
      DFPIt9 = LAG(DFPIt8);
      DFPIt10 = LAG(DFPIt9);
```

```
DFPIt11 = LAG(DFPIt10);
DFPIt12 = LAG(DFPIt11);

REVt1 = LAG(REV);
REVt2 = LAG(REVt1);
REVt3 = LAG(REVt2);
REVt4 = LAG(REVt3);
REVt5 = LAG(REVt4);
REVt6 = LAG(REVt5);
REVt7 = LAG(REVt6);
REVt8 = LAG(REVt7);
REVt9 = LAG(REVt8);
REVt10 = LAG(REVt9);
REVt11 = LAG(REVt10);
REVt12 = LAG(REVt11);

SURt1 = LAG(SUR);
SURt2 = LAG(SURt1);
SURt3 = LAG(SURt2);
SURt4 = LAG(SURt3);
SURt5 = LAG(SURt4);
SURt6 = LAG(SURt5);
SURt7 = LAG(SURt6);
SURt8 = LAG(SURt7);
SURt9 = LAG(SURt8);
SURt10 = LAG(SURt9);
SURt11 = LAG(SURt10);
SURt12 = LAG(SURt11);

DEXPt1 = LAG(DEXP);
DEXPt2 = LAG(DEXPt1);
DEXPt3 = LAG(DEXPt2);
DEXPt4 = LAG(DEXPt3);
DEXPt5 = LAG(DEXPt4);
DEXPt6 = LAG(DEXPt5);
DEXPt7 = LAG(DEXPt6);
DEXPt8 = LAG(DEXPt7);
DEXPt9 = LAG(DEXPt8);
DEXPt10 = LAG(DEXPt9);
DEXPt11 = LAG(DEXPt10);
DEXPt12 = LAG(DEXPt11);

INTER1INDt1 = LAG(INTER1IND);
INTER1INDt2 = LAG(INTER1INDt1);
INTER1INDt3 = LAG(INTER1INDt2);
INTER1INDt4 = LAG(INTER1INDt3);
INTER1INDt5 = LAG(INTER1INDt4);
INTER1INDt6 = LAG(INTER1INDt5);
INTER1INDt7 = LAG(INTER1INDt6);
INTER1INDt8 = LAG(INTER1INDt7);
INTER1INDt9 = LAG(INTER1INDt8);
INTER1INDt10 = LAG(INTER1INDt9);
INTER1INDt11 = LAG(INTER1INDt10);
INTER1INDt12 = LAG(INTER1INDt11);

DRPETROLt1 = LAG(DRPETROL);
DRPETROLt2 = LAG(DRPETROLt1);
DRPETROLt3 = LAG(DRPETROLt2);
DRPETROLt4 = LAG(DRPETROLt3);
DRPETROLt5 = LAG(DRPETROLt4);
DRPETROLt6 = LAG(DRPETROLt5);
```

```
      DRPETROLt7 = LAG(DRPETROLt6);
      DRPETROLt8 = LAG(DRPETROLt7);
      DRPETROLt9 = LAG(DRPETROLt8);
      DRPETROLt10 = LAG(DRPETROLt9);
      DRPETROLt11 = LAG(DRPETROLt10);
      DRPETROLt12 = LAG(DRPETROLt11);
run;

Proc reg data = timeseries13 plots = (criteria sbc);
      model MPROTESTS = DCPIt1 DCPIt2 DCPIt3 DCPIt4 DCPIt5 DCPIt6 DCPIt7
DCPIt8 DCPIt9 DCPIt10 DCPIt11 DCPIt12 DFPIt1 DFPIt2 DFPIt3 DFPIt4 DFPIt5
DFPIt6 DFPIt7 DFPIt8 DFPIt9 DFPIt10 DFPIt11 DFPIt12 DEXPt1 DEXPt2 DEXPt3
DEXPt4 DEXPt5 DEXPt6 DEXPt7 DEXPt8 DEXPt9 DEXPt10 DEXPt11 DEXPt12 REVt1
REVt2 REVt3 REVt4 REVt5 REVt6 REVt7 REVt8 REVt9 REVt10 REVt11 REVt12 SURt1
SURt2 SURt3 SURt4 SURt5 SURt6 SURt7 SURt8 SURt9 SURt10 SURt11 SURt12
INTER1INDt1 INTER1INDt2 INTER1INDt3 INTER1INDt4 INTER1INDt5 INTER1INDt6
INTER1INDt7 INTER1INDt8 DRPETROLt1 DRPETROLt2 DRPETROLt3 DRPETROLt4
DRPETROLt5 DRPETROLt6 DRPETROLt7 DRPETROLt8 DRPETROLt9 DRPETROLt10
DRPETROLt11 DRPETROLt12  / SLSTAY= 0.05 selection=stepwise  r cli clm;
      Title 'Stepwise lagged regression of monthly protests';
run;
```

## M.4 VAR (1) model and forecast

```
proc varmax data = monthly;
      model MPROTESTS FPI RPETROL = INTER1IND / p=1 print = (diagnose
estimates roots);
      causal group1 = (MPROTESTS) group2 = (FPI RPETROL INTER1IND);
      output out=forecast lead=4;
run;
```

*"Education is the most powerful weapon which you can use to change the world"*

*-Nelson Mandela-*