REGIONAL TEMPORAL AND SPATIAL ANALYSIS OF ANTI-DOPING RULE VIOLATIONS OF ATHLETES

MUCHEMI SAMUEL KARUGU

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE OF MASTERS OF SCIENCE IN MEDICAL STATISTICS OF THE UNIVERSITY OF NAIROBI

2019

DECLARATION

This is my original work and has not been submitted for examination in any other university.

Name of student: **SAMUEL KARUGU MUCHEMI**

Registration Number: **W62/88410/2016**

Signed…….. Date…….

Name of supervisor: **DR OSCAR NGESA**

Senior Lecturer,

Department of Mathematics, Statistics and Physical Sciences

Taita Taveta University

Signed…….. Date…….

Name of supervisor: **DR. ANNE WANG'OMBE**

Lecturer, UNITID

Signed…….. Date…….

ABSTRACT

Background: The number of regional athletes reported with Anti-doping Rule Violations (ADRVs) is a cause for international concerns. Understanding the trends and drivers of ADRVs among athletes is important in developing strategies to curb it. Information on doping from secondary sources was retrieved from World Anti-doping Agency (WADA) for a period of 14 years (2003 – 2016) for time series analysis and for 4 years (2013 – 2016) for Regional Anti-doping Organization (RADO) Zone V spatial analysis.

Data analysis: Exploratory data analysis was carried out to identify any time series features in the data then followed by analysis using Auto-regressive Integrated Moving Average (ARIMA) sets of models following Box-Jenkins procedure. Performance Enhancing Drugs (PEDs) were analyzed with their annual relative frequencies and consistency of use. Granger causality analysis of prize money on ADRVs was also performed.

Findings: The ADRVs in Africa showed a decreasing trend. Anabolic steroids were the most frequently and consistently used PEDs in the study period. There was no Granger causality observed regarding prize money and ADRVs for Africa and six other regions globally. However ADRVs from Paris and Sydney WADA approved laboratories were Granger caused by prize money. In the RADO Zone V, the highest computed four year averages of ADRVs were from Kenya (11) and Egypt (21) with Ethiopia which produces many track athletes showing low numbers (3). The spatial distribution of doping data in the RADO Zone V was virtually random.

Conclusion: Although the ADRVs for Africa and Kenya are on the decrease, the numbers in Kenya were considerably high. Ethiopia is comparable to Kenya in terms of athletic prowess yet ADRVs in Kenya were about 4 times those recorded for Ethiopia. This calls for intensified doping control measures in Kenya and benchmarking with Ethiopia.

DEDICATION

To June Esther Wambui

ACKNOWLEDGEMENT

My sincere gratitude to my supervisors Dr Oscar Ngesa and Dr Anne Wang'ombe, for their guidance and support. To Professor Machoki, The Principal, College of Health Sciences, Dr Mwaengo, The Director, UNITID and to Dr Oyugi, The Deputy Director, UNITID, for creating a conducive learning environment. To Linda, the administrator, and to Purity and Teresa, the administrative assistants, for keeping me informed. To my classmates for being friends and colleagues throughout the course.

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AAF: Adverse Analytical Findings

ACF: Autocorrelation Function

ADAK: Anti-doping Agency of Kenya

ADRVs: Anti-doping Rule Violations

ARIMA: Autoregressive Integrated Moving Average

CAS: Court of Arbitration for Sports

EPO: Erythropoietin

IAAF: International Association of Athletics Federations

IOC: The International Olympics Committee

ISPPPI: International Standard for the Protection of Privacy and Personal Information

PACF: Partial Autocorrelation Function

PEDs: Performance Enhancing Drugs

RADO: Regional Anti-doping Organization

TUEs: Therapeutic use exemptions

WADA: World Anti-doping Agency

TABLE OF CONTENTS

Contents **Page**

CHAPTER ONE

INTRODUCTION

1.0 Background

The most important sports genre in the Eastern African region is track and road athletics. Kenya and Ethiopia are household names in global events currently and historically. Most Anti-doping rule violations (ADRVs) in the region come from this genre. Doping is the single biggest threat to the integrity and reputation of sports (Lenskyj, 2018). Sports provide revenue to many athletes and their associates (Cortsen, 2013) and has led to growth of many associated sectors (Amestica, 2016; Satta, 2016; Sanctis, 2017). The money involved could have led to the increased cases of doping (Héas & Régnier, 2016).

The International Olympics Committee (IOC) threatened to pull some countries out of Rio-2016, the latest event, due to lack of compliance in doping regulations (Lenskyj, 2018). The government of Kenya crafted laws to guide the anti-doping process through relevant institutions and was allowed to participate in the 2016 Olympics. Russia, on the other hand, was accused of state sponsored doping and wasn't allowed to participate as a country. Individual athletes who passed doping tests participated independently.

The use of performance enhancing drugs (PEDs) in sports is prohibited and regulations on anti-doping are governed by World Anti-Doping Agency (WADA), which has subsidiaries distributed globally with accredited testing laboratories (Ritchie & Henne, 2018; Malcourant, Vas & Zintz, 2015). It is a meta-organization being formed by other organizations (Malcourant, Vas & Zintz, 2015). It is supported by national and international sporting

organizations, national governments, and the United Nations (Chappelet &Luijk, 2018). Targeted athletes are required to give samples on a regular basis but sometimes they are required impromptu. Testing ensures athletes have no altered physiological capacities and compete equitably (Danylchuk, Stegink & Lebel, 2016).

Three distinct periods can be identified in the war on doping in ascending order of intensity: pre-1967, between 1967 and 1999 and post-1999 (Ritchie & Henne, 2018). WADA, which is supported by United Nations Educational, Scientific and Cultural Organization (UNESCO) anti-doping convention, was formed with the support of the International Olympics Committee (IOC) in 1999 after public outrage over doping (Kayser & Broers, 2012). Several doping scandals have since then been busted in many sports disciplines (Danylchuk, Stegink & Lebel, 2016). The anti-doping activities are comprehensive with a zero tolerance approach.

The IOC which was responsible for the formation of WADA ensures that no athlete who is doping is allowed to participate in the Olympics and Paralympics (Snyder, Fay & DePerno, 2005). Athletes are either suspended before participation or their medals withdrawn if the discovery is made after the competition. The operation of IOC is controlled by member governments, but in return, the IOC also has an influence on the running of government affairs on sports (Lee, 2017). In 1983, the IOC created the Court of Arbitration for Sports (CAS). Whenever a sports person feels justice is not served at national or international level, they are given the CAS as an appeal mechanism (Lenskyj, 2018). The athletes are the key stakeholders of sports organizations and Bamford and Dehe (2016) argue that it is important to ensure fairness and equity. Adverse analytical findings (AAFs) which could translate to ADRVs form the greatest proportion of cases in CAS (Lenskyj, 2018). Forster (2006) observed that the structure of global sporting organizations pose a challenge to the running of the organizations as there are contradictions within. Lewnskyj

(2018) asked whether matters in CAS are civil or criminal, whether it were strictly arbitration or litigation is allowed and whether matters should be made public or conducted confidentially. This has led to controversy in doping as shall be discussed later in this document. Sanctis (2017) observes that due to the complex nature of global organizations involving national, regional and global aspects, they are vulnerable to abuses and organized crime.

Doping is generally defined as the use of PEDs by athletes (Blumrodt & Kitchen, 2015). Kayser and Broers (2012) extend this further and define doping as the use or attempted use of a prohibited substance. This definition also includes possession of such substances. The strict liability rule states that an athlete is responsible for whatever is administered in their body such that the presence of a substance or its metabolite in urine or blood samples leads to a violation. Doping could thus be by design where the violation is intentional or by default where the athlete didn't violate the doping code willingly (Mazanov et al, 2014).

Another aspect is the availability of an athlete, whereby when required to give a sample for testing, must avail themselves within stipulated time failure to which it amounts to a violation. Certain target athletes must also inform relevant authorities of their whereabouts whenever required. This could lead to awkward ambiguity. Christian Ohuruogu, a 400m British athlete, was banned for a year in 2006 for failing to give a sample for testing three times for unannounced and unscheduled out-of-competition testing (Kayser & Broers, 2012). This was in spite of the fact that other agencies had tested her within the same period and found her clean. Yanina Wickmayer, after rising to top 50 on Women Tennis Association, failed to inform Flemish anti-doping authorities of her where about three times and was banned by the authority (Kayser & Broers, 2012).

Claudia Pechstein was banned in 2009 for having unusually high number of oxygen carrying red blood cells (RBCs) even though no banned substance or their metabolites were found in the blood or urine samples (Kayser & Broers, 2012). Other physiological mechanisms other than doping could explain this. For instance, high altitude training and a number of metabolic conditions could also lead to high RBC levels.

In 2015, WADA published the ten-point anti-doping code that was adopted by her subsidiaries globally. It listed the omissions and commissions constituting ADRV. The presence of a banned substance in urine or blood samples constitutes an ADRV, but the code is also violated if there is evidence of possession or use even without a positive sample. The whereabouts of an athlete must be known by authorities at all times and whenever they demand a sample refusal constitutes a violation. Possession and trafficking of banned substances is a violation. Administering the same, complicity with it and even being associated with an individual who is known to be involved in doping is also a violation. Lastly, tampering with any process of anti-doping is an offence (WADA, 2015).

Testosterone is a natural hormone but is also used as a PED. Caster Semenya is a South African athlete majoring on 800m who has led to controversy due to her gender ambiguity and hyperandrogenism which leads to better performance than females with normal levels of sex hormones. Lenskyj (2018) notes that in all sports, it is only in horse racing that males and females compete. Sex is binary and that assumption must be maintained and protected from pollution through appropriate sex testing policies (Erikainen, 2017). Each should compete with its own kind. Pape (2017), draws a separation between gender and sex in sports with regard to fairness. The question is whether a male or a hermaphrodite who carries themselves as a female should be allowed to compete against females. This introduces sociological and political undertones in sports

which is in line with what is happening with the rest of humanity (Burnett, 2016). The current president of the International Association of Athletics Federations (IAAF) developed testosterone rules to remove ambiguities and allow fairness in competition.

Ultimately, there is a need for every country to be able to know its rates of doping so as to enhance prevention and control (Blank et al, 2015). Is doping on the increase or not? If it is on the increase, is it by default or by design? And if it is by default, what PED analogues are involved? And if it is by design, what PEDs are used? How do these rates compare regionally and globally? Does the rise in prize money in various sporting disciplines correlate with the doping rates? These are the questions this study seeks to answer.

1.1 Statement of the problem

The numbers of regional athletes reported with ADRVs in the recent past have been high. This is unlike a few years ago when there were hardly any news of regional athletes associated with doping. It has caused international concerns due to the athletic prowess of the Eastern African region to the extent that the IOC threatened a ban on Kenya from the 2016 Olympics in Brazil (Lenskyj, 2018). The situation was resolved by the establishment of Anti-doping Agency of Kenya (ADAK) via Anti-Doping Act of 2016. Formation of national anti-doping agencies in critical countries is an indicator of government commitment to fight the vice. To monitor athletes, samples are collected and analyzed by WADA accredited laboratories in association with national and regional anti-doping agencies, and the reports on ADRVs released by WADA on an annual basis. The purpose of this study was to analyze temporal and spatial data for regional athletes thus generating a tailored regional statistical perspective. Forecasting of ADRVs for the near future is critical for effective and efficient planning. Information on the most common PEDs in ADRVs is also critical to identify the most important substances in use. Correlation of doping numbers with

financial incentives is an important issue to explore. Spatial autocorrelation for Regional Anti-doping Organization (RADO) Zone V is also vital to identify patterns in the distribution of ADRVs within the region. Kenya is in RADO Zone V along Eastern African nations and Egypt. Kenya and Ethiopia are the biggest contributors of champions in long and middle distance races globally and it is therefore important to look at Kenya's neighborhood to better understand domestic and international patterns of ADRVs. This allows for more focused local and regional doping control by WADA associated agencies.

The international sports community is bent on ensuring fairness in sports and anti-doping activities aim at that. Several initiatives on doping control are in place and more should be encouraged. WADA is the overall custodian of doping data, but more refined and targeted doping information can be derived from it. This study aimed at contributing to the integrity of sports locally, regionally and globally with respect to doping.

1.2 Significance of the study

The most important sports genre in the Eastern African region is track and road athletics and most ADRVs in the region come from this genre. This has caused great embarrassment to the countries involved, and the sporting community more specifically. Elite athletes are, therefore, bound to be subjected to more rigorous monitoring as compared to other lower risk athletes.

The government of Kenya, through ADAK, is therefore working towards control of doping. RADO Zone V is in charge of the greater Eastern African region and Egypt. There are other RADOs in Africa and the world doing similar work. However, more specific information is required to facilitate anti-doping. The most recent information worked out against the historical data in terms of time and space will enable the anti-doping programs to be more evidence based. Budgetary justification for higher levels of funding of anti-doping activities will be facilitated.

Kenya and Ethiopia are recognized globally as the cradle of current and retired champions in athletics. Any success in anti-doping activities in the region will therefore have global significance and will form gold standards in global anti-doping activities especially targeting athletes in middle and long distance racing. At the very least, this study will contribute to the discussion on the utility of WADA data and the various derivatives that could be obtained from it. The discussion on disaggregation of WADA data is also critical in that generalization may not be of benefit where specific policy changes in local and regional anti-doping approaches are required.

1.3 General objective

The general objective of this study was to obtain a temporal-spatial overview of regional doping statistics of athletes.

1.3.1 Specific objectives

1. To establish the regional doping trend for 14 years (2003-2016) and compare it with global trends.
2. To establish the critical PEDs giving positive doping tests in the same period globally.
3. To establish the relationship between the regional doping trends and corresponding prize money in the same period and compare it with global observations.
4. To establish the current doping statistics in Kenya and compare them with the statistics in the RADO Zone V region exploring for spatial autocorrelation.

1.4 Assumptions of the study

It was assumed that the data obtained from this study was obtained under the same conditions. Analysis, therefore, did not control for moderating variables as it was assumed that the conditions in data collection and analysis were generally unbiased.

It was also assumed that the observations were in the ratio scale and a zero was actually a zero. The absence of ADRVs in any region was therefore taken to mean that tests were performed on athletes from the respective countries and all tests gave negative doping results.

Lastly, it was assumed that data obtained from a region was representative of the region. For instance data obtained from Bloemfontein in South Africa and Tunis in Tunisia were assumed to represent Africa with negligible omissions or commissions. The data that could have originated from a region and was analyzed in laboratories outside the region was therefore not considered.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

WADA has classified doping agents as follows: anabolic steroids, masking agents, stimulants, beta blockers, beta agonists, cannabinoids, glucocorticoids, hormones and metabolic modulators, chemical and physical manipulation, narcotics, alcohol, enhancement of oxygen transfer and finally peptide hormones, growth factors and related substances. The techniques applied by WADA affiliates to detect the presence of PEDs are broad spectrum identifying several agents at a go (Deventer et al., 2002; Thörngren, Östervall & Garle, 2008).

It has been known for a long time that hormonal changes affect energy metabolism (Hervey, 1974). Anabolic steroid PEDs such as stanozolol and nandrolone enable growth of muscles and development of typical male characteristics. They are analogues of the hormone testosterone. Muscle growth is desirable in some sports disciplines in that it improves strength of an individual. Theorell (2009), notes that the state of mind is positively correlated with anabolism and regeneration. Boxers, footballers and sprint athletes in general abuse this class of drugs. Medically, anabolic androgenic steroids are useful in hypo-androgenism. This class of drugs has side effects which are harmful to health, but users prefer not to focus on that despite their knowledge (Walker & Joubert, 2011).

Masking agents, most commonly diuretics, are another category of PEDs. Diuretics such as furosemide are medically prescribed in hypertension and edema. Diuretics act by increasing the amount of urine produced thus reducing the amount of water in the body (Cadwallader et al, 2010). They, therefore, dilute urine making any prohibited substance or metabolite excreted in urine

difficult to detect while also reducing the amount of the PED in blood rapidly (Delbeke & Debackere, 1988).

Another category of doping agents is stimulants. They increase alertness and reduce fatigue and include drugs like ephedrine, amphetamine and cocaine (Docherty, 2008). The reason for testing positive according to Docherty (2008) are medicinal use, recreational use or purposefully for performance enhancement. Some food supplements may also contain high levels of stimulants that could yield a failed drug test (Baume et al., 2006).

Erythropoietin (EPO) is another PED in use. It stimulates red blood cell production thus enhancing the oxygen carrying capacity of the blood (Souillard et al., 1996). Endurance sports like marathons are aerobic thus requiring oxygen and the more efficient blood is in oxygen transport the more the energy metabolism thus the better the performance of athletes. Athletes training in areas of low oxygen tension naturally increase erythropoiesis and perform better in respective endurance disciplines. EPO doping serves as a shortcut. An advanced form of doping with EPO is using its gene for doping purposes (Neuberger et al., 2012). This allows the body to produce more EPO thus achieving the same purpose as EPO doping.

Clomifene, tamoxifen, meldonium and insulin are among drugs classified by WADA under hormones and metabolic modulators. Hormones are widely used doping agents, and insulin is of special mention as it is known to increase muscle mass for body builders and weight lifters (Sönksen, 2001; Barroso, Mazzoni & Rabin, 2008). Metabolic modulators, on the other hand, modify the effects or side effects of hormones. Tamoxifen, for example, hides the negative effects of exogenous testosterone used in doping. Excess testosterone use leads to higher production of estrogen leading to gynecomastia. Tamoxifen blocks this effect.

Betamethasone, dexamethasone, cortisone and budesonide are classified as glucocorticoids. Medically, they are pain and inflammation relievers and athletes can use them out of competition. Sporting activities involve vigorous use of muscles and joints which lead to wear and tear causing pain and inflammation. The use of this class of drugs, therefore, allow for an athlete to keep going beyond what is naturally possible, thus the prohibition during competitions (Coll et al 2018).

Beta 2 agonists, for example salbutamol, are prescribed for allergies and asthma but are in the WADA prohibited list. This is because they enhance athletic performance (Heuberger, van Dijkman and Cohen, 2018). They increase muscle performance by anabolic activity and also improve ventilation (Pluim et al 2011; Cairns & Borrani, 2015). They are not useful, however, in endurance sports. Inhaled salbutamol does not have this effect but systemically administered salbutamol is effective and therefore prohibited. Asthmatic athletes could use beta 2 agonists under therapeutic use exemptions (TUEs).

Cannabinoids have been used in their natural state for long for recreation and for medicinal purposes (Adams & Martin, 1996). Recently the world is generally accepting products of cannabis for mainstream medicinal use (Borgelt et al, 2013). They reduce nausea, vomiting and anorexia. They also reduce pain and muscle spasms, and this is the effect that has led to their increased use and legalization globally. These pharmacological effects are desirable for athletes and have therefore led to their use as PEDs.

2.1 Doping scandals in cycling

Doping is common in cycling, maybe it has always been part of cycling, and cyclists know about it and approve of it (Lentillon-Kaestner & Carstairs, 2010). Sponsors of cycling teams adopt preemptive and preventive strategies to doping as it has the potential to harm their businesses (Blumrodt & Kitchen, 2015). However, doping scandals have not affected profitability of sponsors

of teams involved so the general public seem less bothered by such vice (Danylchuk, Stegink & Lebel, 2016; Drivdal, Nordahl & Rønes, 2018). The doping culture in cycling has been passed down the generations at elite level. Although PEDs are associated with adverse effects, it was observed by Lentillon-Kaestner, Hagger and Hardcastle (2012) this knowledge has little effect on the decision to dope in cycling.

Lance Armstrong is the most high profile cyclist in recent times. His name became even bigger after he recovered from cancer and still went on to perform well (Bassham & Krall, 2010). In total, he won seven times at the most prestigious cycling event, *Tour de France*, in the course of his career (Kasdan, 2013). When confronted with doping evidence against him, he chose not to contest the charges and was therefore stripped of all his titles. The doping was well organized and executed by his team, the US Postal Service (de Bruijn, Groenleer & van Ruijven, 2016). In a televised interview, Armstrong confessed to having doped using EPO and human growth hormone.

Another notable name is Michael Rasmussen (Savulescu & Foddy, 2014). He was the leader of the 2007 *Tour de France* and was destined to win but was caught doping. He confessed that he had doped for 12 years and had used EPO, growth hormone, testosterone, dehydroepiandrosterone, insulin, insulin-like growth factor 1, cortisone and blood transfusions. It was a huge embarrassment for the athlete and for the organizers of the tour. It raised the issue of zero tolerance to doping since as noted by Savulescu and Foddy (2014), only 10-15% of professional athletes are subjected to doping tests meaning there is as high as 90% chance of not being tested.

As previously noted, the sport of cycling would not be where it is without doping. The historical greats could all have doped. The current great and multiple winner of the Tour *de France*, Christopher Froome, was investigated for doping using Salbutamol (UCI, 2017). The amounts found in his urine sample were not consistent with the 1000 ng ml^{-1} threshold allowed through

TUE and therefore there was a possibility of doping (Heuberger, van Dijkman & Cohen, 2018). This further puts the sport to scrutiny. Heuberger, van Dijkman and Cohen (2018) have, however, shed light on this issue suggesting that the analysis used in pinning down Froome may not be robust enough to say with certainty that he doped. They demonstrated that random urine tests are not sufficient to predict the amount of salbutamol consumed.

2.2 Individual athletes

Ben Johnson won 100m gold in 1988 Olympics in Seoul, Korea, breaking the world record in the course. An in-competition test gave a positive test for anabolic steroid stanozolol. Marion Jones stands out in recent times (Pfister & Gems, 2015). She confessed to have doped using anabolic steroids. She was doping along with her husband C.J. Hunter who was competing in shot put. She was convicted and sentenced to 6 months in prison for lying to federal prosecutors in 2008. Tim Montgomery who held the world record in 100m sprint was banned for doping in 2005. Currently the controversy in sprints is around Justin Gatlin. After the retirement of Usain Bolt, Gatlin is the man to beat in sprints. However, he has been banned twice for doping offences. In 2001, he was found guilty of using amphetamine, which is a stimulant, but in defense he said it was a prescription drug he had used since childhood. His 2 year ban was reduced on appeal. He was caught again doping using anabolic steroids in 2006. In 2015, Justin Gatlin ran his personal best time in the 100m at the age of 33. It is unusual for athletes to run their personal best times at that age and this further fuels the speculation that something unusual is going on in an athlete who is known to have failed multiple doping tests. The question being asked by the athletics community is whether it is prudent for an athlete with multiple doping offences to be allowed back into active competition and whether that would encourage cheating in the general athlete population.

2.3 Doping in Russia

Alfons Bukowski (1858–1921) is generally referred to as the father of anti-doping (Pokrywka et al, 2010). He is wrongly regarded as Russian, yet he was Polish. He developed ways of detecting doping in horses. Several years later, 2016 to be precise, Russia, the country he is identified with was banned from participating in the Olympics due to state sanctioned doping (Ruiz & Schwirtz, 2015). The doping scandal was so well calculated it puzzled WADA (Duval, 2017). The entire system was designed not to give failed doping tests. In 2014, a German broadcaster aired a documentary shedding light to the extent of doping in Russia (Seppelt, 2015). This dossier was followed by a WADA investigation, The Pound, which confirmed the allegations (WADA, 2015). This was later followed by the McLaren Investigation sanctioned by WADA in 2016 which gave further details of the doping scheme (WADA, 2016). Another study by Wintermantel, Wachsmuth and Schmidt (2016) demonstrated that between year 2000 and 2013, Russia gave the highest proportion of failed doping tests at 10.4% followed by USA at 6.8%. This adds credence to other findings.

One of the most high profile athletes banned around the same time is Maria Sharapova (Sumner, 2017). She is a former world number one in tennis and is the only Russian to win all the four tennis majors. She was caught using meldonium which is a metabolic modulator. Other athletes recently caught doping, as recent as February 2018, from Russia are Nadezhda Sergeeva who was caught using trimetazidine classified by WADA as a stimulant and Alexander Krushelnitsky who was using meldonium. They were caught through an in-competition test.

2.4 Doping in Eastern Africa

The region is known globally for its athletic prowess, especially in middle and long distance races. Kenya and Ethiopia are the most important countries globally in these disciplines. Athletes such

as Kenenisa Bekele, Haile Gabreselasie and Eliud Kipchoge come from this region (Flaherty, O'Connor & Johnston, 2016). Athletes like Ben Jipcho, Kipchoge Keino and Daniel Rudisha put Kenya on the athletics map around the time of Kenya's independence, and since then, Kenyans have done well globally and regionally (Kirui, Simotwo & Kipkoech, 2013). Haile and Andargachew (2018) note that doping is not a common topic in Ethiopian athletic community.

Among the first doping cases involving high profile Kenyans was the case of John Ngugi (Manners, 1997). He was banned for four years for failing to provide a sample for analysis during an unplanned random visit by anti-doping officials. He argued that he didn't recognize the individuals and they didn't adequately identify themselves. It was a huge story since Ngugi was a multiple champion in global and regional events. He came back after the ban but his form was bad and he struggled to regain form.

Since then, several athletes have been caught doping. Among them are Asbel Kiprop. He provided a sample which yielded a positive test for EPO (Bloom, 2018). Lucy Kabuu was caught in 2018 using narcotics which are prohibited by WADA. This was just before Africa Senior Athletics Championships in Nigeria. Another athlete caught around the same meeting was Boniface Mweresa who confessed to have been using a supplement called Yeah Buddy (Omulo, 2018). A Kenyan born athlete competing for Bahrain, Ruth Jebet, was also caught in 2018 using EPO. Matthew Kisorio was banned in 2012 for using steroids. He has the third-best time in the world in Half Marathon. Jemimah Sumgong, the 2016 London Marathon champion, was banned in 2017 for 4 years for using EPO. Rita Jeptoo, a Boston Marathon winner, was also banned in 2014 for two years for using EPO. A Kenyan boxer, David Munyasia, was also banned from the 2004 Athens Olympics for testing positive for cathine which is a stimulant (Gathura, 2018).

2.5 Prize money in sports

Sports has grown to become a big business (Fox, 2009). Sports are used for marketing purposes, especially when they are organized in some form of consistent format with a schedule (Benijts, Lagae & Vanclooster, 2011). Soccer leagues like *La Liga* in Spain, English Premier League and *Seria A* in Italy are examples. The Diamond League in field and track athletics is another example. Sponsorship of sporting entities is one of the marketing tactics used by companies (Wilber, 1988). Individuals could also market themselves as brands and gain from sponsorships (Cortsen, 2013). Athletes like Cristiano Ronaldo, Lionel Messi, Tiger Woods and Usain Bolt are big brands.

Participation in sports is motivated by monetary rewards (Wheatcroft, 2016), so it is important to know the extent to which money relates to ADRVs. Money is not the only incentive as there are other sources of motivation (Zhou et al, 2018). Frick (2003) sees sports events as not only an avenue to showcase athletic prowess but also a means to motivate athletes to perform better than others and do their personal best through compensation of their efforts. The effort put by athletes depends on how much they are going to get (Frick & Humpreys, 2011).This has been tested in team sports and individual sports and found to be true. In an attempt to cut on excessive spending, the European football governing body introduced financial fair play, which among other things, limits the amounts paid to players in Europe (Freestone & Manoli, 2017). Leaving wages and transfer fees purely to market forces is likely to raise pay to levels that damage clubs, and football in general.

Monetary rewards for performance is one explanation given to the growth of vices in sports. These include match fixing associated with betting and doping. Match fixing has been shown to be linked to illegal betting, weak individuals who can be exploited as well as poor governance of the

involved sports (Tak, Sam & Jackson**,** 2018). Jones (2010) argues that alcohol manufacturers' sponsorship could lead to drug abuse. Corruption has also been observed in sports (de Sanctis, 2017).

2.6 Time series analysis

A number of models have been developed that deal with data that change with respect to time (Wei, 2006). This class of data has time as an important component, and by definition is longitudinal. A time series is a set of data recorded in chronological order. Anderson (2011) points out that while many statistical analyses are based on independence of individual observations, this may not be the case in time series. Observations are dependent based on the time they occurred.

When describing a time series plot, you describe it in terms of the trend, the seasonal and/or cyclic component, as well as the irregular component (Durbin & Koopman, 2012). The trend has to do with the general direction that the plot takes as to whether it is increasing, stagnating or decreasing. For instance the general trend of world population is an increase but individual countries on the globe may have different trends and this may be useful in standardizing of doping data. If within a year there is a repetitive pattern of the plot we call the data seasonal and when a pattern repeats itself in a frequency longer than a year we call that cyclic data. The number of athletes caught doping may pike during certain events such as the Olympics which comes after four years for example.

One important property of a time series is stationarity. Rosca (2011) states that most time series analyses are based on stationary data, and in cases where the series is not stationary, it is transformed to become stationary before analysis. A stationary series is one whose properties, for instance, mean, variance and autocorrelation do not change over time. Another important binary term is domain, which could be time or frequency domain. Time domain shows how observations

change over time. Frequency-domain shows the number of observations at various represented rate of occurrence over time where each rate goes with the number of times it is reported (Shumway & Stoffer, 2017).

Autoregressive Integrated Moving Average (ARIMA) models have found widespread application in many areas of health science (Sarpong, 2013; Ling & Mahadevan, 2012; Liu, et al., 2011). The models are popular linear models used in time series forecasting. They have used singly and also in combination with other models to construct higher accuracy hybrid models (Khashei & Bijari, 2011; Christodoulos, C., Michalakelis, C., & Varoutas, D. (2010).). Lifting scheme and ARIMA models combine aspects in the spatial domain and time series domain, creating a hybrid that has enhanced forecasting accuracy (Lee & Ko, 2011). Other types of models could also be combined with ARIMA models for specific types of analyses.

ARIMA models have three components p, d and q. The d refers to the number of times a process needs to be differenced to make it stationary. The p and q components refer to the autoregressive (AR) and moving average (MA) orders, respectively. Autocorrelation function (ACF) gives the values of auto-correlation of a series with its lagged values. By plotting these values with the confidence band you obtain an ACF plot. Partial autocorrelation function (PACF) examines for correlation of the residuals with the next lag value and could be plotted just like ACF. For an AR process, the ACF plot dies off, while PACF plot cuts off at the order ‘p’. For a MA process the PACF plot dies off while the ACF plot cuts off at order ‘q’. However modern analytical tools are available that perform these analyses automatically (Hyndman & Athanasopoulos, 2018).

2.6.1 Granger causality

Granger causality (G-causality) has gained popularity in health sciences in the past decade (Bressler & Seth, 2011; Friston, Moran & Seth, 2013; Seth, Barrett & Barnett, 2015). It is the

statistical equivalent of correlation and linear regression modeling when it comes to time series. If X_1 Granger-causes X_2 then values of X_1 can be used to estimate values of X_2. It applies to both time and frequency domains. Prediction is based on vector autoregressive (VAR) modelling (Barnett & Seth, 2014).

Let's take the set F_t as $(x_t, z_t, x_{t-1}, z_{t-1}, \ldots, x_1, z_1)$ and x_t and z_t are vectors. We take that z_t includes y_t and z_t might include other variables other than y_t. Then x_t is G-causes y_t with respect to Ft if the variance of the optimal linear predictor of yt+h based on Ft has smaller variance than the optimal linear predictor of y_{t+h} based on z_t, z_{t-1}, ... for any h. Therefore x_t G-causes y_t if x_t helps predict y_t in the future. When x_t G-causes y_t often y_t Granger causes x_t. (Sørensen, 2005).

Zaiontz (2019) cautions that correlation doesn't necessarily mean causation. It was therefore appropriate to say x G-causes y as opposed to say x causes y. He states that G-causality test is based on the OLS regression model below:

$$\mathbf{y_i} = \boldsymbol{\alpha_0} + \sum_{\mathbf{j=1}}^{\mathbf{m}} \boldsymbol{\alpha_j} \mathbf{y_{i-j}} + \sum_{\mathbf{j=1}}^{\mathbf{m}} \boldsymbol{\beta_j} \mathbf{x_{i-j}} + \boldsymbol{\varepsilon_i}$$

Where: α_j and β_j ~ regression coefficients

ε_i ~ error term.

The test is based on:

$$\mathbf{H_0}: \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \cdots = \boldsymbol{\beta_m} = \mathbf{0}$$

When the null hypothesis is rejected then *x* Granger-causes y.

2.7 Spatial autocorrelation

The concept of spatial autocorrelation was fathered by Michael F. Dacey of University of Washington in the 1950s. It seeks to establish whether there is a relationship between measures of variables of adjacent sub-regions on the map of a region (Getis, 2010; Griffith, 2013). In East Africa for example you may want to know if the number of ADRV in Kenya has any relationship

with numbers in Uganda and Tanzania for example. Several authors have used this technique to analyse biological phenomena and therefore this technique could be used for doping analyses (Debarsy & Ertur, 2010; De Knegt et. al., 2010).

Some of the commonly used measure is Moran's I (Getis & Ord, 2010). Geary's C is also used and is similar to Moran's I. Moran's I is a more universal and is sensitive to extreme values of x. Geary's C is more appropriate in situations of differences in small neighborhoods. Generally Moran's I and Geary's C give similar conclusions. Moran's I is preferred since it is more powerful. The equation for Moran's I is given by:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Where:

n is the number of locations

$\bar{x}$ is the mean of x

w_{ij} are the elements of the weight matrix

$$\boldsymbol{S_0 = \sum_i \sum_j w_{ij}}$$

Geary's C which is similar to the Moran's I is given by:

$$C = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}$$

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter discusses how the secondary data was obtained and analyzed to give useful information. It details the study design, study area description, the population and sampling, the variables under study and data management. Ethical consideration is also discussed.

3.1 Study design

The study adopted an observational study design where data on cases of doping that have been recorded in the past were clustered by year of occurrence to form time series data. Records of all doping cases for the period of 2003 to 2016 were analyzed. For spatial analysis, data for RADO Zone V on ADRVs for the period of 2013 to 2016 was analyzed.

3.2 Study area description

This study focused on Africa and then gave attention to RADO Zone V. The study was therefore two-fold: time series analysis of Africa's ADRVs analyzed along with data from other regions for comparison and then spatial analysis of RADO Zone V ADRVs.

3.3 Study population

The study population were regional elite athletes under the jurisdiction of WADA. These athletes therefore fell within the jurisdictions of the national and regional doping agencies. All active elite athletes in the senior category were included. Retired and juniors were excluded. The statistics only focused on adjudicated cases which had been given final verdict on appeal where applicable.

3.4 Sample size determination

Due to the nature of the study and the aggregation of data, analysis was determined by the availability of specific data. The laboratories that had been giving doping statistics consistently from 2003 were used. A census was conducted. For time series analysis, all the 14 year data (2003 – 2016) was used for Africa. The WADA certified laboratories reporting doping statistics for the region were Bloemfontein and Tunis. Table 3.1 below shows the number of samples analyzed for Africa.

Table 3. 1: Samples analyzed in Bloemfontein and Tunis

Year	Analyzed samples
2003	4294
2004	4652
2005	5125
2006	4657
2007	4785
2008	4672
2009	5752
2010	6313
2011	5238
2012	4245
2013	4029
2014	3658
2015	4,132
2016	1392

For comparison purposes, the corresponding data sets for 8 other regions were analyzed. These regions are USA (Los Angeles and Salt Lake laboratories), Canada (Montreal laboratory), South America (Rio laboratory), Australia (Sydney laboratory), Russia (Moscow and Sochi laboratories), China (Beijing laboratory), France (Paris laboratory) and UK (London and Cambridge laboratories). For spatial analysis, all the RADO Zone V data for 4 years (2013 – 2016) was obtained for the 10 member countries. These are Kenya, Uganda, Tanzania, Burundi, Rwanda, Ethiopia, Eritrea, Sudan, Somalia and Egypt.

3.5 Variables

For descriptive time series analysis, the dependent variable was the ADRV count while the independent variable was time whose unit was years. Doping statistics were also paired with prize money for Granger causality analysis. The dependent variable was the ADRV count, while the prize money was the independent variable. Spatial autocorrelation analysis was carried out where the ADRV count was the dependent variable while the spatial distance was the independent variable.

3.6 Data Sources

The doping data is publicly available in WADA annual reports and these are Anti-Doping Testing Figures and ADRVs Report therefore secondary data was obtained via data mining. The data on prize money is also publicly available from the respective bodies responsible for Chicago, Boston and New York marathons. The choice of these prize monies was informed by the fact that the data from the three marathon majors was consistently available throughout the study period and the money is in USA dollars so no need for conversions across currencies.

3.7 Quality assurance procedures

The data mining and analysis was guided by standard procedures and in conjunction with my supervisors. This is an academically acceptable way of ensuring the results of the study are valid and reliable.

3.8 Ethical Consideration

The data is publicly available from WADA website. The gathering of the data by WADA is based on World Anti-Doping Code's International Standard for the Protection of Privacy and Personal Information (ISPPPI). The data is anonymized and collated to counts per year hence no athlete's information is divulged in the data or during analysis. Therefore there was no risk of harming individuals through breaching confidentiality and anonymity. The study was ratified by the ethical review committee of the University of Nairobi and Kenyatta National Hospital.

3.9 Data management

The data was mined and recorded in excel sheets. The data was stored in multiple devices for security purposes. It was analyzed using R soft-ware. Box-Jenkins analysis, Granger-causality analysis of ADRVs and prize money, and ultimately spatial autocorrelation via Moran's I and Geary's C were conducted. The results were presented in tables and figures.

3.10 Data analysis

The plots of ADRVs obtained in the study period enabled the description of the data in terms of the trend, cyclic nature and the random component (Gerbing, 2016). The seasonal component is excluded since the data is collected annually. Multiplicative relationship was assumed as described below:

$Y_t = T_t \,.\, C_t \,.\, E_t$

Where: Y_t is the available data

$T_t.C_t$ are the predictable trend and cyclic component respectively

E_t is the random component

The Box-Jenkins procedure involves understanding these components within the data and appreciating the implicit patterns within the data to enable forecasting through an appropriate ARIMA model (Din, 2016). For ADRVs it would be important to know current and predicted future values, all factors held constant, to enable planning of control measures.

The autoregressive (AR) component of the model is based on the dependent relationship between observations and lagged observations (Kang, 2017).

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

Where:

$y_{t-1}, y_{t-2} \ldots y_{t-p}$ are the past series values

ε_t is white noise

δ is defined by the following equation $\delta = (1 - \sum_{i=1}^{p} \emptyset_i)\mu$

μ is the process mean

The integrated component relates to transformation by differencing to make the data stationary. This is achieved by subtracting observations from the previous time steps. The equation below explains it:

$$y_t^* = y_t' - y_{t-1}'$$

The moving average component uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. The finite MA model is always stationary. The following equation describes it:

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Where:

q is the order of the process

$\theta_j \varepsilon_{t-j}$ are past values

The general forecasting equation therefore is given by:

$$y_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

The selection of an appropriate model in Box-Jenkins method is based on Akaike's information criterion (AIC). The model with the lowest AIC is the most appropriate model. The AIC values for this study were generated using R software.

AIC = -2(log-likelihood) + 2K

Where:

K is the number of model parameters

Log-likelihood is a measure of model fit.

Box-Ljung test is based on the null hypothesis that the chosen ARIMA model fits (Ljung & Box, 1978). Rejecting the hypothesis means the model doesn't fit. It is performed on residuals of the model and is based on the following equation:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{n-k}$$

Where: r_k is the estimated autocorrelation of the series at lag k

m is the number of lags being tested.

Granger causality looks for correlation between sets of time series data and in this study it related ADRVs to prize money. You could check whether Y_t Granger-causes X_t and also if X_t Granger-causes Y_t. The applied process was automated using R and is based on the following equations:

$$Y_t = \alpha_0 + \sum_{i=1}^{k1} \alpha_i Y_{t-i} + \sum_{i=1}^{k2} \beta_i X_{t-i} + \varepsilon_t$$

$$X_t = x_0 + \sum_{i=1}^{k3} x_i X_{t-i} + \sum_{i=1}^{k4} \delta_i Y_{t-i} + v_t$$

Where:

K is the number of lags

ε and v are error terms

$\alpha, \beta, \chi \; and \; \delta$ are coefficients

The test in R is based on the null hypothesis that X_t does not Granger-cause Y_t and vice versa and therefore a p-value less than 0.05 leads to the rejection of the null hypothesis and the conclusion that Granger-causality does exist.

Moran's I and Geary's C relating the ADRVs observed and their distribution within RADO Zone V are computed based on the equations shown below:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$C = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}$$

Where:

n is the number of locations

$\bar{x}$ is the mean of x

w_{ij} are the elements of the weight matrix

$S_0 = \sum_i \sum_j w_{ij}$

Moran's I values range from -1 to +1 where negative values indicate negative spatial autocorrelation and positive values indicate positive spatial autocorrelation. The values of Geary's C range from zero to unspecified values greater than 1. Values close to zero indicate high positive spatial autocorrelation while values much greater than 1 indicate high negative spatial autocorrelation (Getis, 2010).

CHAPTER FOUR

FINDINGS

4.0 Introduction

This chapter presents the findings from the analysis. It is arranged in the order of the specific objectives. The general objective of this study was to obtain a temporal-spatial overview of regional doping statistics of athletes.

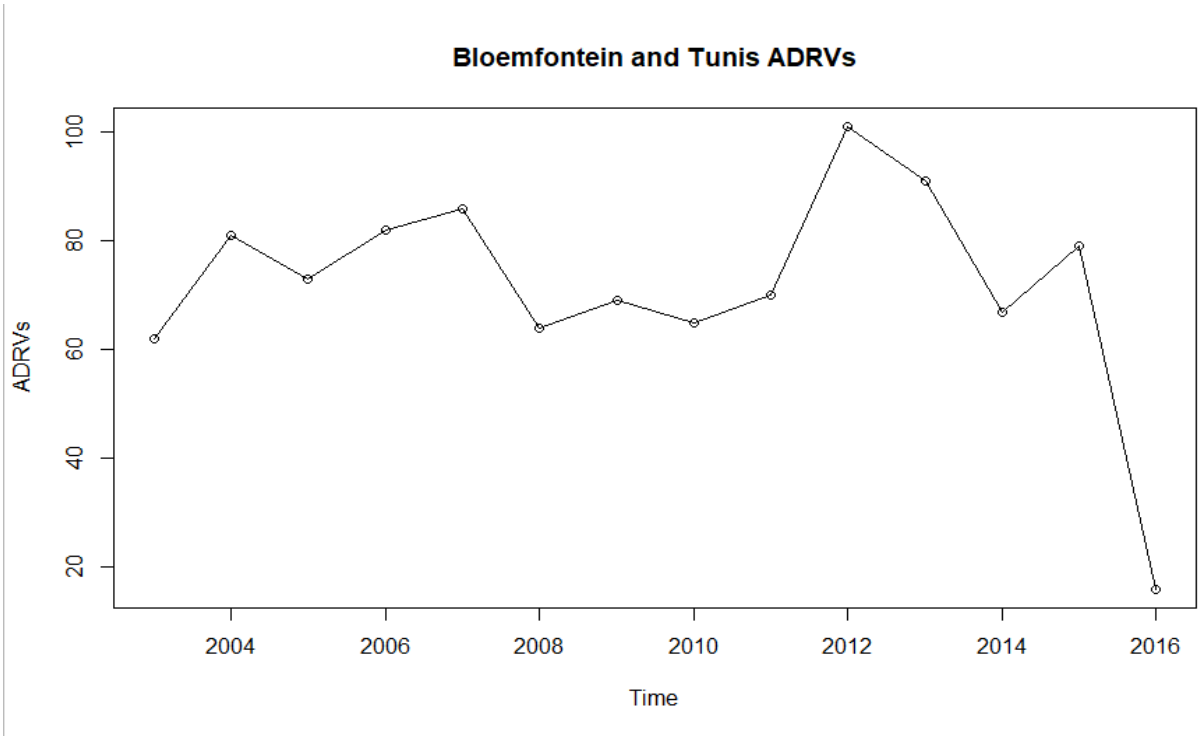
4.1 Regional and global doping trends

The first specific objective was to establish the regional doping trend for 14 years (2003-2016) and compare it with global trends. Each plot is followed by Box-Jenkins procedure. This involves model identification and estimation where one model is selected for further analysis. This model is then validated using a residuals plot and Box-Ljung test. This is followed by forecasting of ADRV values of the next 5 years.

4.1.1 Trend of Bloemfontein and Tunis ADRVs

The data from Bloemfontein and Tunis laboratories indicated a decreasing trend. The cyclic aspect repeats after two years on three occasions, three years and four years on one occasion. This is illustrated in Figure 4.1 below.

Figure 4. 1: Plot of Bloemfontein and Tunis ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.2 and Figure 4.3 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The output showed the possible ARIMA models and their respective AIC values. The model with the lowest AIC value was ARIMA(0,2,1) and was chosen as the most appropriate model. This is illustrated in Table 4.1 below.

Figure 4. 2: ACF Plot of Bloemfontein and Tunis ADRVs

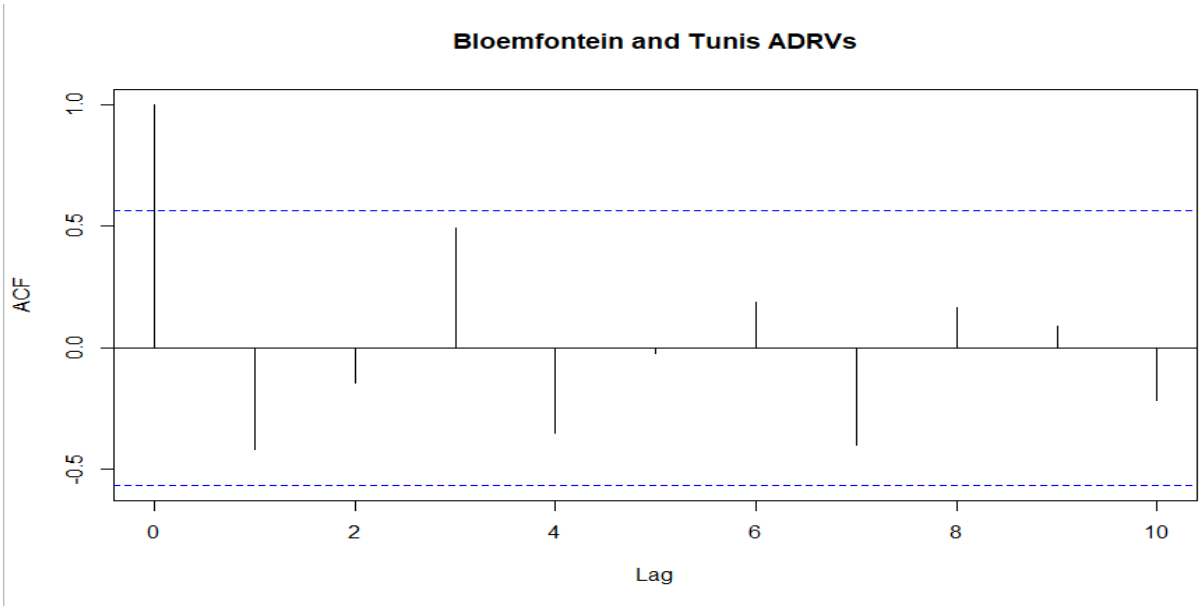


Figure 4. 3: PACF Plot of Bloemfontein and Tunis ADRVs

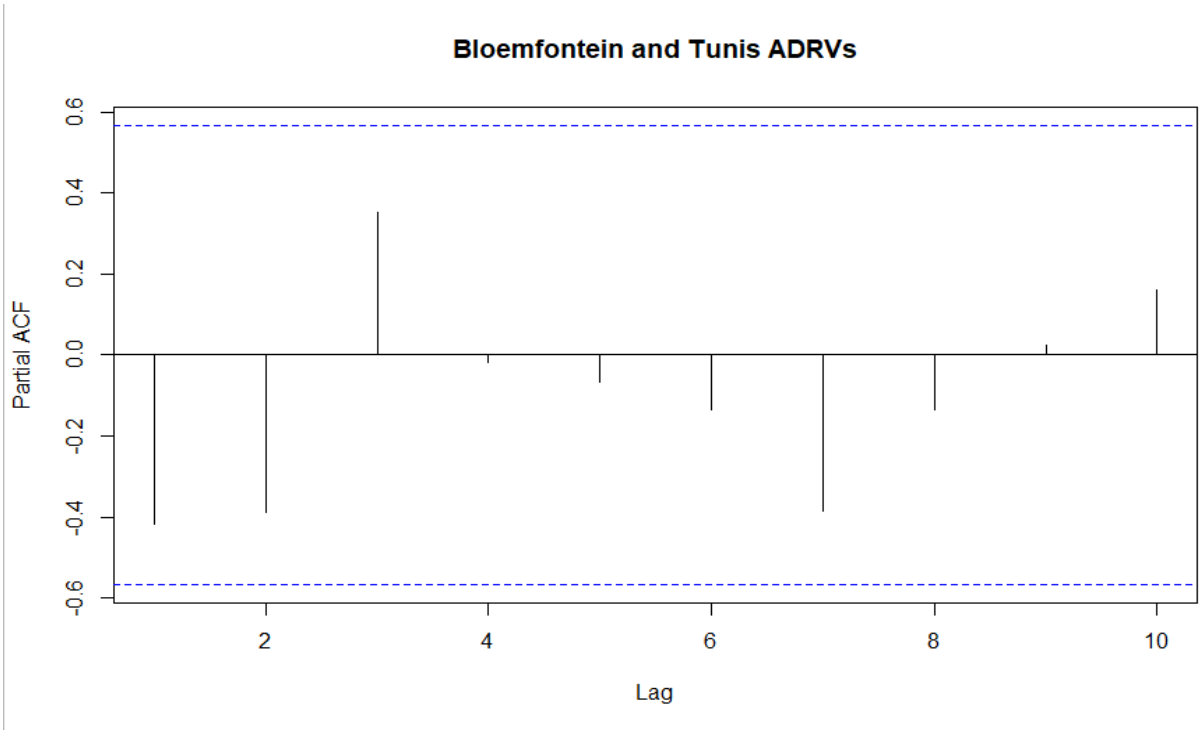


Table 4. 1: ARIMA models of Bloemfontein and Tunis ADRVs

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(0,2,0)	31.56149	
ARIMA(1,2,0)	29.76763	Best Model
ARIMA(0,2,1)	Infinite	
ARIMA(2,2,0)	31.76741	

Diagnostics and validation

The chosen model ARIMA(0,2,1) was validated using a residuals plot as illustrated in Figure 4.4 below. The plot vibrated around zero meaning the model was acceptable. This was further confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.2 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 4: Residuals Plot of Bloemfontein and Tunis ADRVs

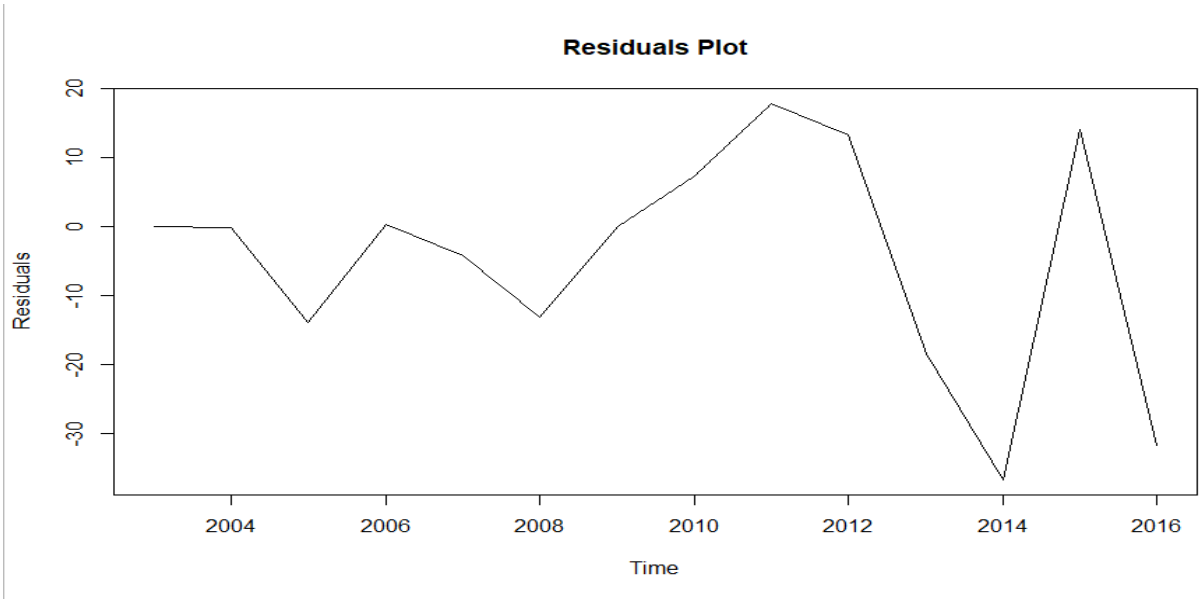


Table 4. 2: Box-Ljung test of Bloemfontein and Tunis ADRVs

Lags	P-value	Inference
5	6379	Not significant
10	0.7633	Not significant

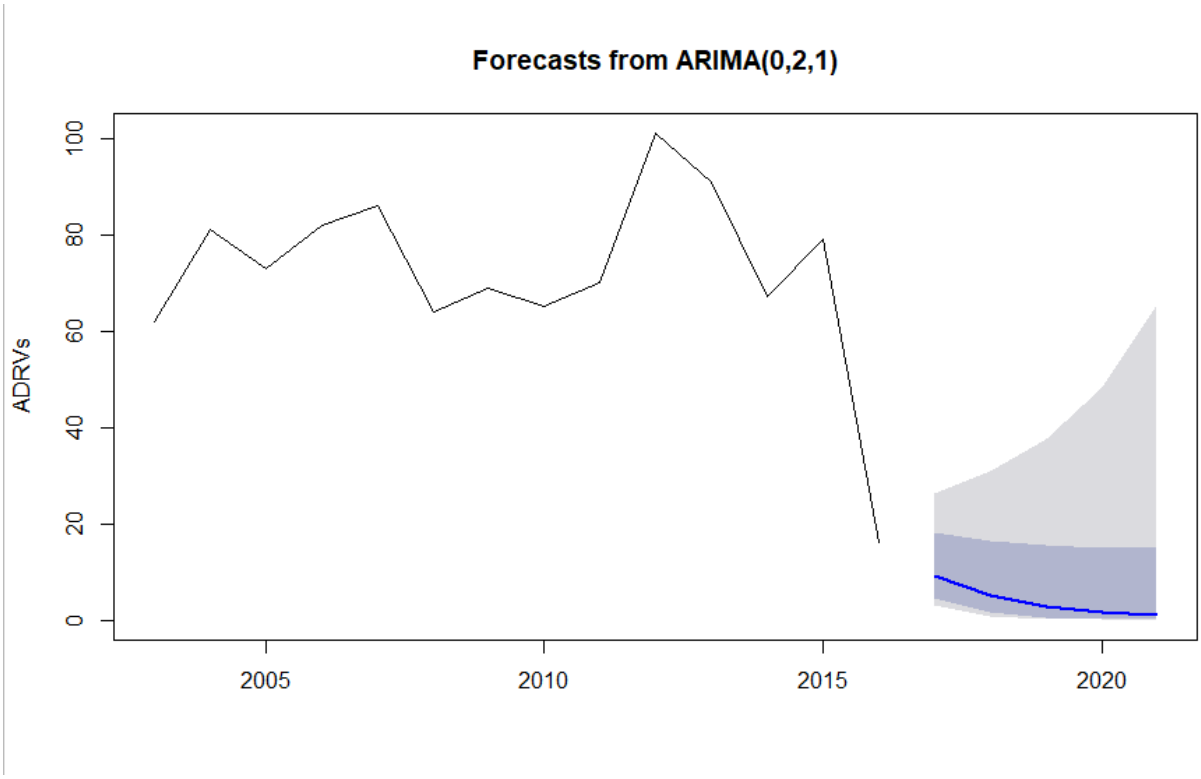
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.3. The data was also plotted as indicated in Figure 4.5. The predicted number of ADRVs fell gradually from 9 in 2017 to 1 ADRV in 2021.

Table 4. 3: Five years forecasts of Bloemfontein and Tunis ADRVs

Year	Forecast value
2017	9
2018	5
2019	3
2020	2
2021	1

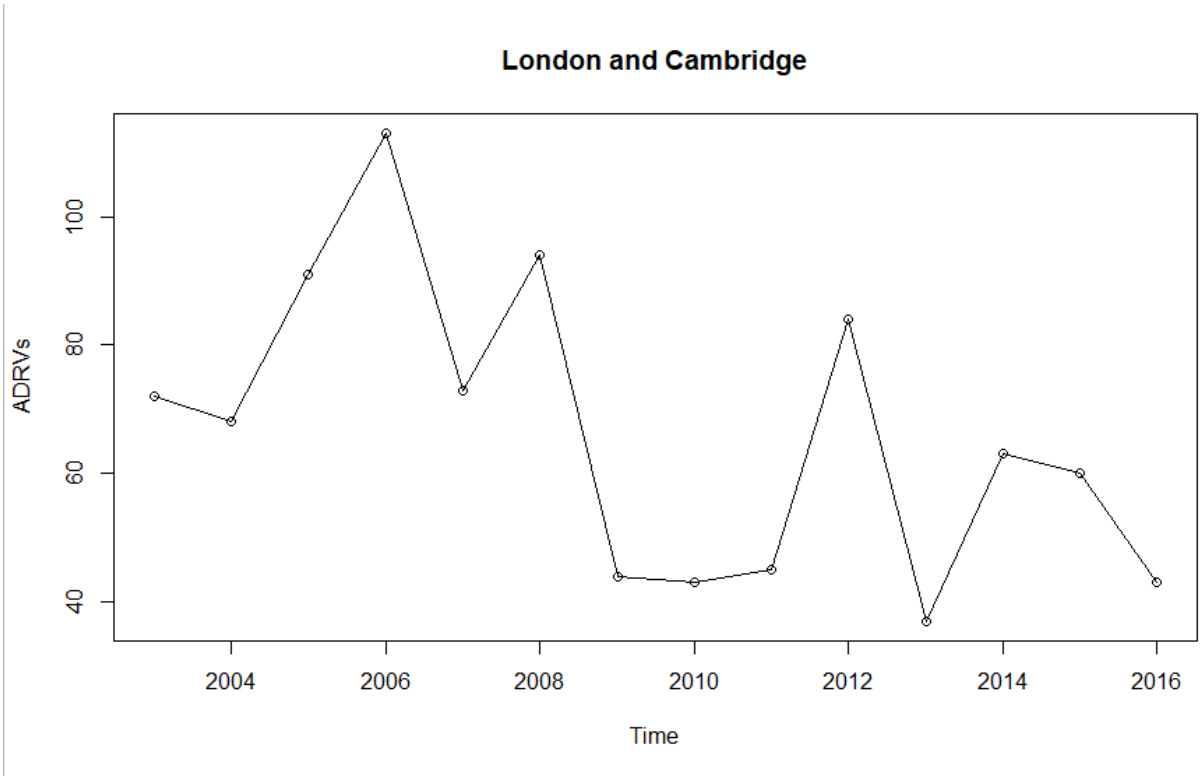
Figure 4. 5: Forecasts of Bloemfontein and Tunis ADRVs



4.1.2 Trend of London and Cambridge ADRVs

The data from London and Cambridge laboratories indicated a decreasing trend. The cyclic component repeated after three years. The pattern was irregular as illustrated Figure 4.6 below.

Figure 4. 6: Plot of London and Cambridge ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.7 and Figure 4.8 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(0,1,1) and was chosen as the most appropriate model. This is illustrated in Table 4.4 below.

Figure 4. 7: ACF Plot of London and Cambridge ADRVs

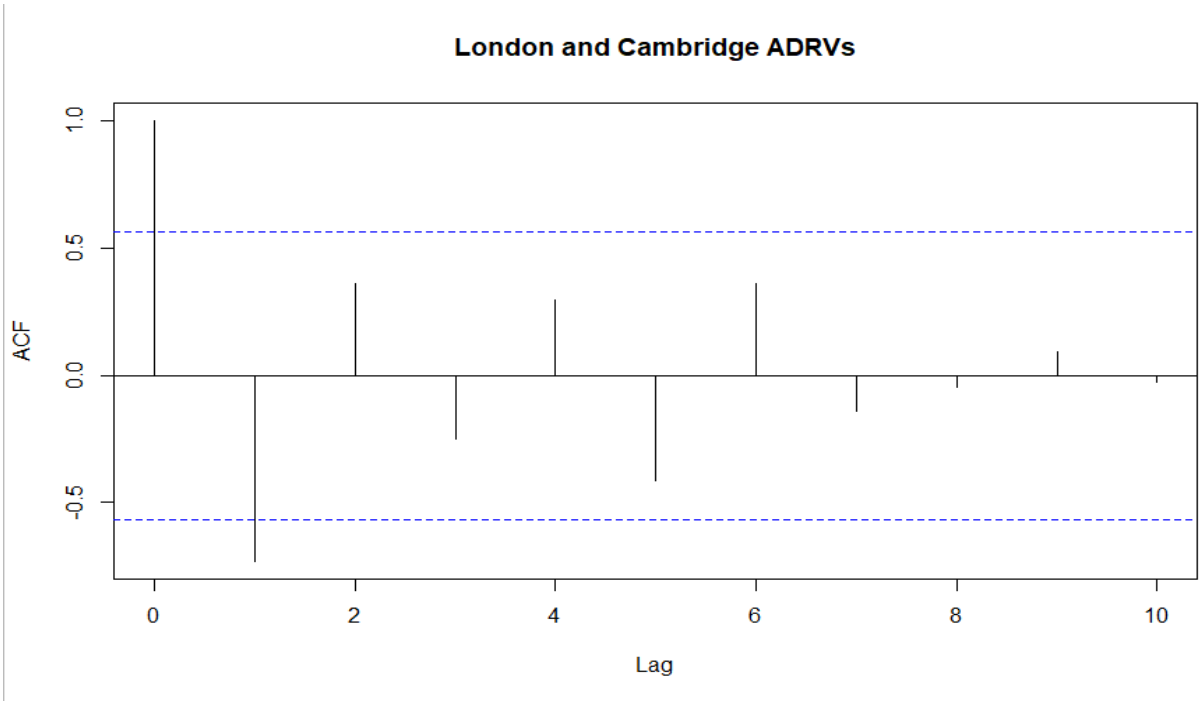


Figure 4. 8: PACF Plot of London and Cambridge ADRVs

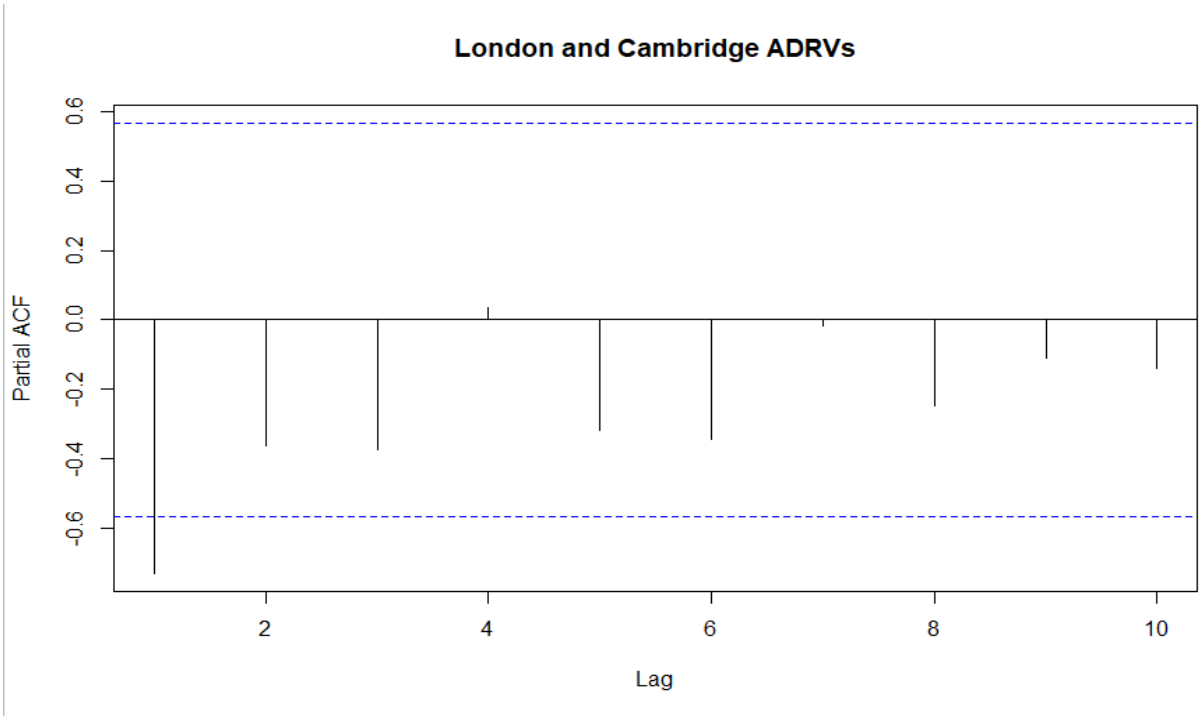


Table 4. 4: ARIMA models of London and Cambridge ADRVs

Model type	AIC	Inference
ARIMA(1,1,1)	16.18169	
ARIMA(0,1,1)	14.21715	Best Model
ARIMA(0,1,2)	16.19265	
ARIMA(1,1,2)	18.01143	
ARIMA(2,1,0)	16.56524	

Diagnostics and validation

The chosen model ARIMA(0,1,1) was validated using a residuals plot as illustrated in Figure 4.9 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.5 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 9: Residuals Plot of London and Cambridge ADRVs

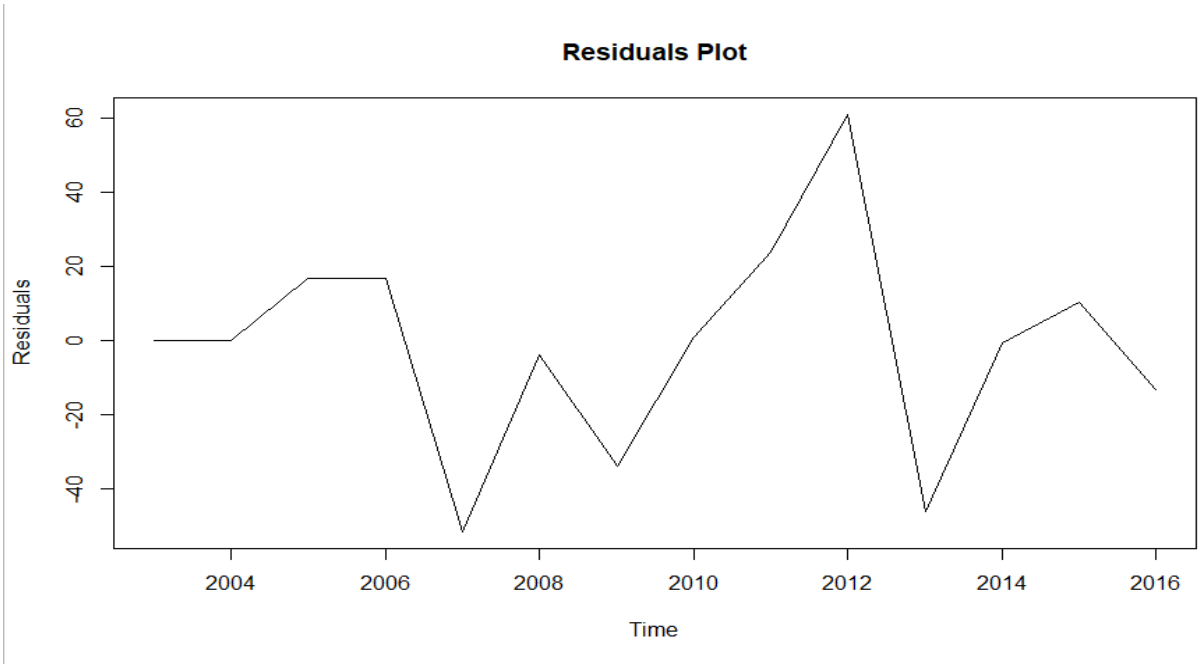


Table 4. 5: Box-Ljung test of London and Cambridge ADRVs

Lags	P-value	Inference
5	0.5283	Not significant
10	0.8301	Not significant

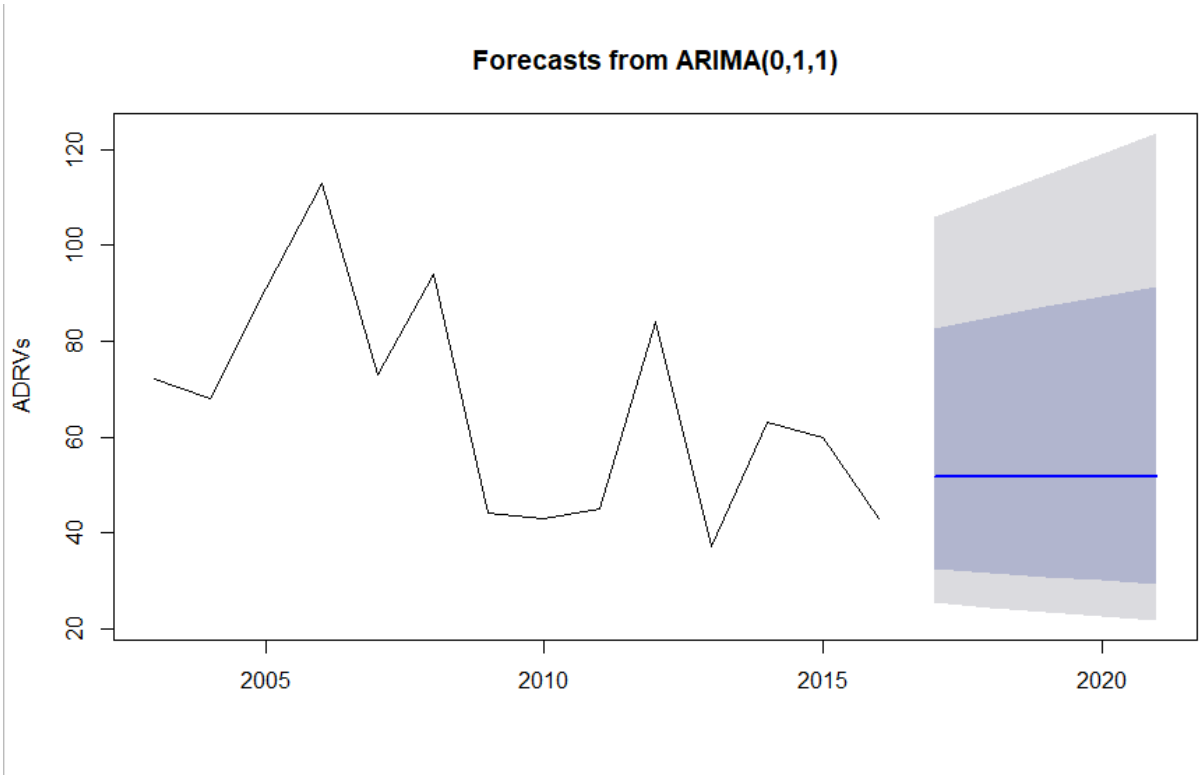
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.6 below. The data was also plotted as indicated in Figure 4.10. The predicted number of ADRVs remained 52 throughout the 5 year period.

Table 4. 6: Forecasting of London and Cambridge ADRVs

Year	Forecast value
2017	52
2018	52
2019	52
2020	52
2021	52

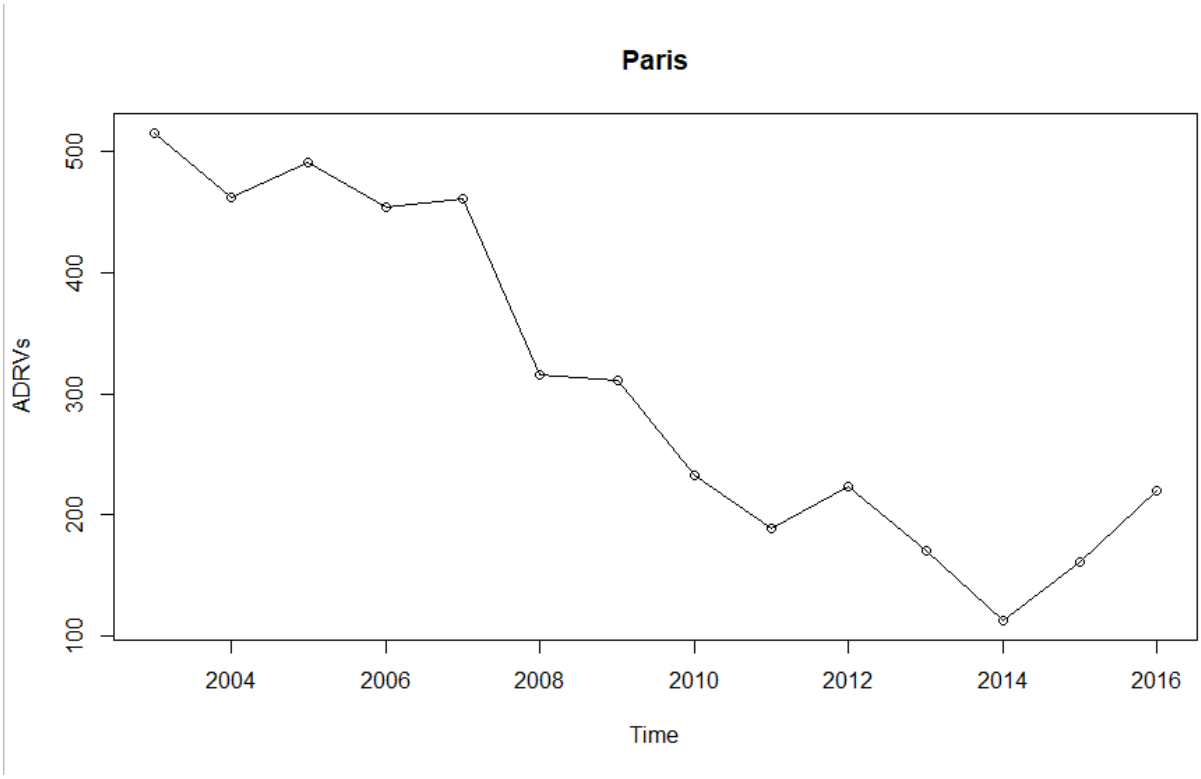
Figure 4. 10: Forecast of London and Cambridge ADRVs



4.1.3 Trend of Paris ADRVs

The data from Paris laboratory indicated a decreasing trend. The cyclic component repeated after two years. This is illustrated in Figure 4.11 below.

Figure 4. 11: Plot of Paris ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.12 and Figure 4.13 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(0,1,0) and was chosen as the most appropriate model. This is illustrated in Table 4.7 below.

Figure 4. 12: ACF Plot of Paris ADRVs

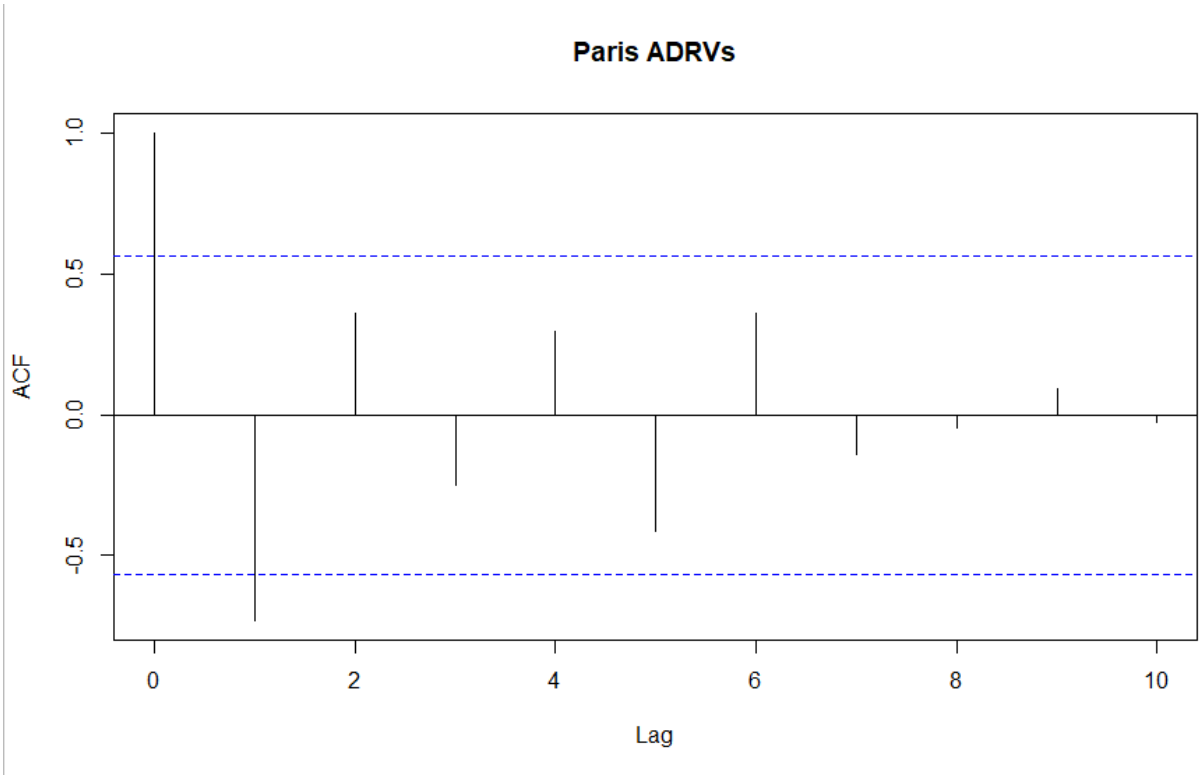


Figure 4. 13: PACF Plot of Paris ADRVs

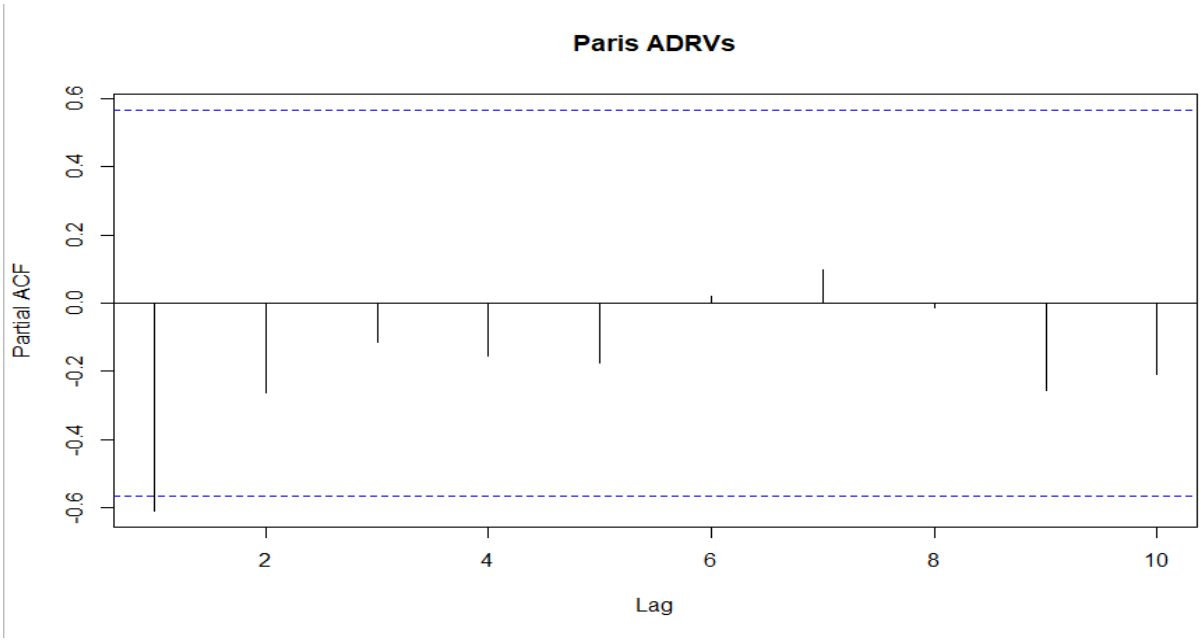


Table 4. 7: ARIMA models of Paris ADRVs

Model type	AIC	Inference
ARIMA(2,1,2)	Infinite	
ARIMA(1,1,0)	5.386184	
ARIMA(0,1,1)	5.326848	
ARIMA(0,1,0)	2.360534	Best model
ARIMA(1,1,1)	Infinite	

Diagnostics and validation

The chosen model ARIMA(0,1,0) was validated using a residuals plot as illustrated in Figure 4.14 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.8 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 14: Residuals Plot of Paris ADRVs

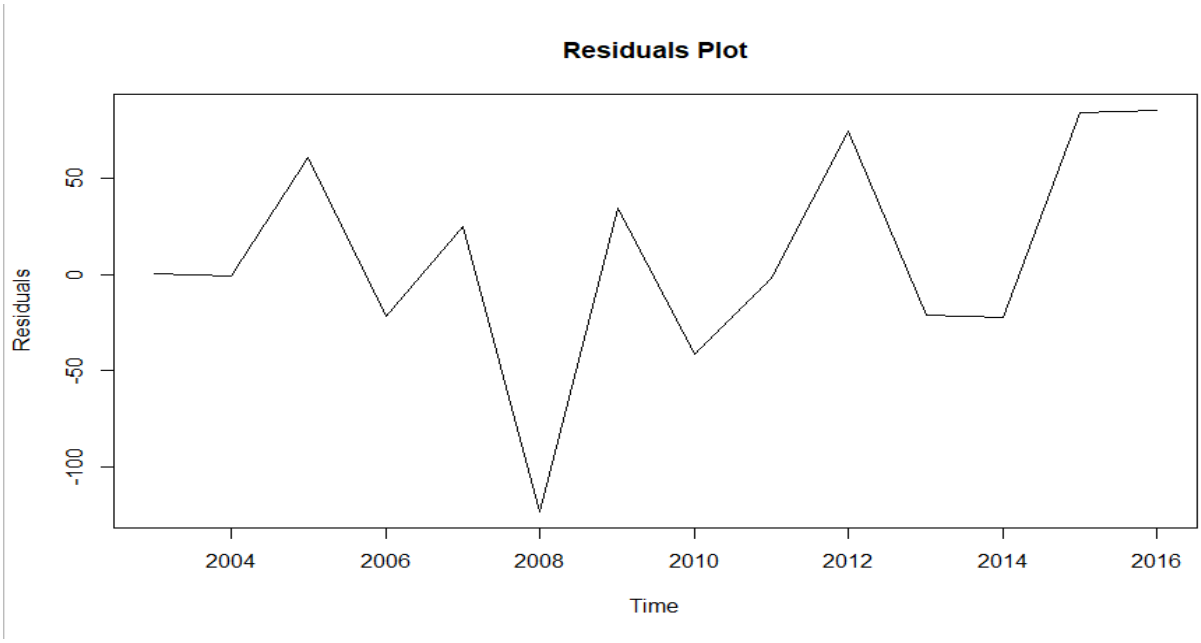


Table 4. 8: Box-Ljung test of Paris ADRVs

Lags	P-value	Inference
5	0.6898	Not significant
10	0.8724	Not significant

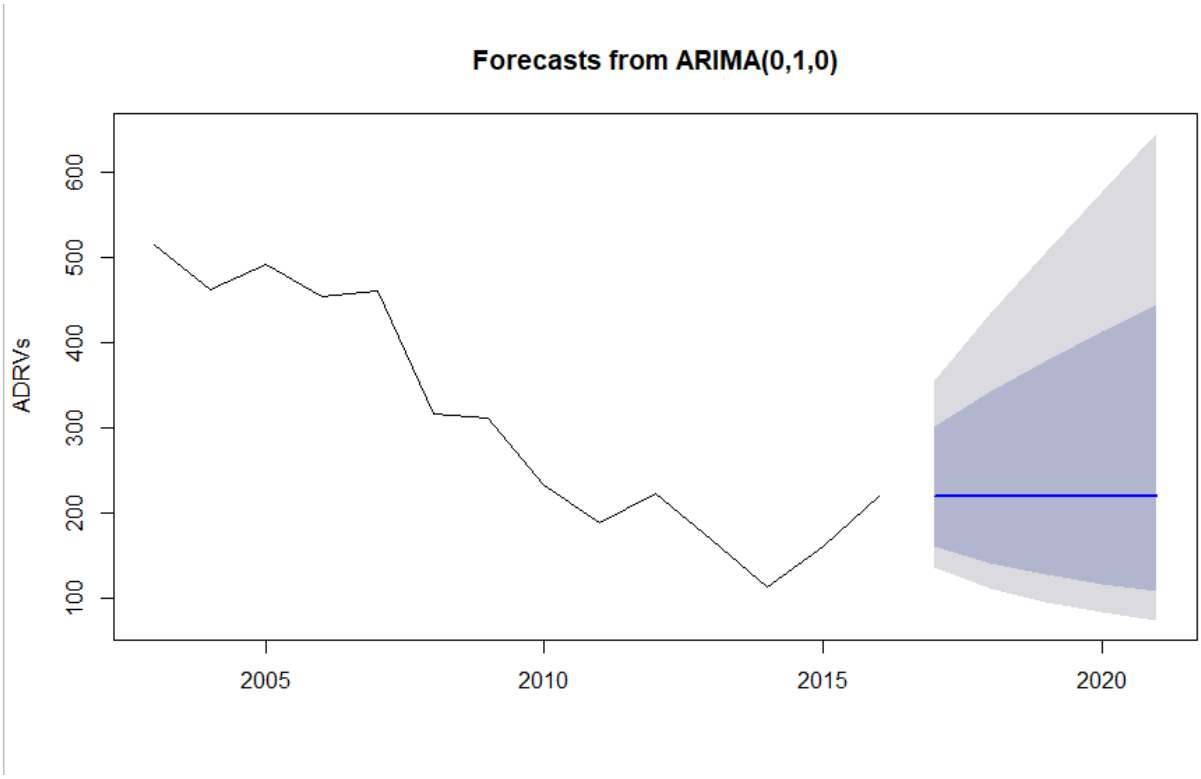
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.9. The data was also plotted as indicated in Figure 4.15. The predicted number of ADRVs remained 220 throughout the 5 year period.

Table 4. 9: Forecasting of Paris ADRVs

Year	Forecast value
2017	220
2018	220
2019	220
2020	220
2021	220

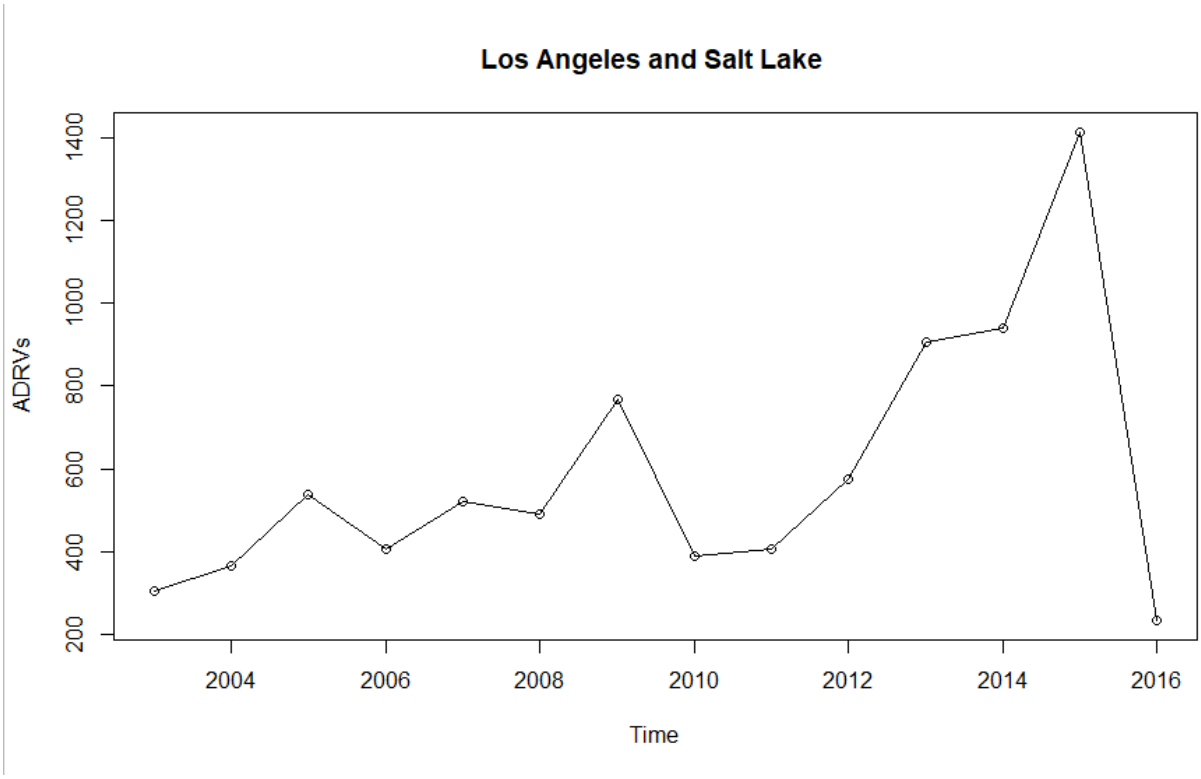
Figure 4. 15: Forecasts of Paris ADRVs



4.1.4 Trend of Los Angeles and Salt Lake ADRVs

The data from Los Angeles and Salt Lake laboratories indicated an increasing trend. The cyclic aspect repeated after two years. This is illustrated in Figure 4.1 below.

Figure 4. 16: Plot of Los Angeles and Salt Lake ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.17 and Figure 4.18 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(1,2,0) and was chosen as the most appropriate model. This is illustrated in Table 4.10 below.

Figure 4. 17: ACF Plot of Los Angeles and Salt Lake ADRVs

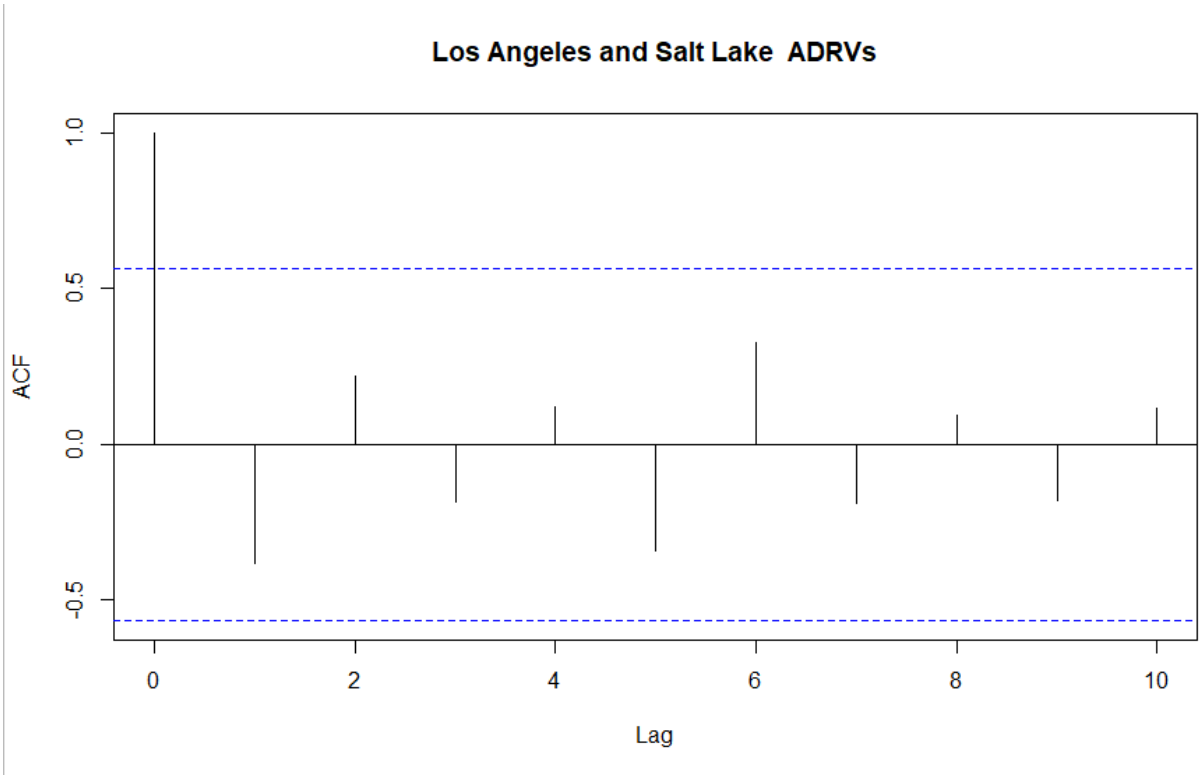


Figure 4. 18: PACF Plot of Los Angeles and Salt Lake ADRVs

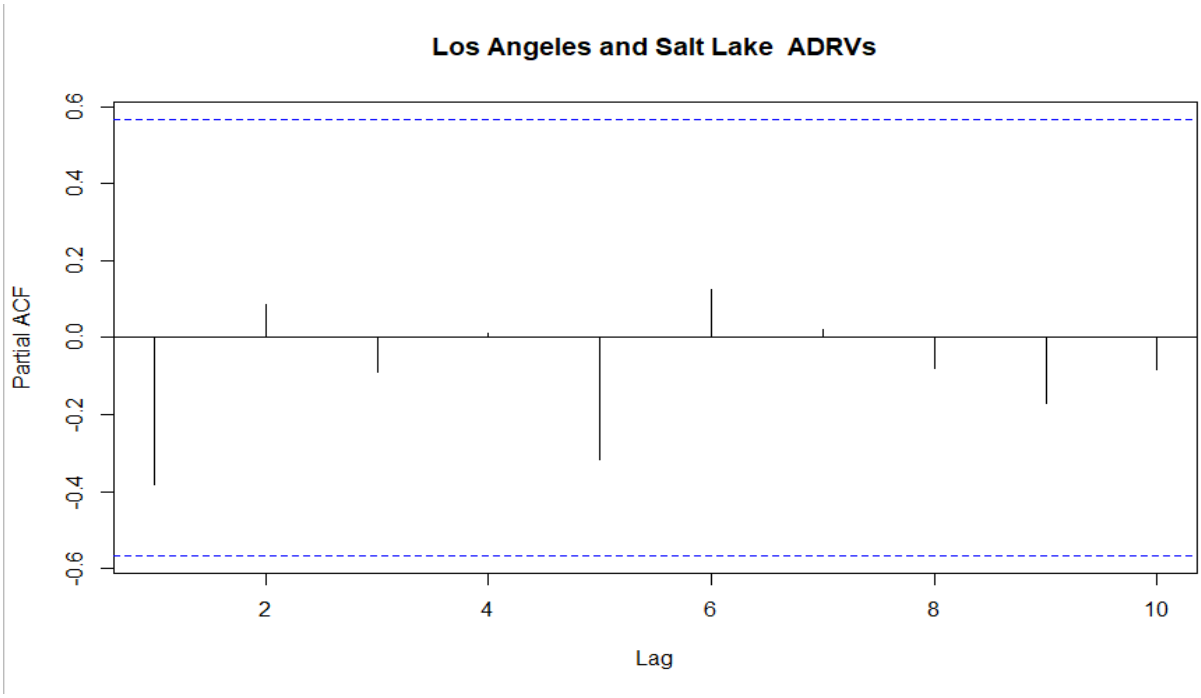


Table 4. 10: ARIMA models of Los Angeles and Salt Lake ADRVs

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(0,2,0)	31.56149	
ARIMA(1,2,0)	29.76763	Best model
ARIMA(0,2,1)	Infinite	
ARIMA(1,2,1)	Infinite	

Diagnostics and validation

The chosen model ARIMA(1,2,0) was validated using a residuals plot as illustrated in Figure 4.19 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.11 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 19: Residuals Plot of Los Angeles and Salt Lake ADRVs

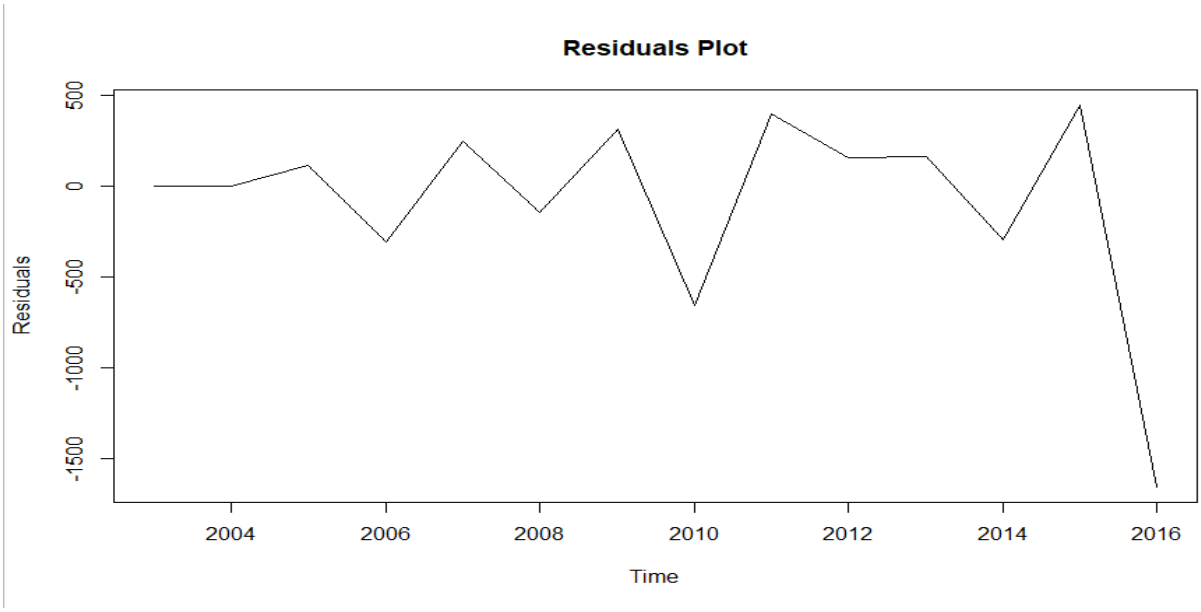


Table 4. 11: Box-Ljung test of Los Angeles and Salt Lake ADRVs

Lags	P-value	Inference
5	0.9604	Not significant
10	0.9539	Not significant

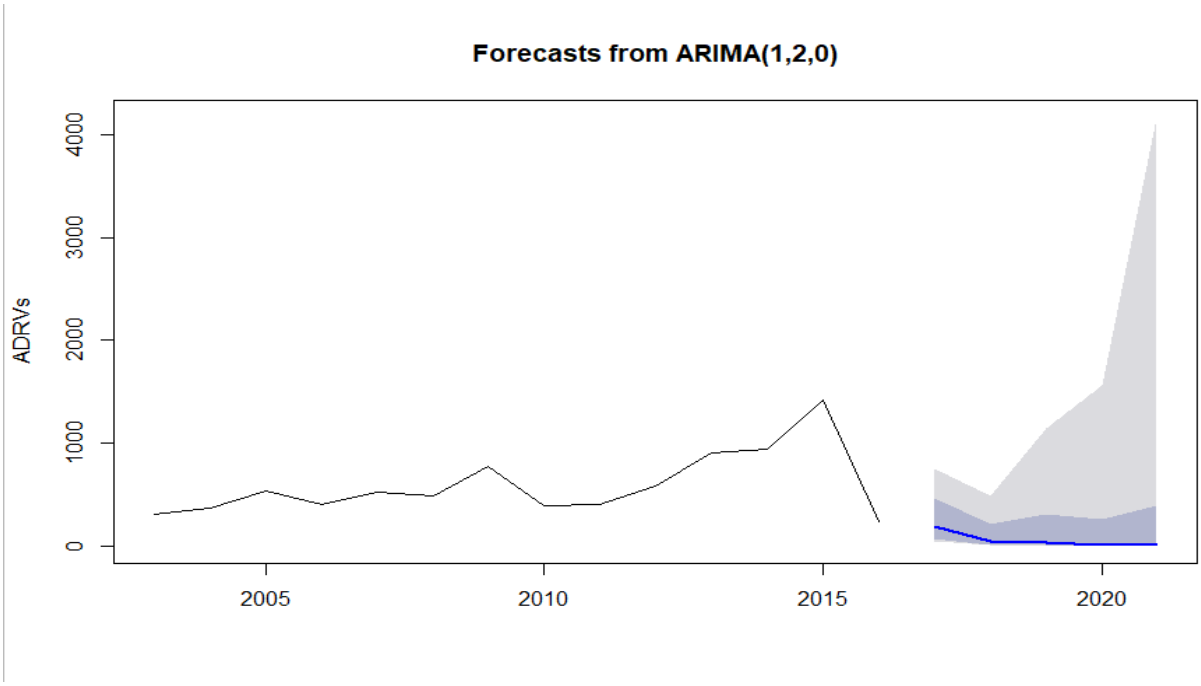
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.3. The data was also plotted as indicated in Figure 4.5. The predicted number of ADRVs fell from 183 in 2017 to 4 ADRVs in 2021.

Table 4. 12: Forecasting of Los Angeles and Salt Lake ADRVs

Year	Forecast value
2017	183
2018	48
2019	27
2020	9
2021	4

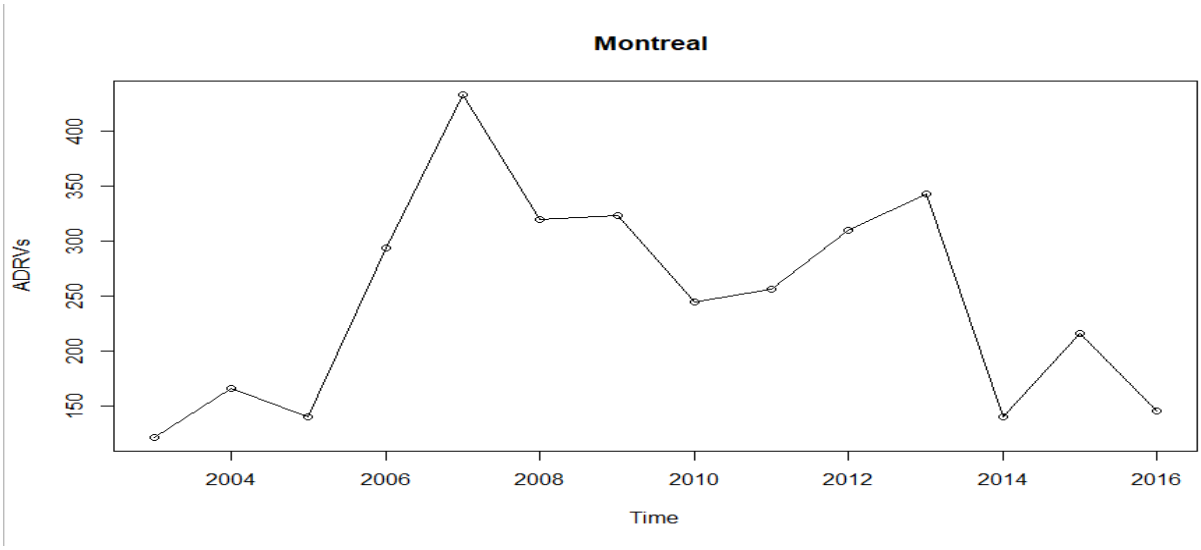
Figure 4. 20: Forecasts of Los Angeles and Salt Lake ADRVs



4.1.5 Trend of Montreal ADRVs

The data from Montreal laboratory indicated a marginally increasing trend. The cyclic aspect repeats in a blend of two and three years. This is illustrated in Figure 4.21 below.

Figure 4. 21: Plot of Montreal ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.22 and Figure 4.23 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(0,2,1) and was chosen as the most appropriate model. This is illustrated in Table 4.13 below.

Figure 4. 22: ACF Plot of Montreal ADRVs

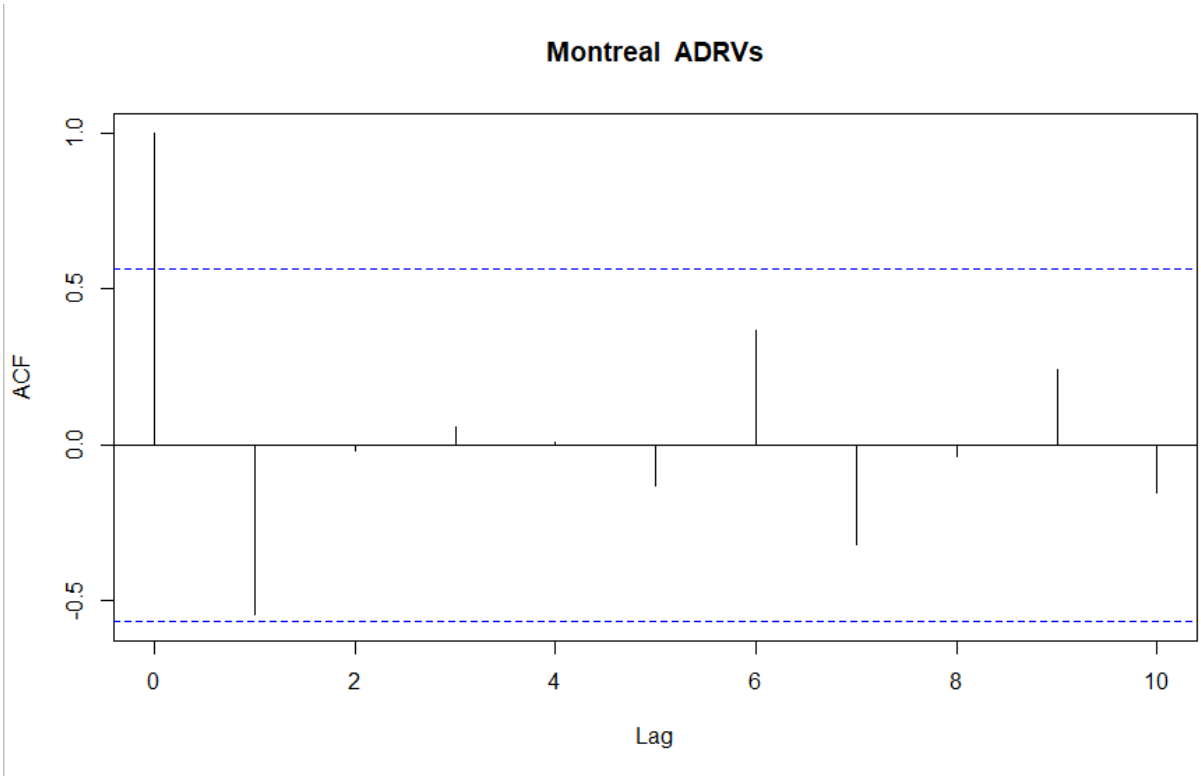


Figure 4. 23: PACF Plot of Montreal ADRVs

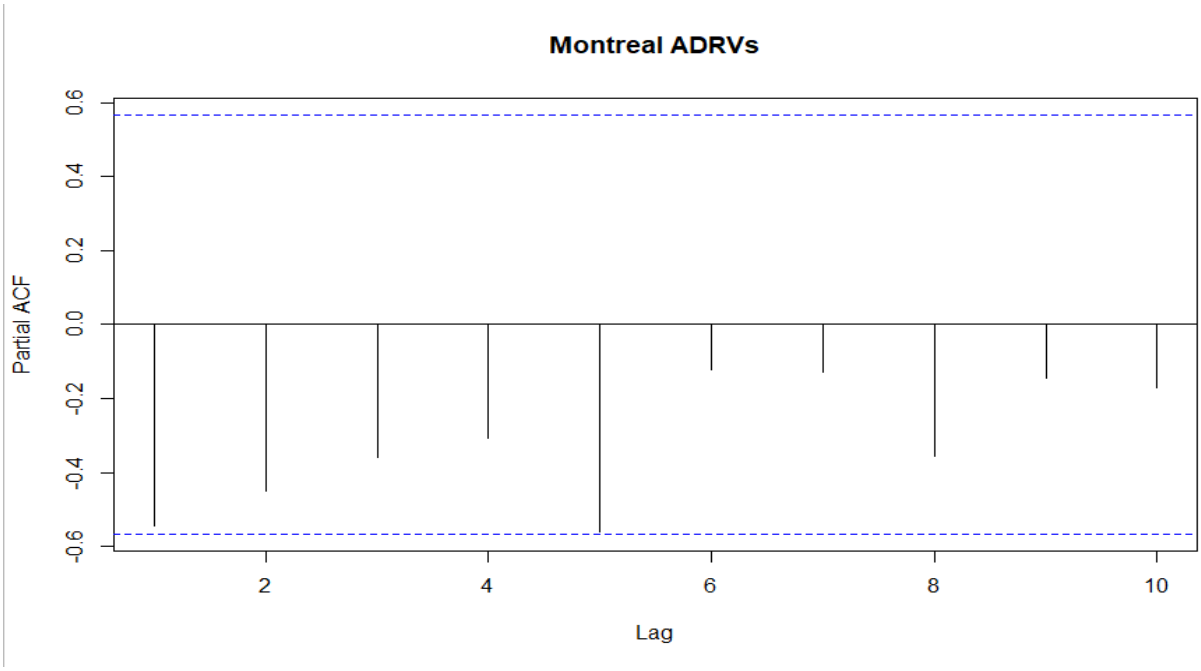


Table 4. 13: ARIMA models of Montreal ADRVs

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(2,2,0)	26.51135	
ARIMA(0,2,1)	20.1272	Best model
ARIMA(1,2,1)	20.96767	
ARIMA(0,2,2)	20.45785	
ARIMA(1,2,2)	Infinite	

Diagnostics and validation

The chosen model ARIMA(0,2,1) was validated using a residuals plot as illustrated in Figure 4.24 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.14 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 24: Residuals Plot of Montreal ADRVs

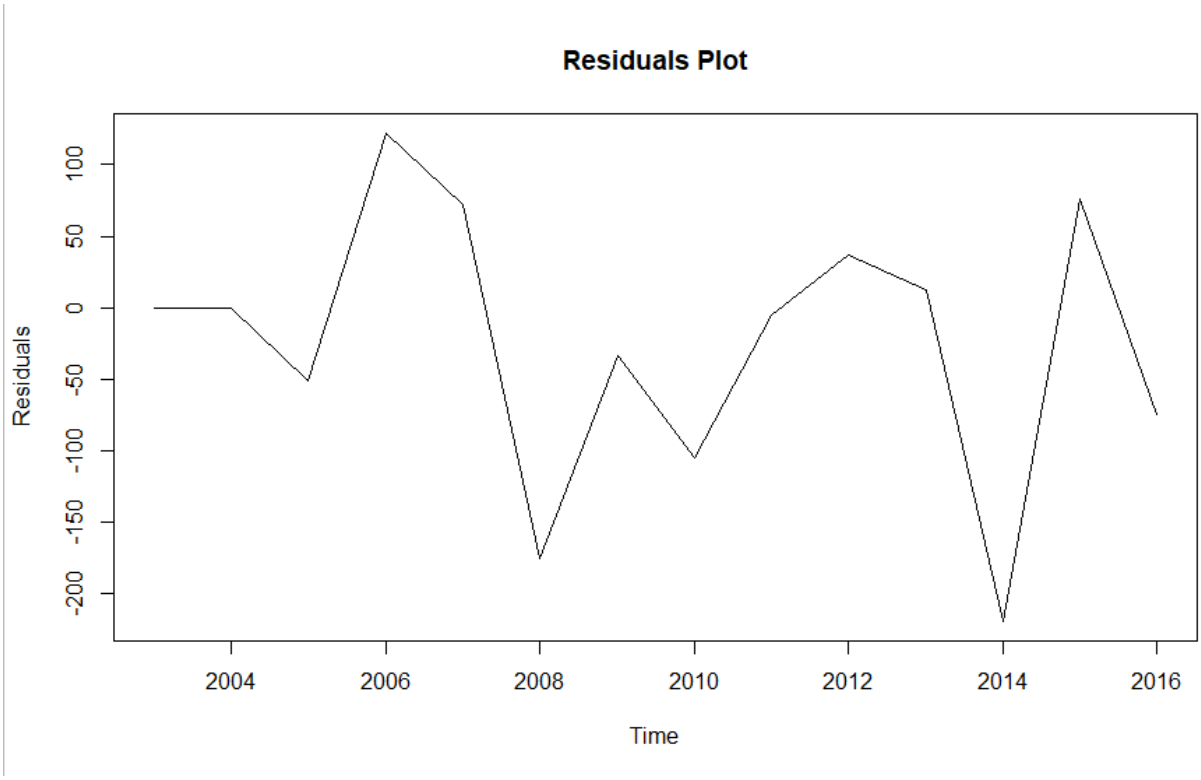


Table 4. 14: Box-Ljung test of Montreal ADRVs

Lags	P-value	Inference
5	0.7196	Not significant
10	0.2419	Not significant

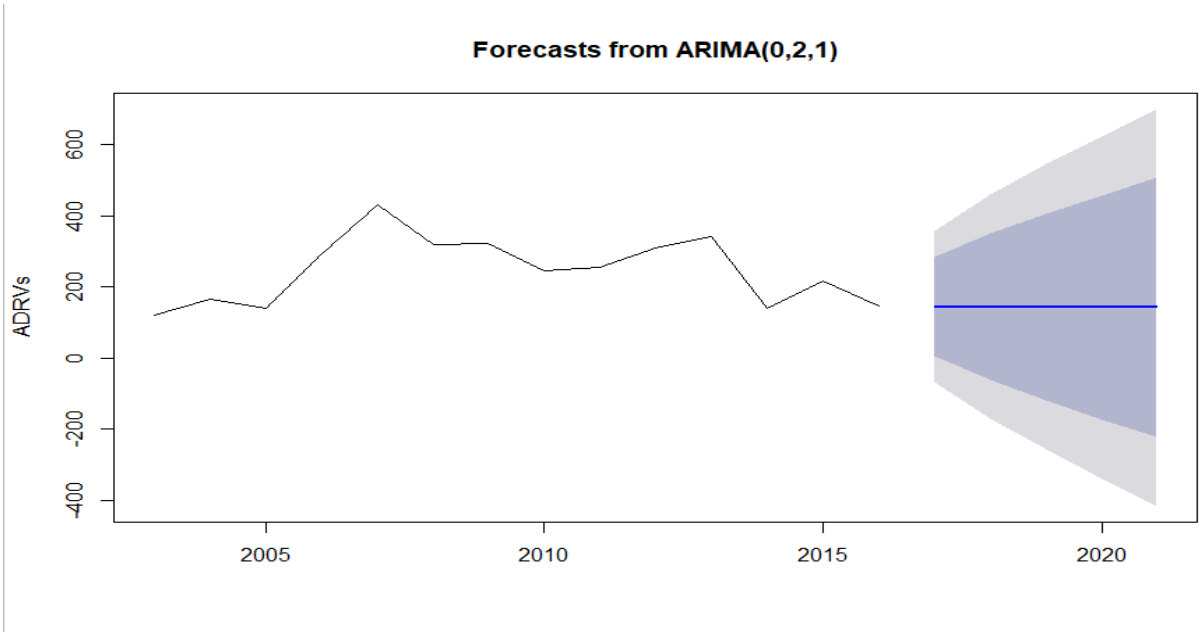
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.15. The data was also plotted as indicated in Figure 4.25. The predicted number of ADRVs fell slightly from 142 in 2017 to 126 ADRVs in 2021.

Table 4. 15: Forecasting of Montreal ADRVs

Year	Forecast value
2017	142
2018	138
2019	133
2020	129
2021	126

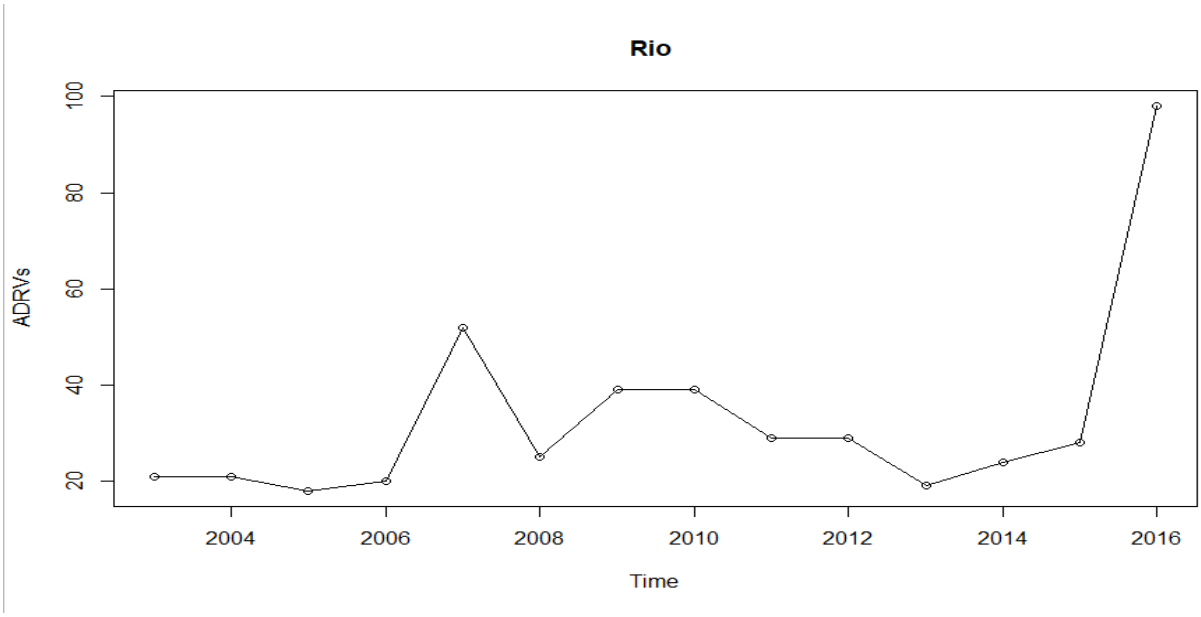
Figure 4. 25: Forecasts of Montreal ADRVs



4.1.6 Trend of Rio ADRVs

The data from Rio laboratory indicated an increasing trend. The cyclic aspect repeated in a blend of two and three years. This is illustrated in Figure 4.26 below.

Figure 4. 26: Plot of Rio ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.27 and Figure 4.28 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(1,2,0) and was chosen as the most appropriate model. This is illustrated in Table 4.16 below.

Figure 4. 27: ACF Plot of Rio ADRVs

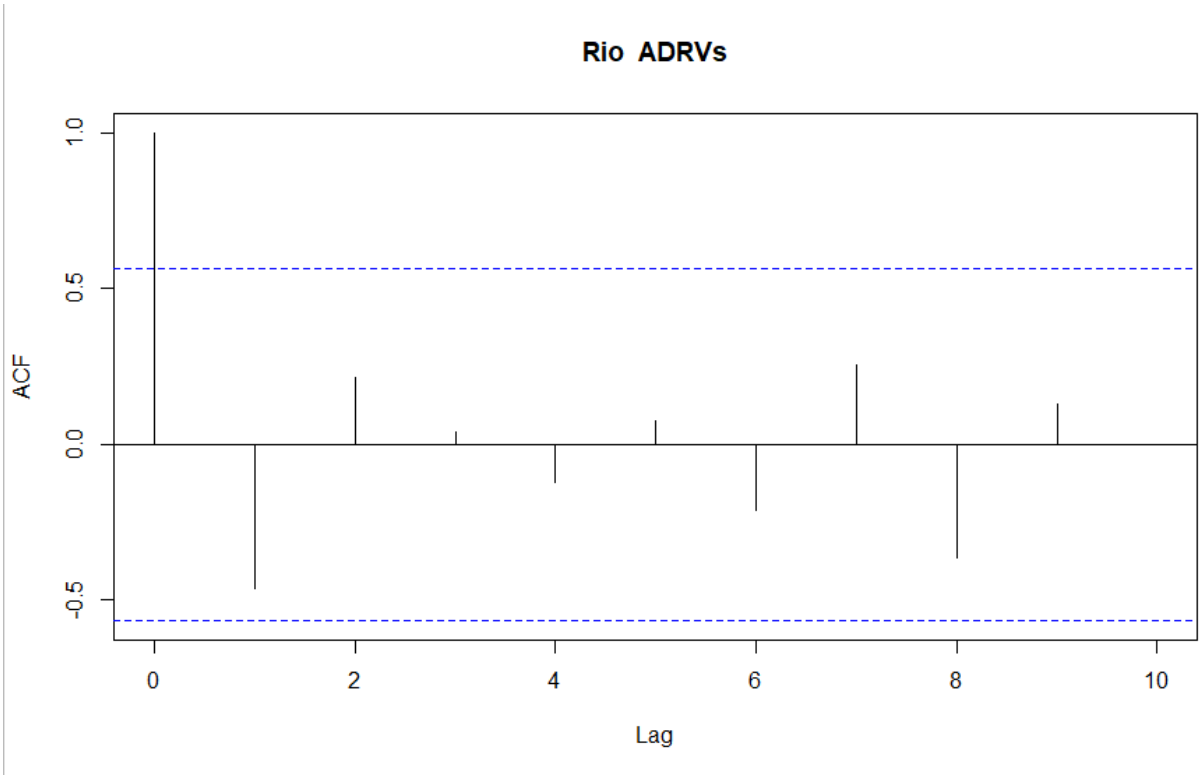


Figure 4. 28: PACF Plot of Rio ADRVs

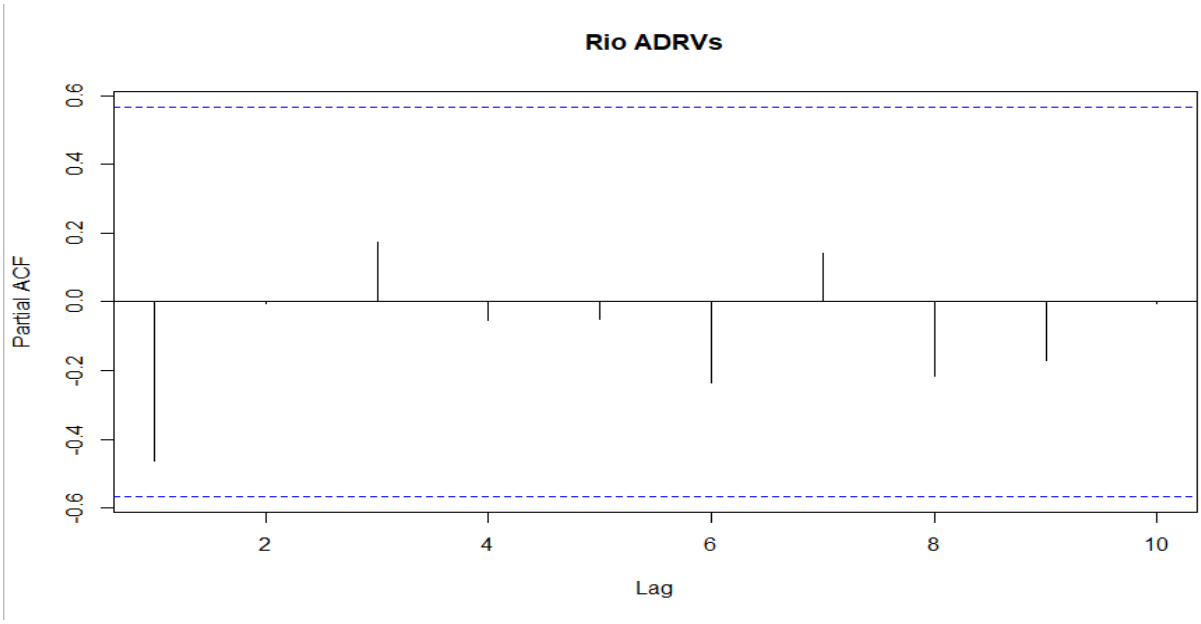


Table 4. 16: ARIMA models of Rio ADRVs

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(0,2,0)	29.99868	
ARIMA(1,2,0)	26.12929	Best model
ARIMA(0,2,1)	Infinite	
ARIMA(2,2,0)	27.10327	
ARIMA(1,2,1)	Infinite	

Diagnostics and validation

The chosen model ARIMA(1,2,0) was validated using a residuals plot as illustrated in Figure 4.29 below. The plot vibrated around zero meaning the model was acceptable. This was farther

confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.17 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 29: Residuals Plot of Rio ADRVs

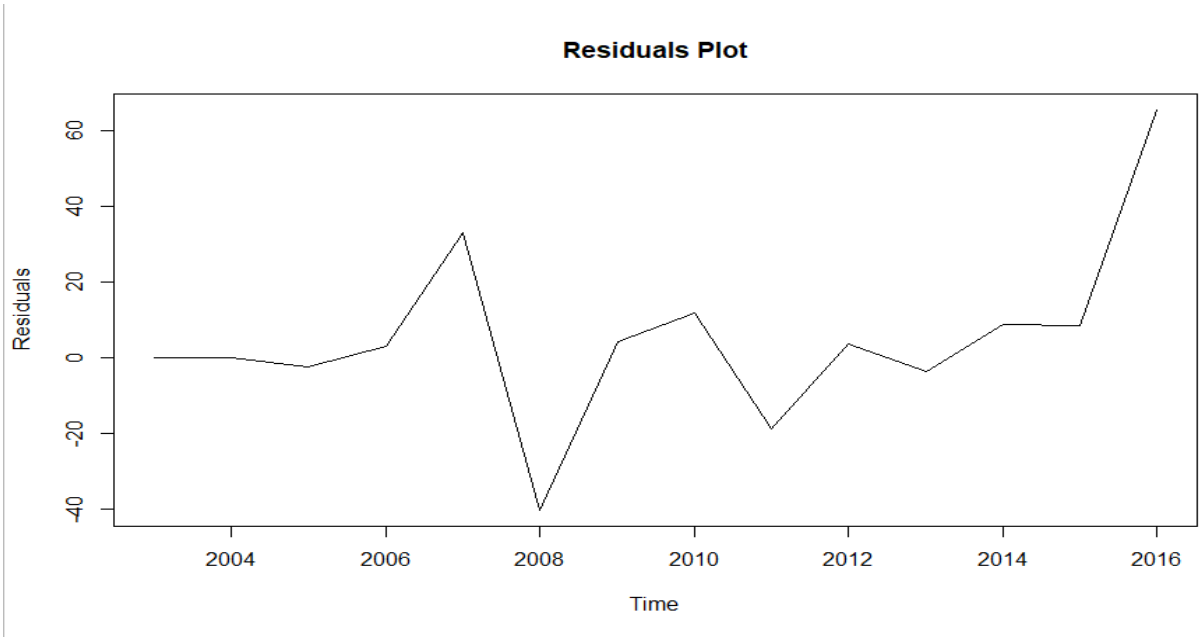


Table 4. 17: Box-Ljung test of Rio ADRVs

Lags	P-value	Inference
5	0.8279	Not significant
10	0.8414	Not significant

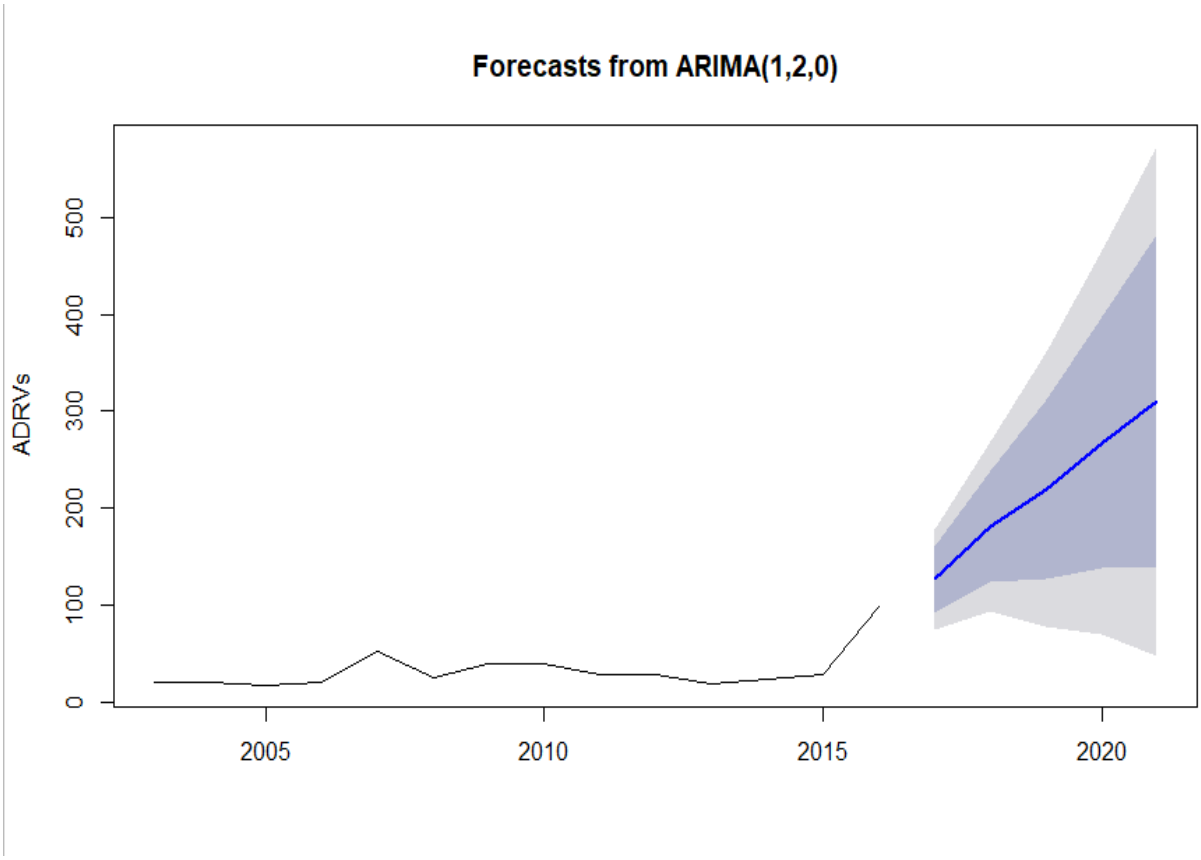
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.18. The data was also plotted as indicated in Figure 4.30. The predicted number of ADRVs rose gradually from 127 in 2017 to 310 ADRVs in 2021.

Table 4. 18: Forecasting of Rio ADRVs

Year	Forecast value
2017	127
2018	181
2019	220
2020	268
2021	310

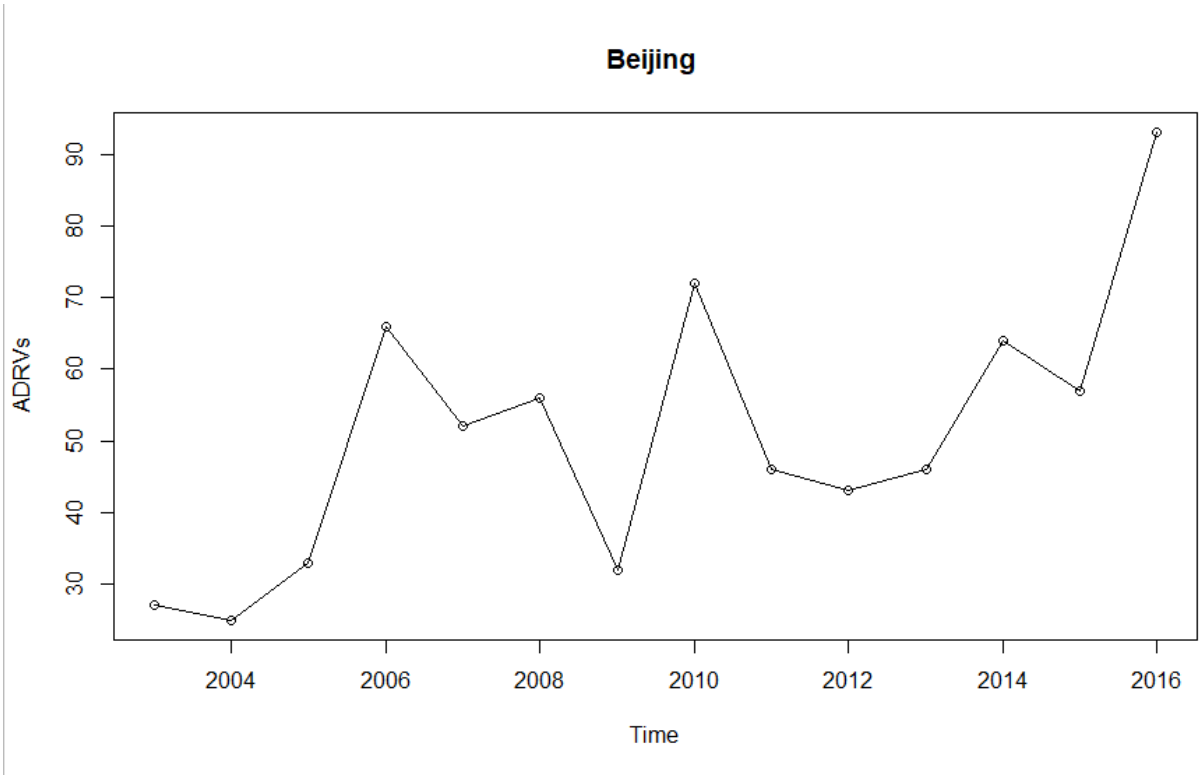
Figure 4. 30: Forecasts of Plot Rio ADRVs



4.1.7 Trend of Beijing ADRVs

The data from Beijing laboratory indicated an increasing trend. The cyclic aspect repeated after three years. This is illustrated in Figure 4.1 below.

Figure 4. 31: Plot of Beijing ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.32 and Figure 4.33 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(1,2,0) and was chosen as the most appropriate model. This is illustrated in Table 4.19 below.

Figure 4. 32: ACF Plot of Beijing ADRVs

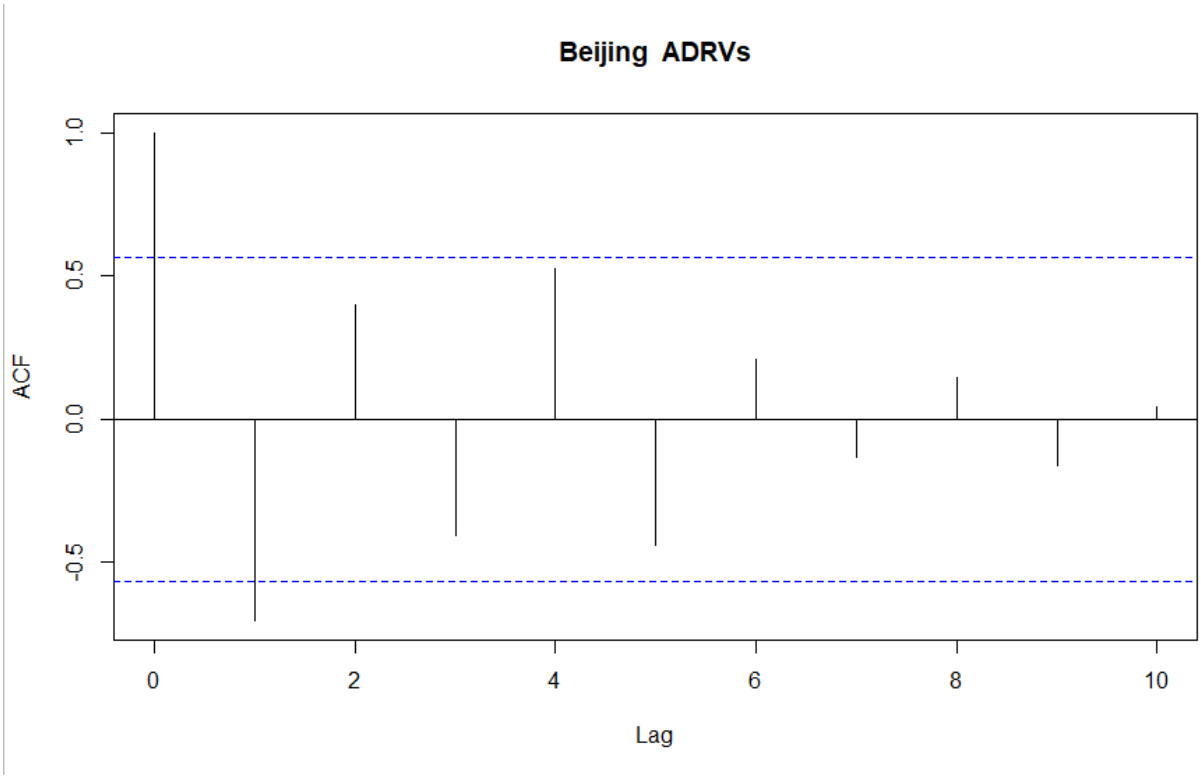


Figure 4. 33: PACF Plot of Beijing ADRVs

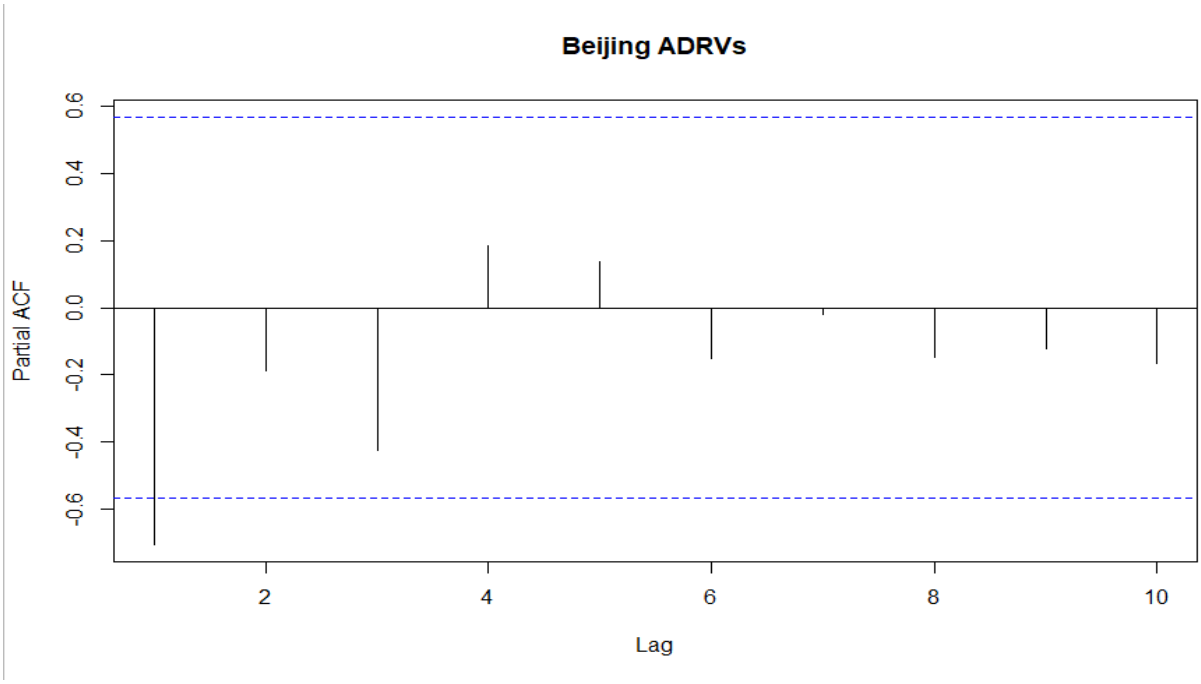


Table 4. 19: ARIMA models of Beijing ADRVs

Model type	AIC	Inference
ARIMA(0,2,0)	27.5886	
ARIMA(1,2,0)	21.45768	Best model
ARIMA(2,2,0)	22.27818	
ARIMA(1,2,1)	Infinite	
ARIMA(2,2,2)	Infinite	

Diagnostics and validation

The chosen model ARIMA(1,2,0) was validated using a residuals plot as illustrated in Figure 4.34 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.20 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 34: Residuals Plot of Beijing ADRVs

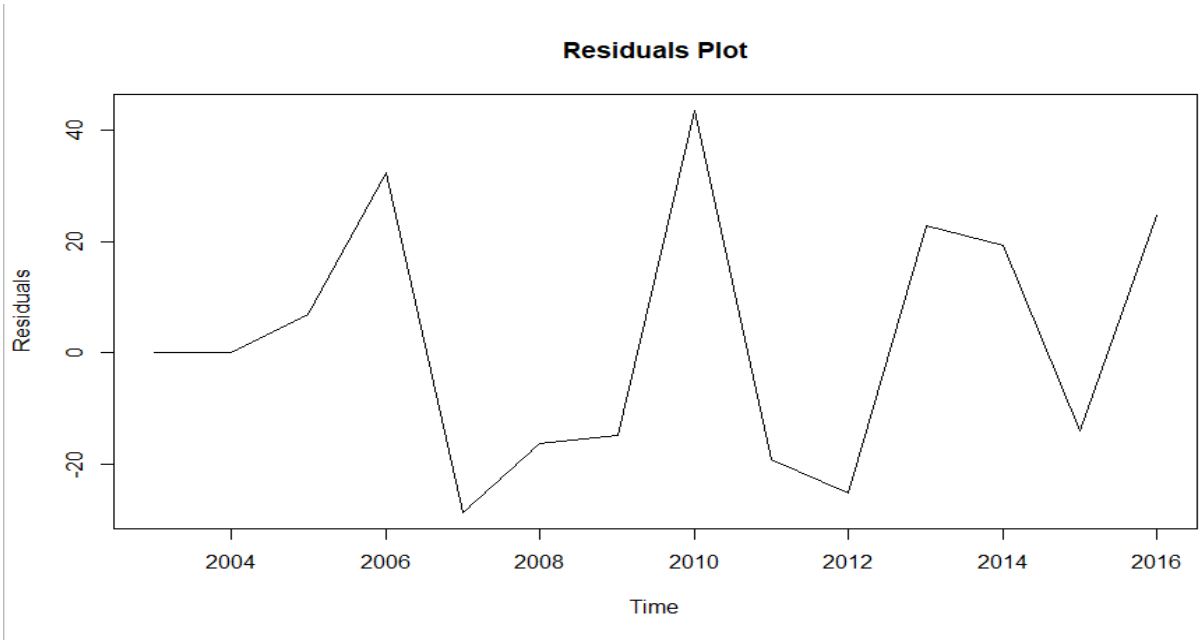


Table 4. 20: Box-Ljung test of Beijing ADRVs

Lags	P-value	Inference
5	0.2424	Not significant
10	0.5575	Not significant

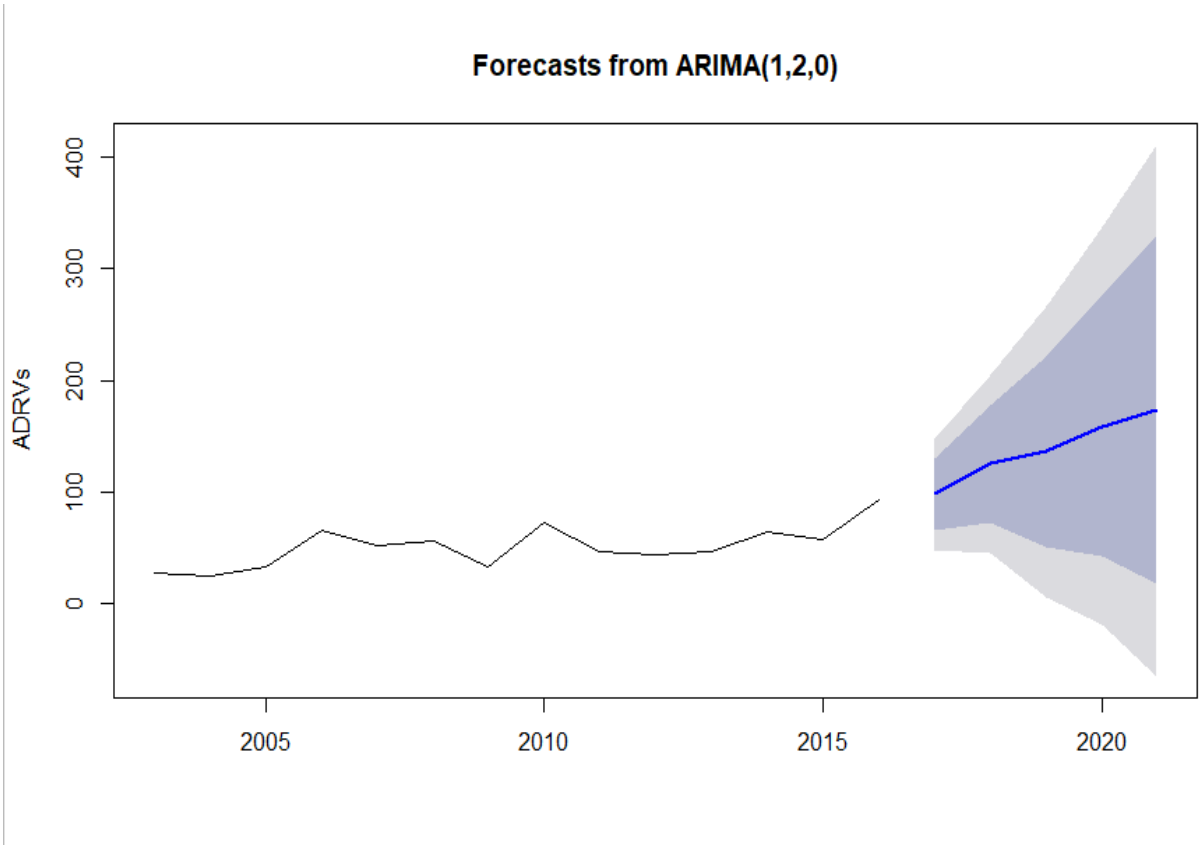
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.21. The data was also plotted as indicated in Figure 4.35. The predicted number of ADRVs rose gradually from 98 in 2017 to 173 ADRVs in 2021.

Table 4. 21: Forecasting of Beijing ADRVs

Year	Forecast value
2017	98
2018	125
2019	137
2020	159
2021	173

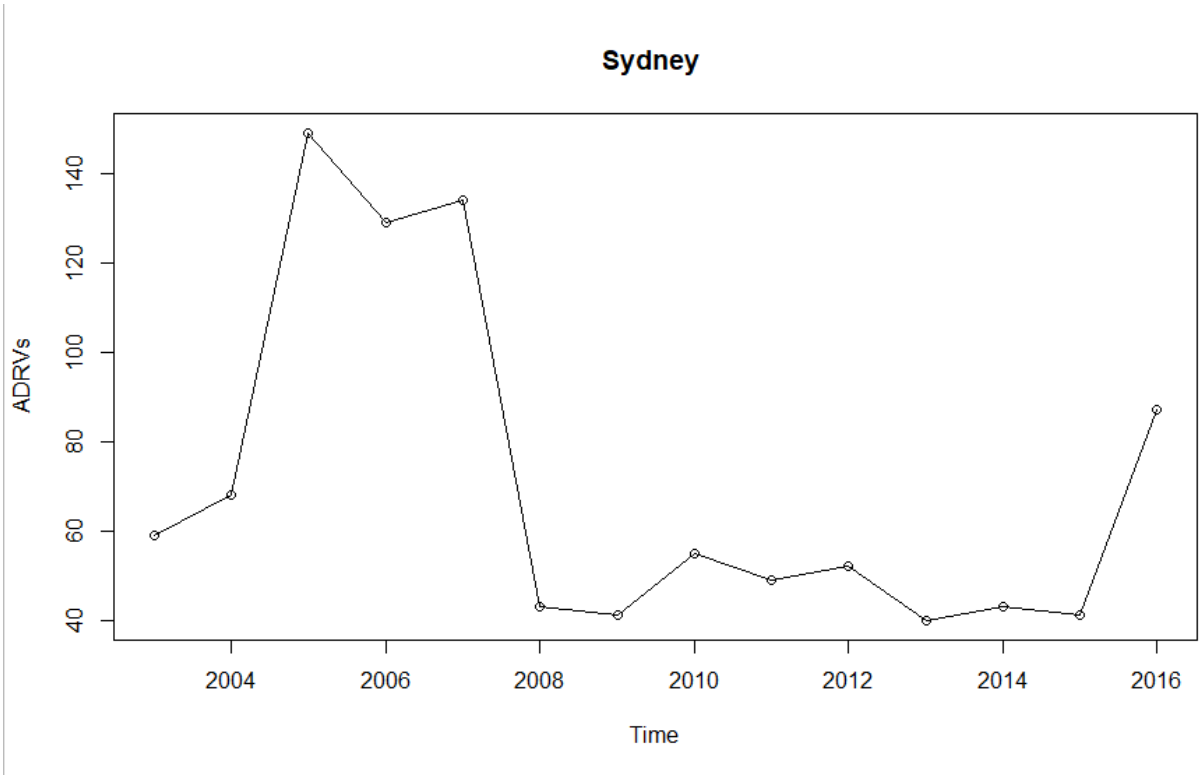
Figure 4. 35: Forecasts of Beijing ADRVs



4.1.8 Trend of Sydney ADRVs

The data from Sydney laboratory indicated a marginally increasing trend. The cyclic aspect repeated after three years twice then after two years three times. This is illustrated in Figure 4.1 below.

Figure 4. 36: Plot of Sydney ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.37 and Figure 4.38 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(1,2,0) and was chosen as the most appropriate model. This is illustrated in Table 4.22 below.

Figure 4. 37: ACF Plot of Sydney ADRVs

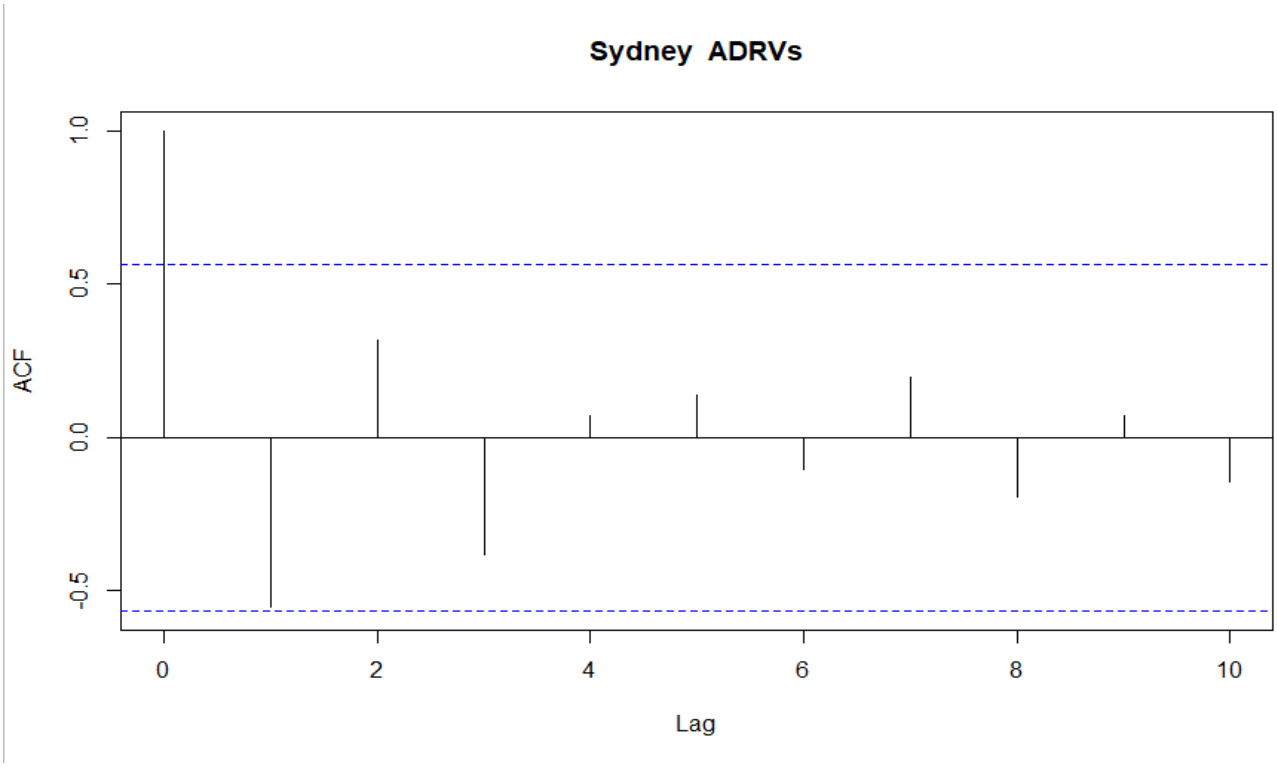


Figure 4. 38: PACF Plot of Sydney ADRVs

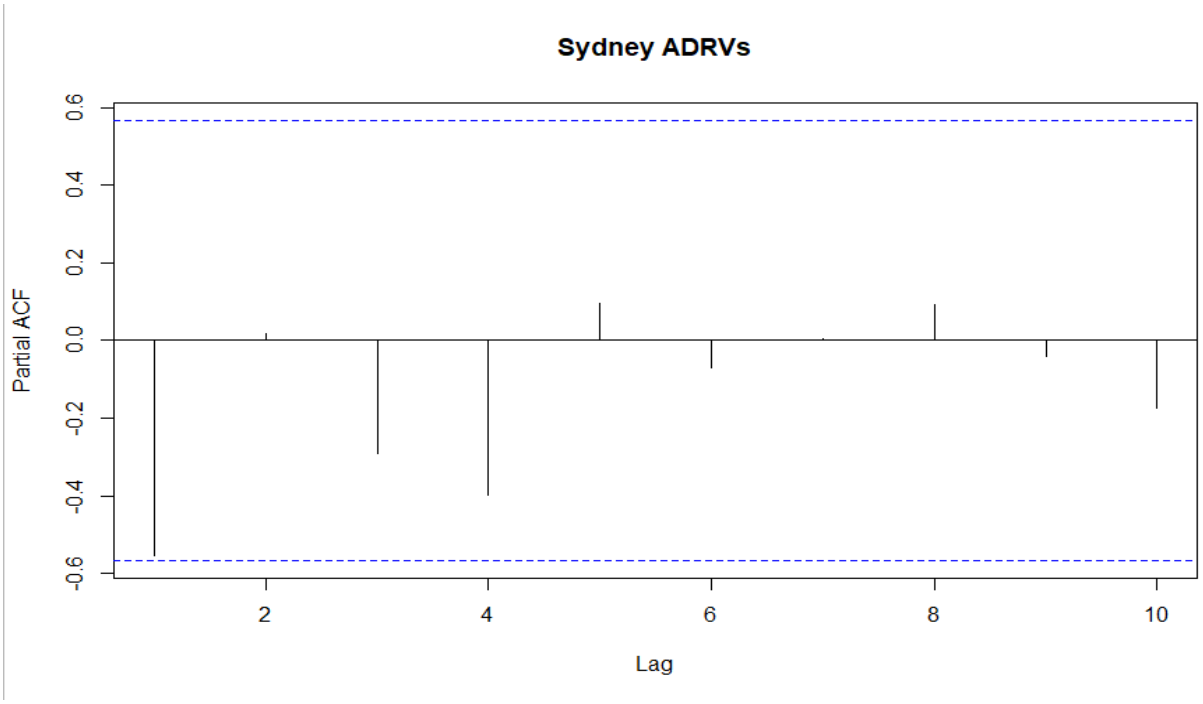


Table 4. 22: ARIMA models of Sydney ADRVs

Model type	AIC	Inference
ARIMA(0,2,0)	25.68856	
ARIMA(1,2,0)	24.07155	Best model
ARIMA(2,2,0)	26.06661	
ARIMA(1,2,1)	26.01954	

Diagnostics and validation

The chosen model ARIMA(1,2,0) was validated using a residuals plot as illustrated in Figure 4.39 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.23 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 39: Residuals Plot of Sydney ADRVs

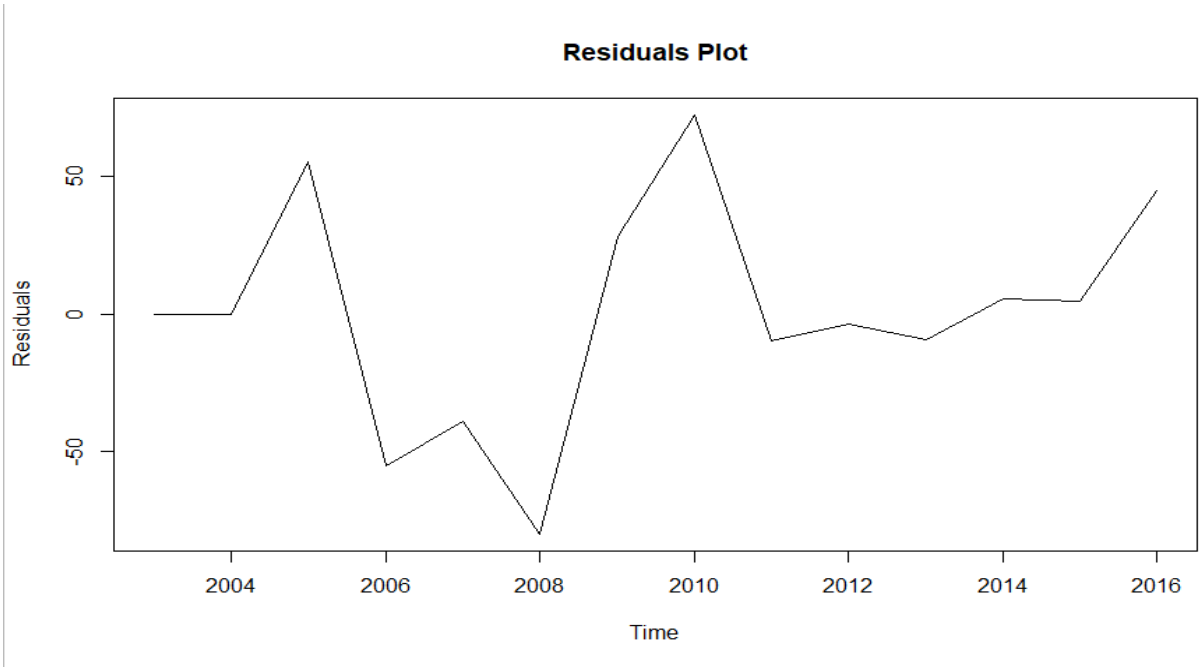


Table 4. 23: Box-Ljung test of Sydney ADRVs

Lags	P-value	Inference
5	0.2996	Not significant
10	0.5394	Not significant

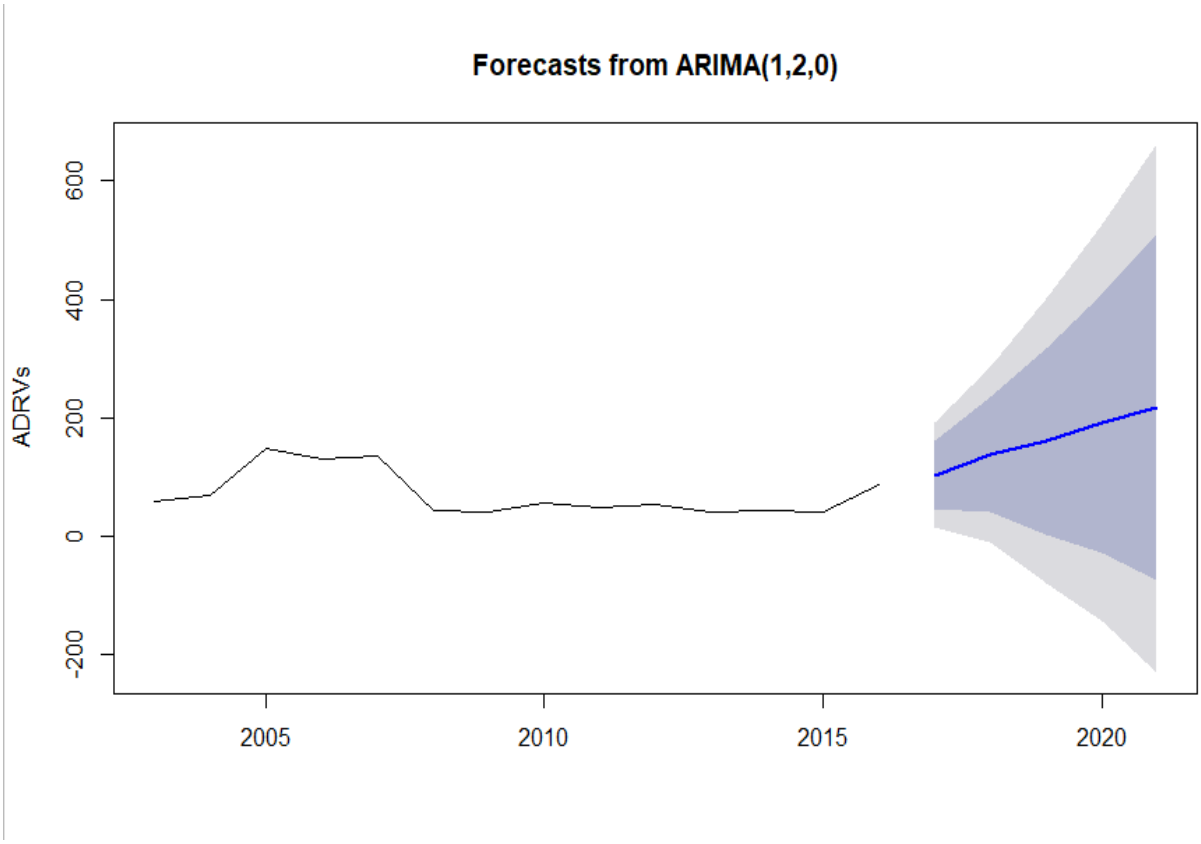
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.24. The data was also plotted as indicated in Figure 4.40. The predicted number of ADRVs rose gradually from 103 in 2017 to 216 ADRVs in 2021.

Table 4. 24: Forecasting of Sydney ADRVs

Year	Forecast value
2017	103
2018	137
2019	160
2020	190
2021	216

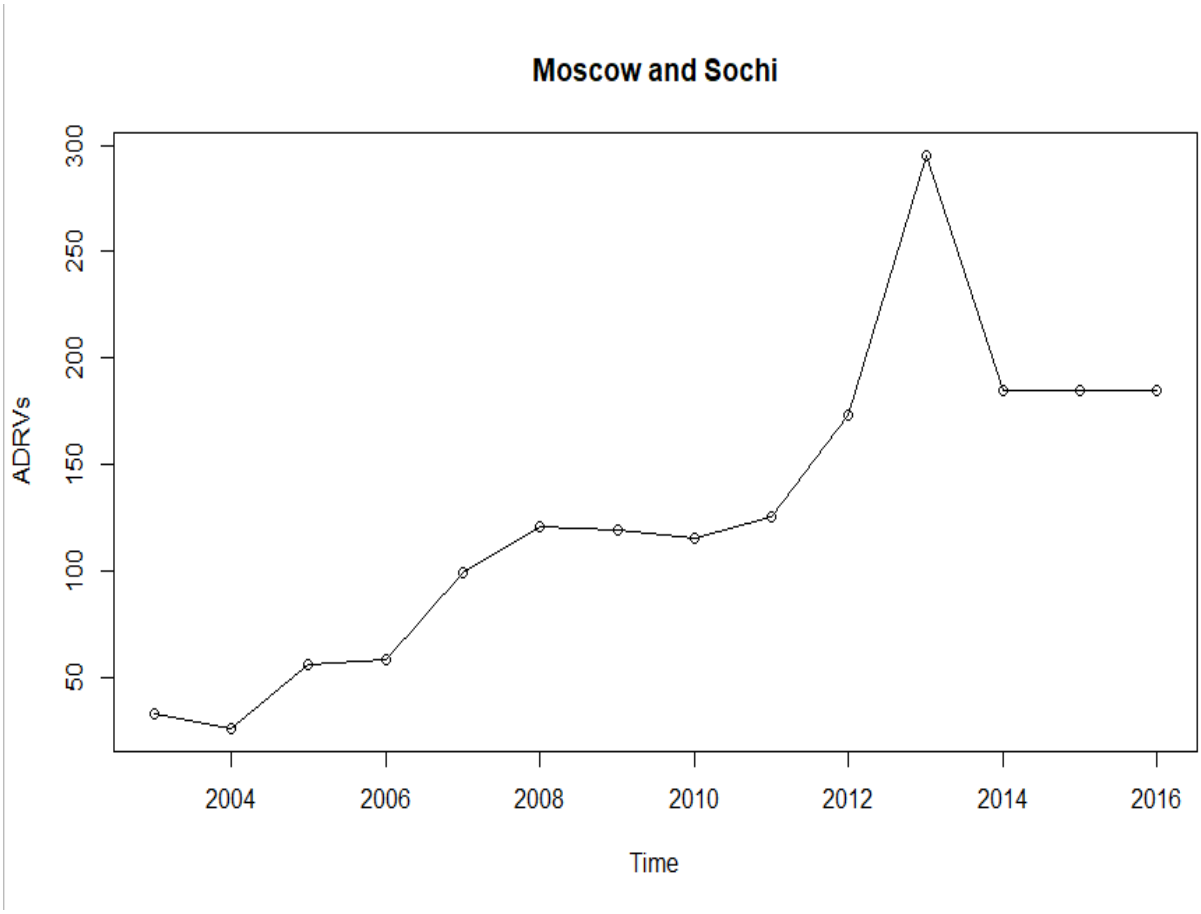
Figure 4. 40: Forecasts of Sydney ADRVs



4.1.9 Trend of Moscow and Sochi ADRVs

The data from Moscow and Sochi laboratories indicated an increasing trend. The cyclic aspect repeated after two years on three occasions then four years on one occasion. This is illustrated in Figure 4.41 below.

Figure 4. 41: Plot of Moscow and Sochi ADRVs



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.42 and Figure 4.43 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(0,2,2) and was chosen as the most appropriate model. This is illustrated in Table 4.25 below.

Figure 4. 42: ACF Plot of Moscow and Sochi ADRVs

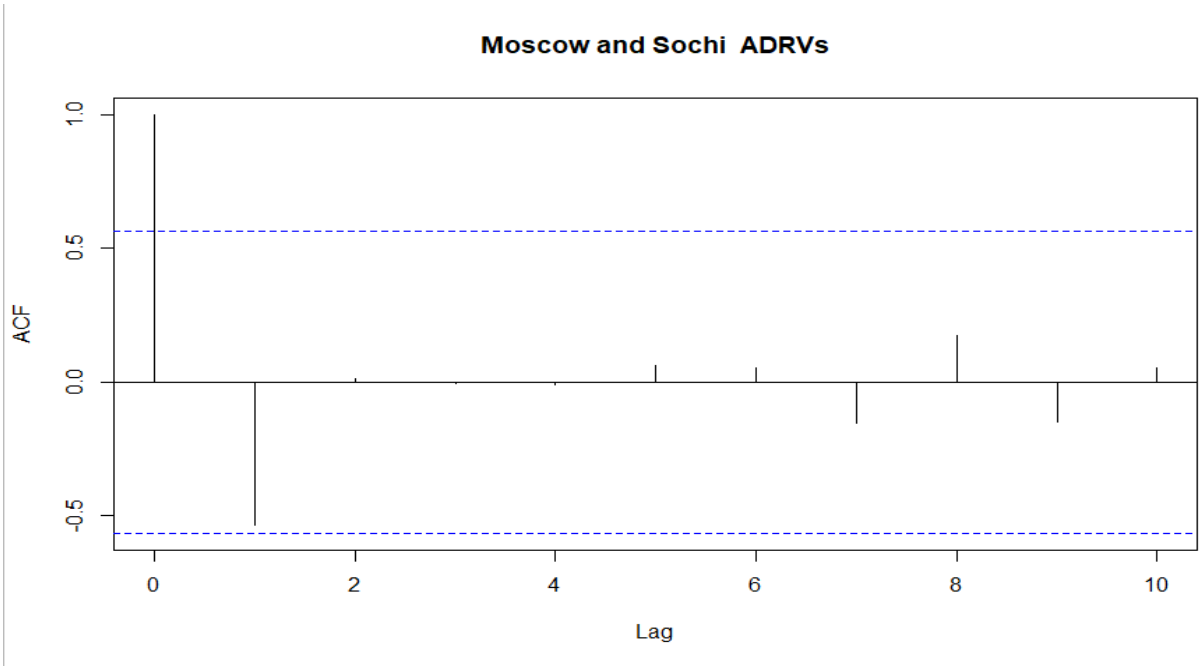


Figure 4. 43: PACF Plot of Moscow and Sochi ADRVs

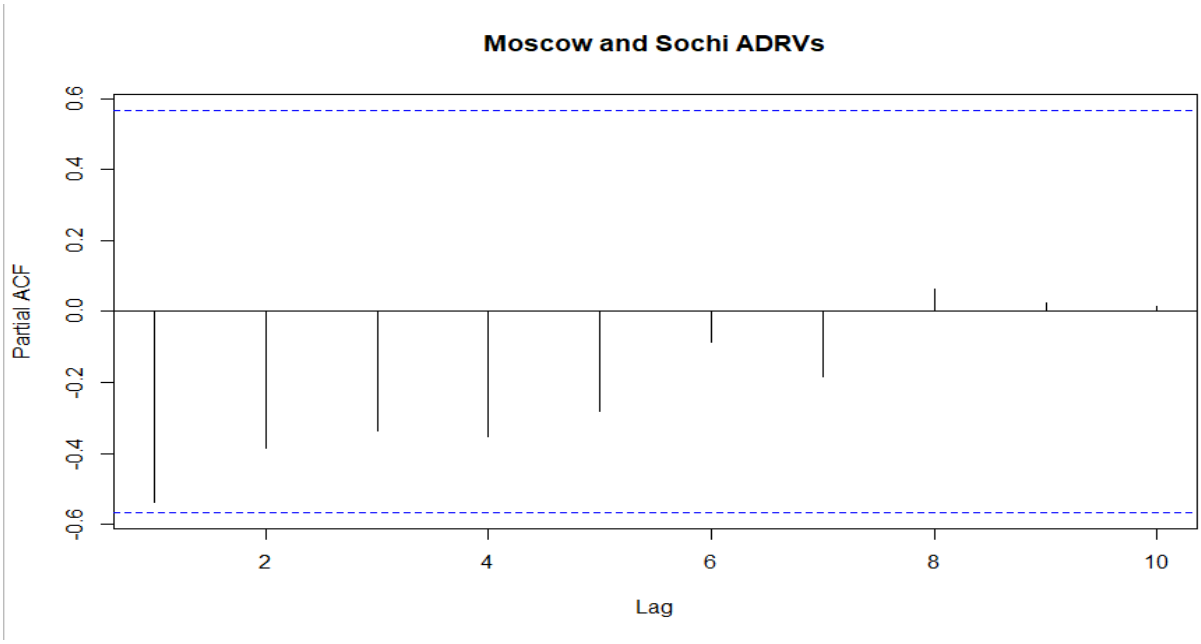


Table 4. 25: ARIMA models of Moscow and Sochi ADRVs

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(0,2,0)	20.52705	
ARIMA(1,2,0)	16.60393	
ARIMA(0,2,1)	Infinite	
ARIMA(0,2,2)	15.88393	Best model
ARIMA(2,2,1)	17.82284	

Diagnostics and validation

The chosen model ARIMA(0,2,2) was validated using a residuals plot as illustrated in Figure 4.44 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.26 below. The p-values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 44: Residuals Plot of Moscow and Sochi ADRVs

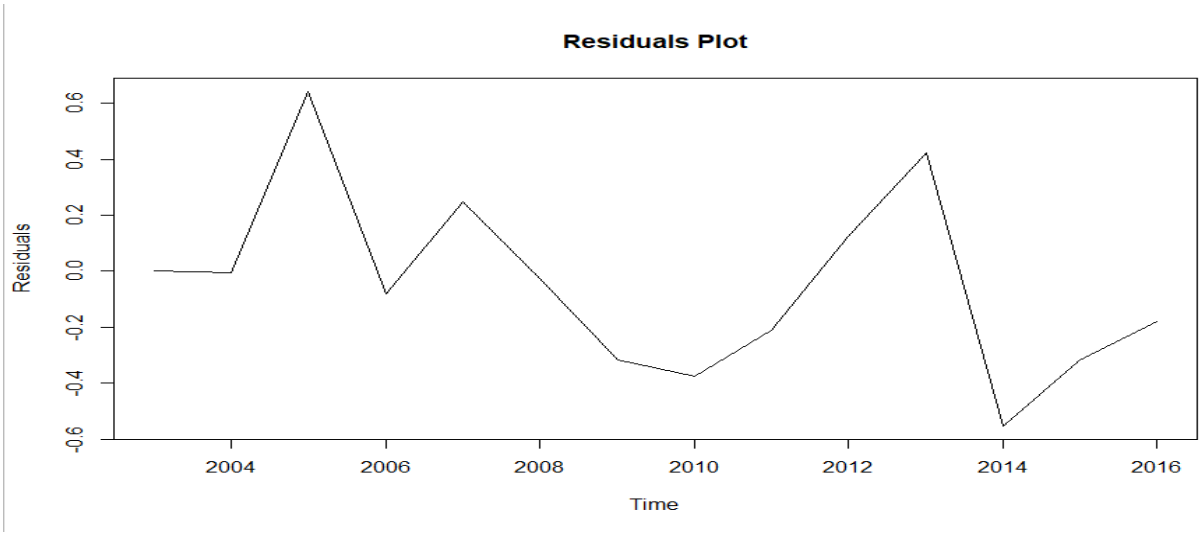


Table 4. 26: Box-Ljung test of Moscow and Sochi ADRVs

Lags	P-value	Inference
5	0.9139	Not significant
10	0.7341	Not significant

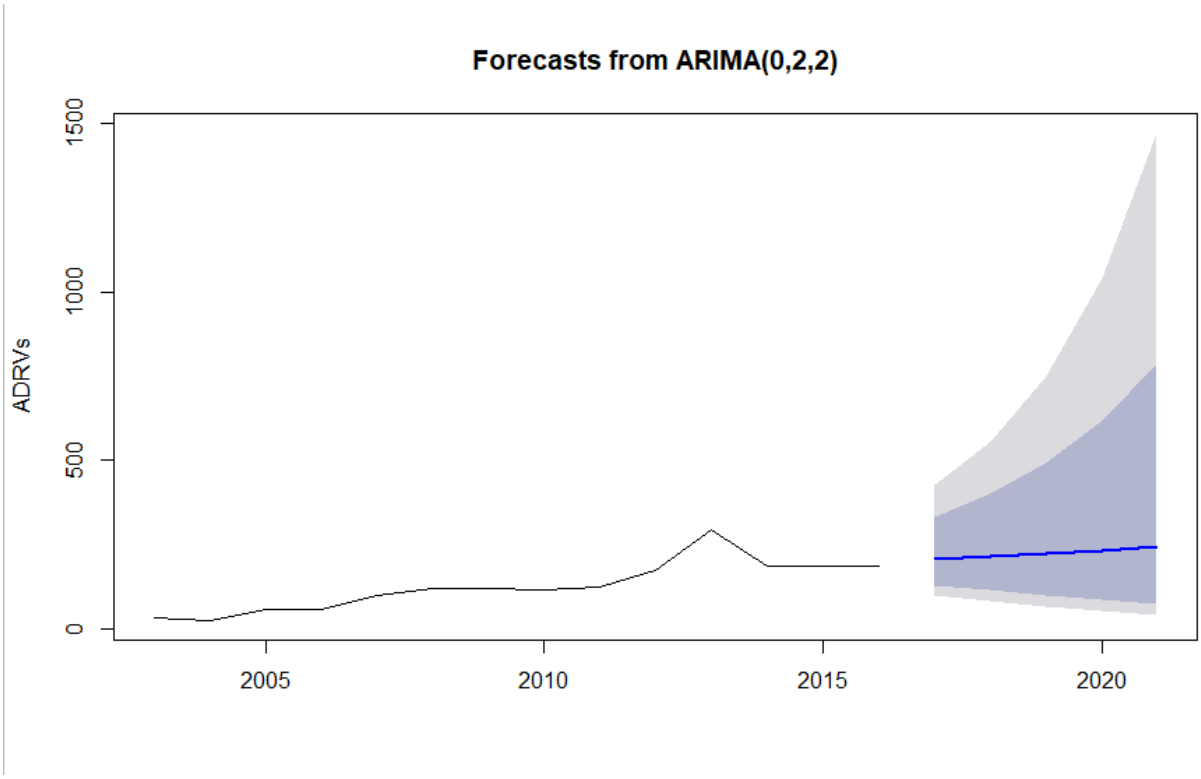
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.27. The data was also plotted as indicated in Figure 4.45. The predicted number of ADRVs rose gradually from 205 in 2017 to 243 ADRVs in 2021.

Table 4. 27: Forecasting of Moscow and Sochi ADRVs

Year	Forecast value
2017	205
2018	214
2019	223
2020	233
2021	243

Figure 4. 45: Forecasts of Moscow and Sochi ADRVs



4.2 PEDs associated with ADRVs

The second specific objective was to establish the PEDs giving positive doping tests in the 14 year period globally. The classification of PEDs is as given by WADA. The results showed that the most commonly used PEDs were anabolic steroids followed by stimulants and cannabinoids. This is illustrated in Figure 4.46, Figure 4.47 and Table 4.28 below.

Figure 4. 46: Doping Agents in ADRVs

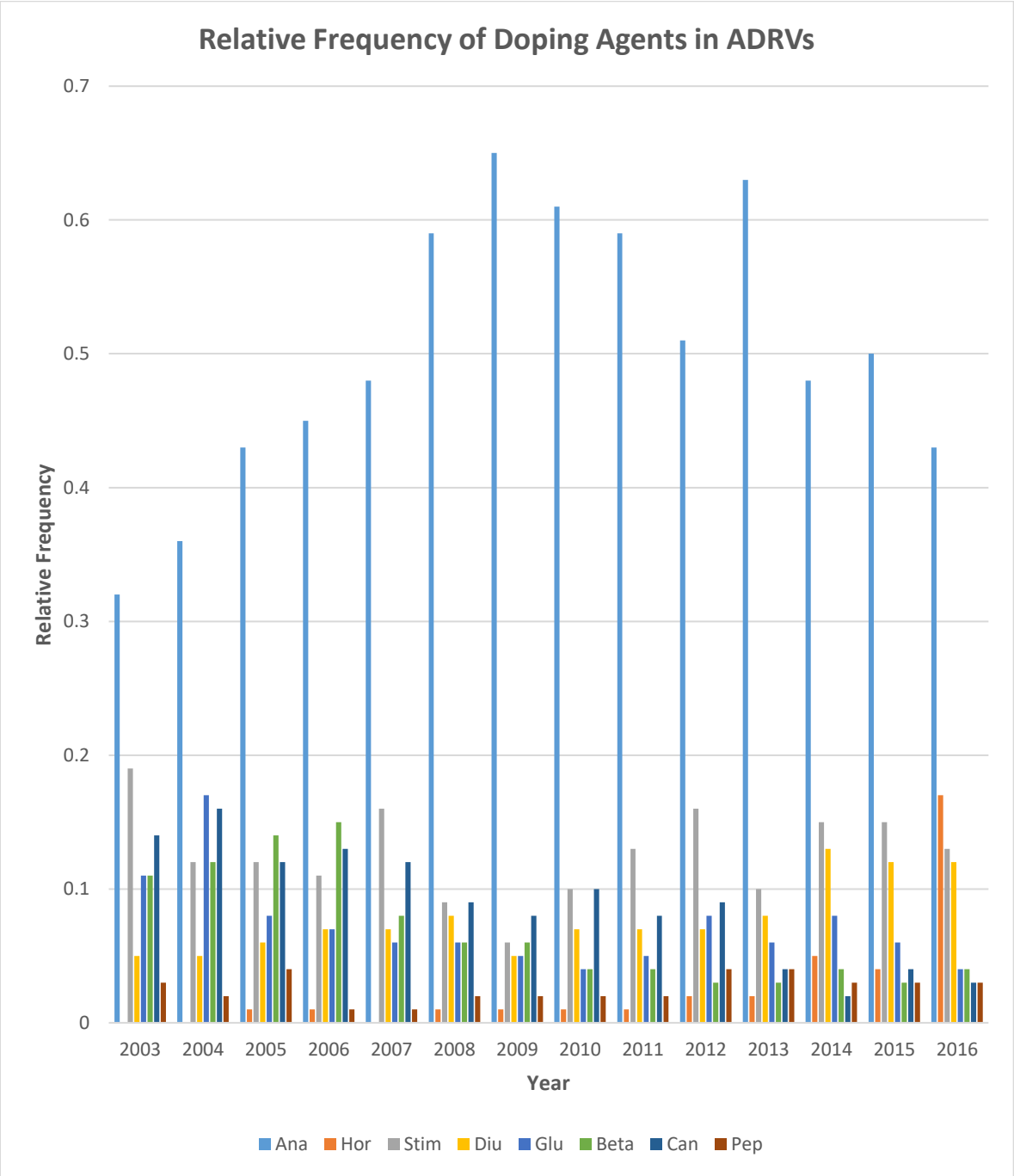


Figure 4. 47: Olympics Years Doping Agents in ADRVs

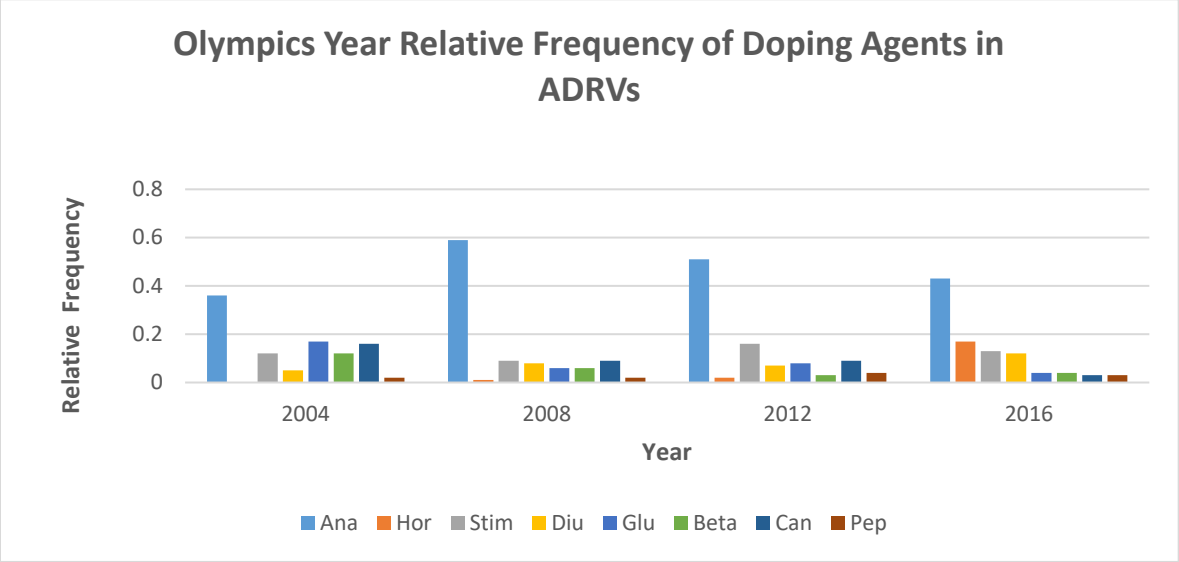


Table 4. 28: Yearly Top-three ranked PEDs during study period

Class of agents	**Frequency**
Ana	14
Stim	10
Can	9
Beta	4
Glu	4
Diu	3
Hor	1

Where:

Ana - Anabolic Steroids

Hor - Hormone and Metabolic Modulators

Stim - Stimulants

Diu - Diuretics and Other Masking Agents

Glu - Glucocorticosteroids

Beta - Beta-2 Agonists

Can – Cannabinoids

Pep - Peptide Hormones, Growth Factors and Related Substances

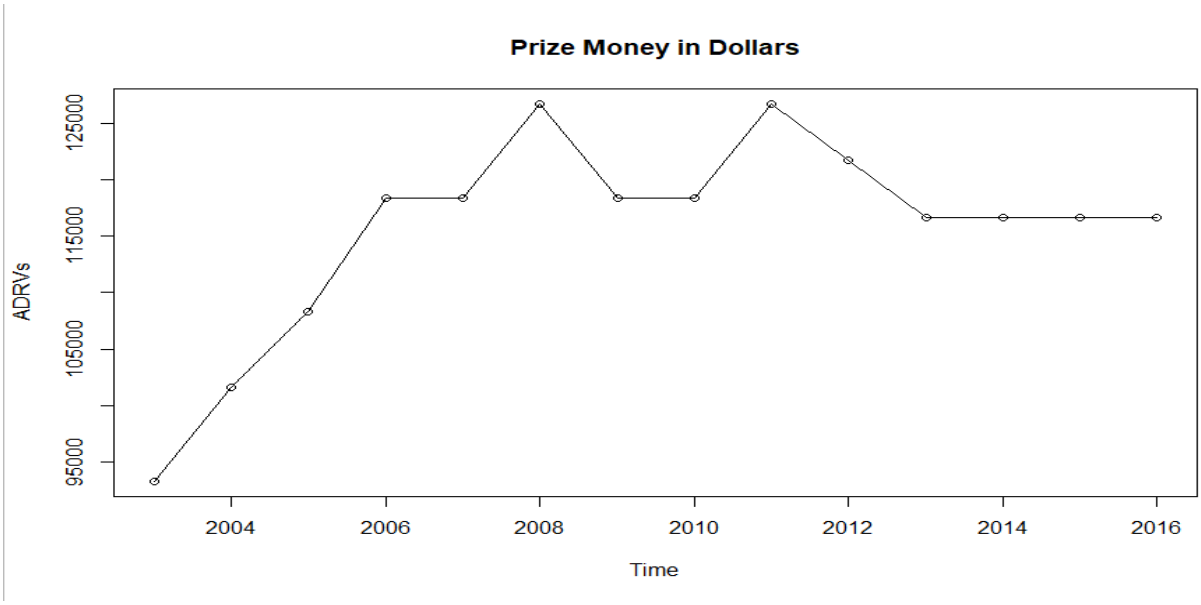
4.3 Granger-causality analysis

The third specific objective was to establish whether prize money Granger caused the corresponding ADRVs in the 14 year period. Analysis was preceded by Box-Jenkins procedure on the prize money data.

4.3.1 Trend of Prize Money

The data for prize money indicated an increasing trend. The cyclic aspect repeated after two years on three occasions then after four years once. This is illustrated in Figure 4.49 below.

Figure 4. 48: Plot of Prize Money



Identification and estimation

Plots of ACF and PACF were generated from the data as illustrated in Figure 4.50 and Figure 4.51 below. The plots were not providing clear cut solutions to the types of models to be used hence an automated model search based on AIC implemented in R using 'auto.arima' function was used. The model with the lowest AIC value was ARIMA(0,2,1) and was chosen as the most appropriate model. This is illustrated in Table 4.29 below.

Figure 4. 49: ACF Plot of Prize Money

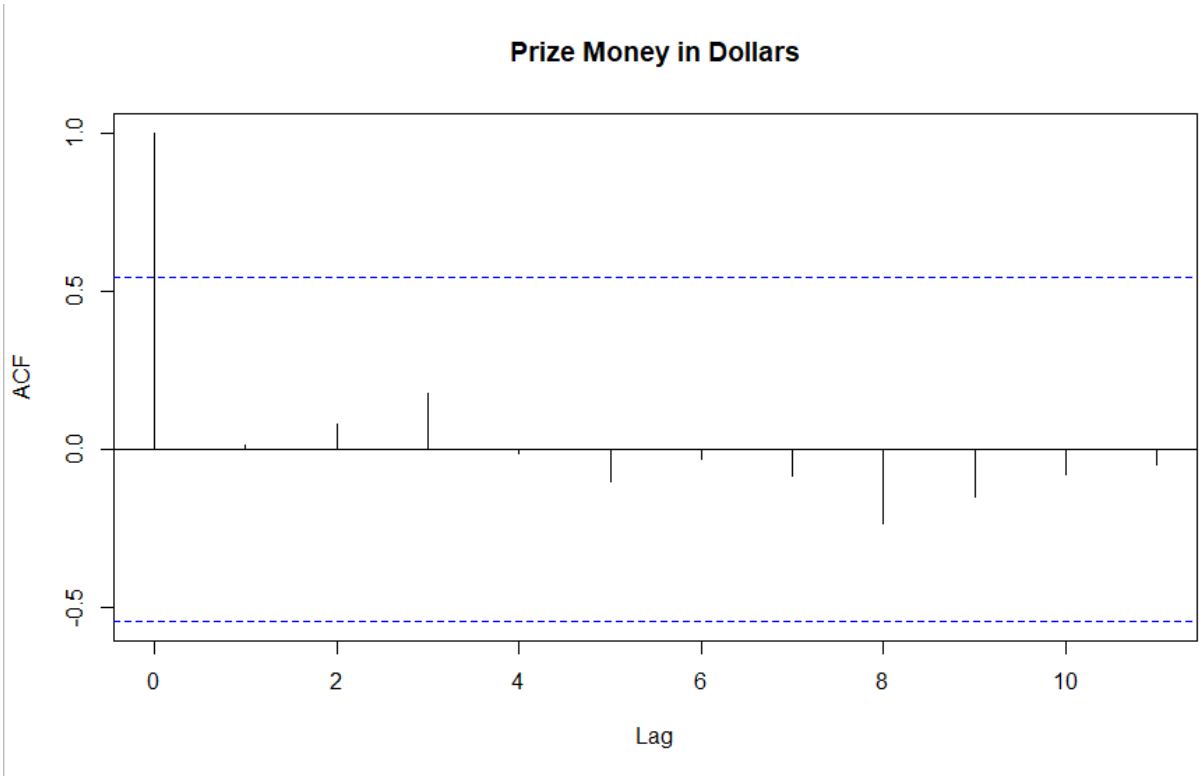


Figure 4. 50: PACF Plot of Prize Money

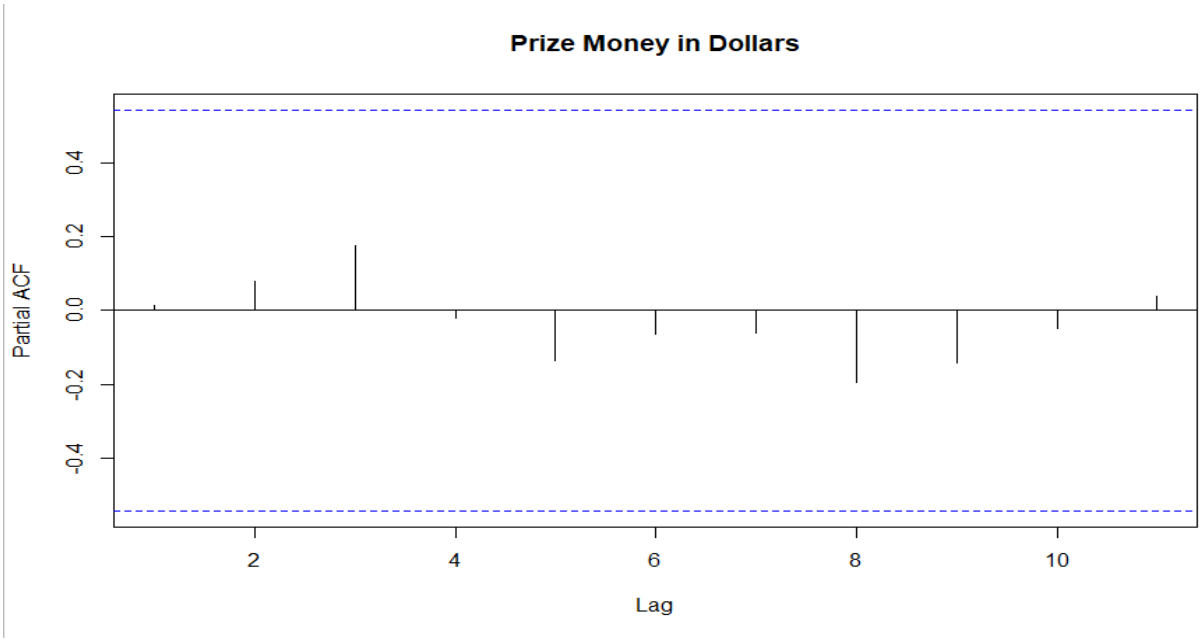


Table 4. 29: ARIMA models of Prize Money

Model type	AIC	Inference
ARIMA(2,2,2)	Infinite	
ARIMA(0,2,0)	-28.6106	
ARIMA(1,2,0)	-30.5641	
ARIMA(0,2,1)	-32.06682	Best model
ARIMA(1,2,2)	infinite	

Diagnostics and validation

The chosen model ARIMA(0,2,1) was validated using a residuals plot as illustrated in Figure 4.52 below. The plot vibrated around zero meaning the model was acceptable. This was farther confirmed by the Box-Ljung test at lag 5 and lag 10 as illustrated in Table 4.30 below. The p-

values were higher than 0.05 thus the chosen model was acceptable and could be used for forecasting.

Figure 4. 51: Residuals Plot of Prize Money

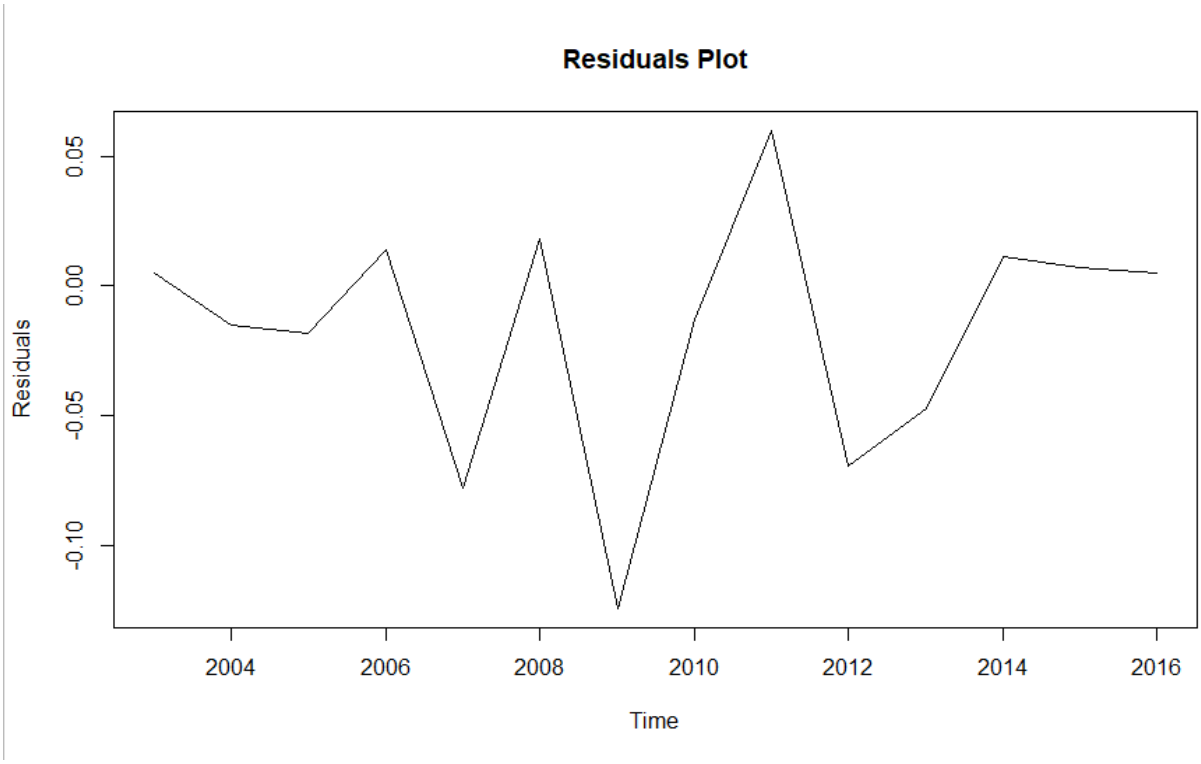


Table 4. 30: Box-Ljung test of Prize Money

Lags	P-value	Inference
5	0.567	Not significant
10	0.8755	Not significant

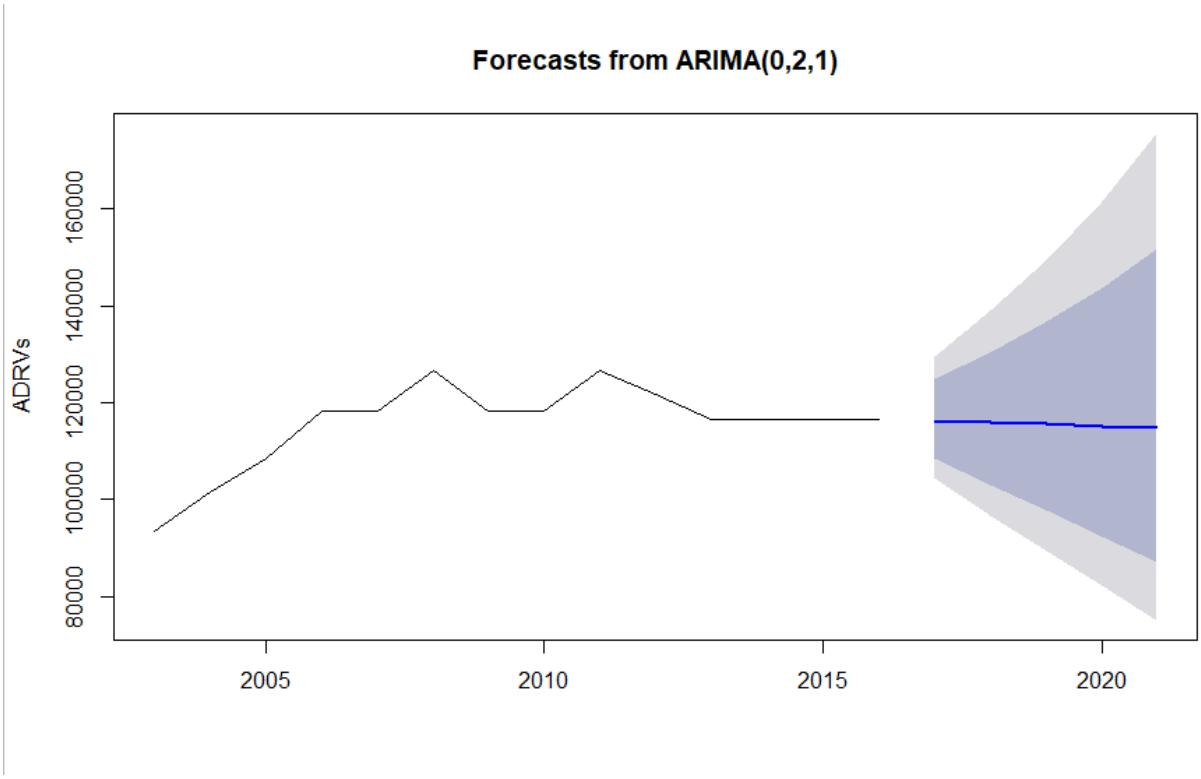
Forecasting

The validated model was used to forecast values of the next five years. They are tabulated in Table 4.31. The data was also plotted as indicated in Figure 4.53. The predicted prize money fell marginally from USD 116, 301 in 2017 to USD 114, 851 in 2021.

Table 4. 31: Forecasting of Prize Money

Year	Forecast value in USD
2017	116, 301
2018	115, 937
2019	115, 574
2020	115,212
2021	114, 851

Figure 4. 52: Forecasts of Prize Money in Dollars



4.3.2 Findings on Granger-causality

The data on prize money was compared with the ADRVs from the nine regions analyzed. Granger causality was examined and results tabulated in Table 4.32. ADRVs data from Paris and Sydney laboratories were Granger caused by prize money. Data from the other seven regions did not indicated Granger causality.

Table 4. 32: Granger causality test results

Accredited testing Centre	P-Value (order = 2)	Inference
Bloemfontein and Tunis	0.5649	Not significant
London and Cambridge	0.07536	Not significant
Paris	0.00353	Significant
Los Angeles and Salt Lake	0.8679	Not significant
Montreal	0.1655	Not significant
Rio	0.734	Not significant
Beijing	0.1568	Not significant
Sydney	0.01272	Significant
Moscow	0.4323	Not significant

4.4 RADO Zone V doping statistics

The fourth specific objective was to establish the current doping statistics in Kenya and compare them with the statistics in the RADO Zone V region checking for spatial correlation. The results from 2013 to 2016 were plotted as shown in Figure 4.54. Moran's I and Geary's C based on the mean of countries' ADRVs were used together as tabulated in Table 4.33 showing the spatial

distribution as random. The individual countries' ADRVs was used to generate Figure 4.55 showing color coded spatial distribution. The pattern was similar throughout the period.

Figure 4. 53: Plot of ADRVs in the RADO Zone V

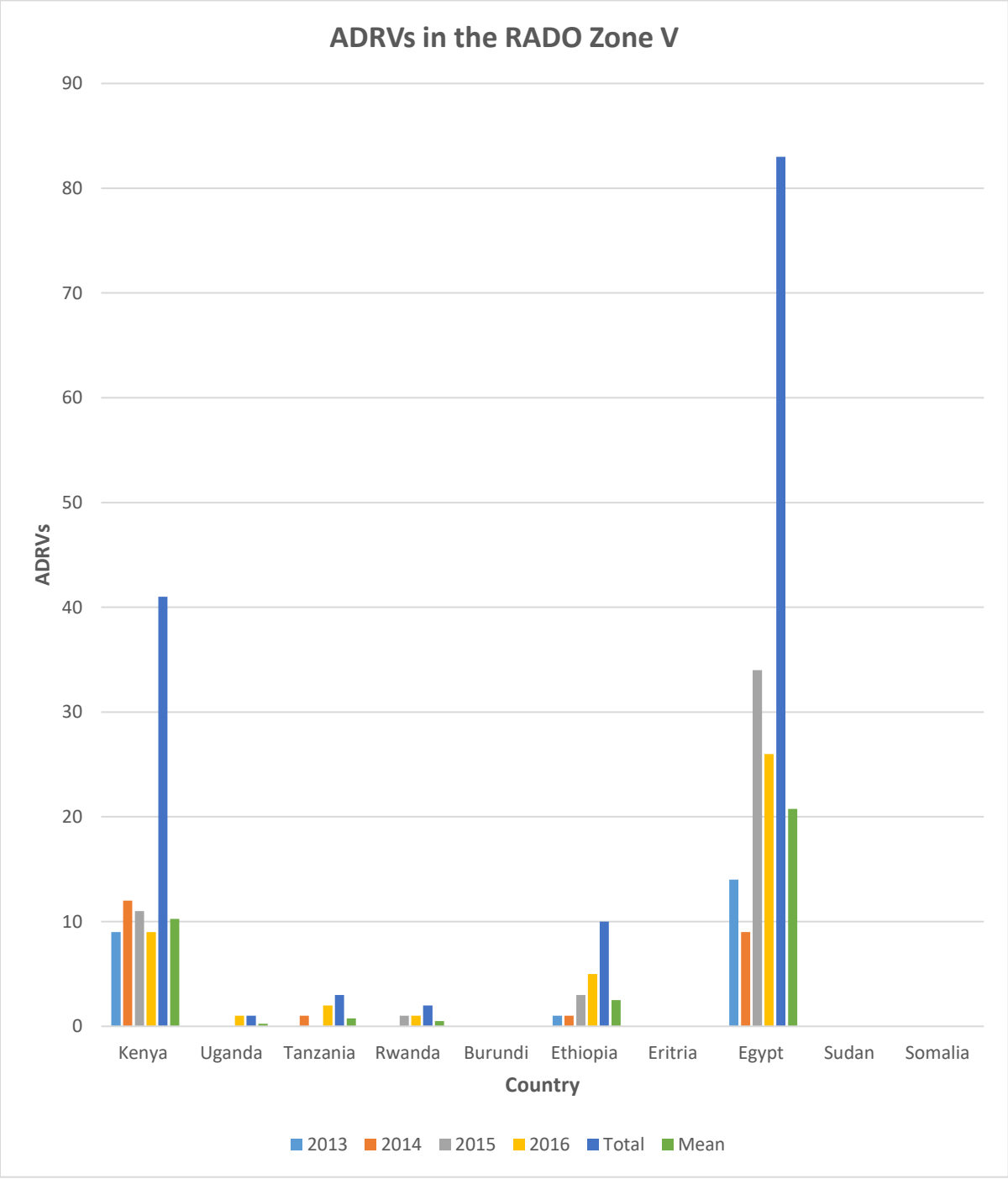
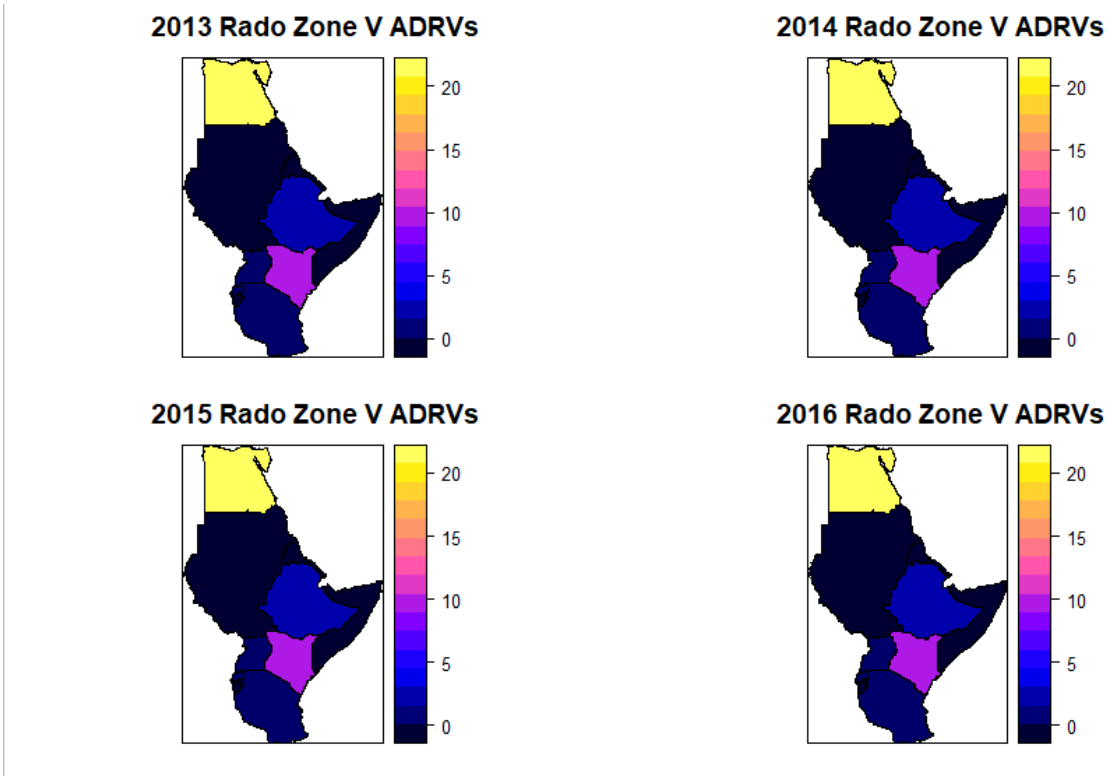


Table 4. 33: Auto-spatial Correlation results

Test Done	Result	Expectation	Variance
Moran's I	-0.1096	0.1111	0.0216
Geary's C	0.8116	1	0.0666

Figure 4. 54: Spatial distribution of ADRVs in the RADO Zone V



CHAPTER FIVE

DISCUSSION AND CONCLUSION

5.0 Introduction

This chapter discusses the findings detailed in the previous chapter. It shows the limitations encountered in the process, the conclusion, the recommendations and finally areas for future research.

5.1 Discussion

The first specific objective was to establish the regional doping trend for 14 years (2003-2016) and compare it with global trends. The data is annual so it has no seasonal component. The plots had trend, cyclic and random components. Bloemfontein and Tunis for Africa, London and Cambridge for UK, Paris for France, and Sydney for Australia had a downward trend indicating a fall in the number of ADRVs. Montreal Canada had a visually flat trend. Los Angeles and Salt Lake in the USA, Rio in Brazil for South America, Beijing for China and Moscow and Sochi in Russia had an upward trend showing an increase in ADRVs. All the ARIMA models were validated using residuals plots and Box-Ljung test and forecasts made for the next 5 years. The data for Bloemfontein and Tunis corresponded to ARIMA (0,2,1). Forecasts indicated that for the next five years, ADRVs in the African region would fall from 9 in 2017 to 1 in 2021. The other sets of data showing a predicted drop in ADRVs are Los Angeles and Salt Lake as well as Montreal. London and Cambridge as well as Paris data forecasted a virtually constant number of ADRVs in the five-year period. Data from Rio, Beijing, Sydney as well as Moscow and Sochi predicted a rising

number of ADRVs from 2017 to 2021. The data for prize money corresponded to ARIMA (0,2,1) and forecasts fell slightly from USD 116,667 to 114, 851.

The second-specific objective was to establish the critical PEDs giving positive doping tests in the same period globally. WADA classifies the substances as anabolic steroids, hormones and metabolic modulators, stimulants, diuretics and other masking agents, glucocorticosteroids, beta-2 agonists, cannabinoids and peptide hormones, growth factors and related substances. In the 14 year period, the categories with the highest frequency in the yearly top three agents are as follows in descending order: anabolic steroids, stimulants, cannabinoids, beta-2 agonists, glucocorticosteroids, diuretics and other masking agents, and hormones and metabolic modulators. The class of peptide hormones, growth factors and related substances did not appear in the top three during the period.

The third-specific objective was to establish the relationship between the global doping trends and corresponding prize money in the same period. Granger causality was used for the analysis. The data from Bloemfontein and Tunis for Africa, London and Cambridge for UK, Montreal Canada, Los Angeles and Salt Lake in the USA, Rio in Brazil for South America, Beijing for China and Moscow and Sochi in Russia gave P-values higher than 0.05 indicating no Granger causality was observed from the data. The data from the laboratory in Paris for France and Sydney for Australia gave P-value less than 0.05 meaning the prize money Granger caused ADRVs.

The fourth-specific objective was to establish the current doping statistics in Kenya and compare them with the statistics from the RADO Zone V, checking for spatial autocorrelation. The zone has the following countries: Kenya, Uganda, Tanzania, Burundi, Rwanda, Ethiopia, Eritrea, Sudan, Somalia and Egypt. The data analyzed was from 2013 to 2016. Eritrea, Sudan and Somalia did not have any ADRVs in the period. Kenya and Egypt showed the highest number of ADRVs. The

population of Kenya, Ethiopia and Egypt respectively are estimated as 52, 112 and 100 million, respectively. The population of Egypt is nearly twice that of Kenya while the average ADRVs for the period for Egypt is 21 while Kenya is 11 therefore the doping numbers in Kenya per capita are about the same as Egypt. The average for Ethiopia is 3. The other countries have a mean of less than one ADRV. The Moran's I value for the RADO zone V was -0.1096 while the Geary's C value was 0.8116. For Moran's I values around zero indicate random distribution of observations in space. For Geary's C values around one indicate the same distribution.

5.2 Limitations

This study was based on data obtained from WADA reports. The data was therefore limited to the genres and tiers of sports WADA had compiled and not necessarily what was actually going on the ground in terms overall doping. This was further compounded by the fact that the effectiveness and efficiency of WADA had been changing over the study period.

WADA data was based on testing of priority countries and priority athletes. The fact that some countries report zero doping may not necessarily mean zero doping. It could actually be an indicator of reduced government funding of anti-doping activities.

Lastly, this study was done under time and financial constraints. The data covers a vast region and thus would require vast resources to collect and verify some data not explicitly available from WADA reports. This was not possible.

5.3 Conclusion

This study showed that generally ADRVs in Africa were on a decline which was the same case as Kenya. Anabolic steroids were the most critical PEDs in terms of consistency throughout the period and relative frequency each year. Spatial autocorrelation was not observed in the data for

RADO Zone V. This indicates that countries in the region with high doping numbers such as Kenya and Egypt have not influenced their neighboring countries either to dope or to avoid doping. Prize money was not shown to influence numbers of ADRVs in Africa and six other regions studied with the exception of Australia and France.

5.4 Recommendations

The data from Rio, Beijing, Sydney as well as Moscow and Sochi predicted a rising number of ADRVs from 2017 to 2021, and WADA needs to take specific actions in these regions so that the vice does not spread to other regions globally. The numbers in Africa have been falling. Kenya's ADRVs have been falling since 2014, but still, the country is at the top of the list in the region along with Egypt. Kenya is juxtaposed with Ethiopia that has about a quarter the number of ADRVs in Kenya while its population is more than twice that of Kenya. This indicates that Ethiopia could be doing something right as far as anti-doping is concerned. The class of doping agents requiring urgent attention is anabolic steroids, and individual countries as well as WADA should address it.

5.5 Further Research

Some countries observed an increase in ADRVs while others observed a decline. Studies should be conducted that standardize the absolute numbers against the population of respective countries and the general uptake of anti-doping activities in the regions. This takes care of the possibility that increase in ADRVs could be as a result of more vigilance with modern equipment as opposed to higher rates of doping. Lower numbers could be caused by lower participation in sports as well as lower vigilance of athletes in some regions, among other reasons and therefore, studies should be conducted that control for such variables. There is need to address the reasons behind anabolic

steroids being the leading class of PEDs in ADRVs consistently in the study period and studies should also be conducted in this area. Finally, there is also need for studies to be conducted on the global spatial autocorrelation of doping numbers.

REFERENCES

Adams, I. B., & Martin, B. R. (1996). Cannabis: pharmacology and toxicology in animals and humans. *Addiction*, *91*(11), 1585-1614.

Aguilar, M., Muñoz-Guerra, J., Plata, M. D. M., & Del Coso, J. (2017). Thirteen years of the fight against doping in figures. *Drug testing and analysis*, *9*(6), 866-869.

Amestica, M.C. (2016). Sociology of Sport: Chile. *In Sociology of Sport: A Global Subdiscipline in Review* (pp. 405 – 419). Emerald Group Publishing Limited.

Anderson, T. W. (2011). *The statistical analysis of time series* (Vol. 19). John Wiley & Sons.

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of neuroscience methods*, *223*, 50-68.

Barroso, O., Mazzoni, I., & Rabin, O. (2008). Hormone abuse in sports: the antidoping perspective. *Asian journal of andrology*, *10*(3), 391-402.

Bassham, G., & Krall, C. (2010). Lance Armstrong and True Success. *Cycling-Philosophy for Everyone: A Philosophical Tour de Force*, 56-67.

Baume, N., Mahler, N., Kamber, M., Mangin, P., & Saugy, M. (2006). Research of stimulants and anabolic steroids in dietary supplements. *Scandinavian journal of medicine & science in sports*, *16*(1), 41-48.

Blank, C., Leichtfried, V., Schaiter, R., Fürhapter, C., Müller, D., & Schobersberger, W. (2015). Doping in sports: Knowledge and attitudes among parents of Austrian junior athletes. *Scandinavian journal of medicine & science in sports*, 25(1), 116–124.

Bloom, B. (2018). Asbel Kiprop claims sabotage and extortion by doping control officers after failed drugs test. Retrieved October 10, 2018, from

https://www.telegraph.co.uk/athletics/2018/05/03/asbel-kiprop-claims-sabotage-extortion-doping-officers-failed/.

Blumrodt, J., & Kitchen P.J. (2015). The Tour de France: corporate sponsorships and doping accusations. *Journal of Business Strategy*, Vol. 36 (2), 41-48.

Borgelt, L. M., Franson, K. L., Nussbaum, A. M., & Wang, G. S. (2013). The Pharmacologic and Clinical Effects of Medical Cannabis. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *33*(2), 195-209.

Bressler, S. L., & Seth, A. K. (2011). Wiener–Granger causality: a well-established methodology. *Neuroimage*, *58*(2), 323-329.

Burnett, C. (2016). Sociology of Sport: South Africa. In *Sociology of Sport: A Global Subdiscipline in Review* (pp. 3-19). Emerald Group Publishing Limited.

Cadwallader, A. B., De La Torre, X., Tieri, A., & Botrè, F. (2010). The abuse of diuretics as performance-enhancing drugs and masking agents in sport doping: pharmacology, toxicology and analysis. *British journal of pharmacology*, *161*(1), 1-16.

Cairns, S. P., & Borrani, F. (2015). β-Adrenergic modulation of skeletal muscle contraction: key role of excitation–contraction coupling. *The Journal of physiology*, *593*(21), 4713-4727.

Chappelet, J. L., & Van Luijk, N. (2018). The Institutional Governance of Global Hybrid Bodies: The Case of the World Anti-Doping Agency. In *Hybridity in the Governance and Delivery of Public Services* (pp. 167-191). Emerald Publishing Limited.

Christodoulos, C., Michalakelis, C., & Varoutas, D. (2010). Forecasting with limited data: Combining ARIMA and diffusion models. *Technological forecasting and social change*, *77*(4), 558-565.

Cortsen, K. (2013). Annika Sörenstam – a hybrid personal sports brand. *Sport, Business and Management: An International Journal*, 3(1), 37-62.

Cox, R. W. (2009). The Business of Sports. *Reference Reviews*.

Crompton, J.L. (2015). Potential negative outcomes from sports sponsorship. *International Journal of Sports Marketing and Sponsorship*, 16(3), 20-34.

Danylchuk, K., Stegink, J., & Lebel K. (2016). Doping scandals in professional cycling: impact on primary team sponsor's stock return. *International Journal of Sports Marketing and Sponsorship*, 17(1), 37-55.

de Bruijn, H., Groenleer, M., & van Ruijven, T. (2016). The dynamics of doping: Lance Armstrong, the United States Anti-Doping Agency and the regulatory governance of professional cycling. *Regulation & Governance*, 10, 284–297.

De Knegt, H. J., van Langevelde, F. V., Coughenour, M. B., Skidmore, A. K., De Boer, W. F., Heitkönig, I. M. A., ... & Prins, H. H. T. (2010). Spatial autocorrelation and the scaling of species–environment relationships. *Ecology*, *91*(8), 2455-2465.

De Sanctis, F. M. (2017). Olympic Games, Football Championships, and Corruption in the Sports Industry. In *The Handbook of Business and Corruption: Cross-Sectoral Experiences* (pp. 423-452). Emerald Publishing Limited.

Debarsy, N., & Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics*, *40*(6), 453-470.

Delbeke, F. T., & Debackere, M. (1988). The influence of diuretics on the excretion and metabolism of doping agents. Part IV-caffeine. *Biopharmaceutics and Drug Disposition*, 9, 137-145.

Deventer, K., Delbeke, F. T., Roels, K., & Van Eenoo, P. (2002). Screening for 18 diuretics and probenecid in doping analysis by liquid chromatography–tandem mass spectrometry. *Biomedical Chromatography*, *16*(8), 529-535.

Din, M.A. (2016). *ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education.* Athens: ATINER'S Conference Paper Series, No: EMS2015- Time

Docherty, J. R. (2008). Pharmacology of stimulants prohibited by the World Anti-Doping Agency. *British Journal of Pharmacology*, 154, 606-622.

Drivdal, M.H., Nordahl H.A., & Rønes H. (2018.) Sponsoring of professional cycling: what does it mean for stock prices? *International Journal of Sports Marketing and Sponsorship*, 19(1), 74-90.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Duval, A. (2017). The Russian doping scandal at the court of arbitration for sport: lessons for the world anti-doping system. *The International Sports Law Journal*, *16*(3-4), 177-197.

Erikainen, S. (2017). Hybrids, Hermaphrodites, and Sex Metamorphoses: Gendered Anxieties and Sex Testing in Elite Sport, 1937–1968. In *Gender Panic, Gender Policy* (pp. 155-176). Emerald Publishing Limited.

Flaherty, G., O'Connor, R., & Johnston, N. (2016). Altitude training for elite endurance athletes: A review for the travel medicine practitioner. *Travel medicine and infectious disease*, *14*(3), 200-211.

Forster, J. (2006). Global sports organizations and their governance. *Corporate Governance: The international journal of business in society*, *6*(1), 72-83.

Freestone, C. J., & Manoli, A. E. (2017). Financial fair play and competitive balance in the Premier League. *Sport, Business and Management: An International Journal*, *7*(2), 175-196.

Frick, B. (2003). Contest theory and sport. *Oxford Review of Economic Policy*, *19*(4), 512-529.

Frick, B., & Humphreys, B. R. (2011). *Prize structure and performance: Evidence from nascar* (No. 2011-12).

Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current opinion in neurobiology*, *23*(2), 172-178.

Gathura, M. (2018). 5 popular Kenyan athletes who got banned for doping. Retrieved October 10, 2018 from https://hivisasa.com/posts/5-popular-kenyan-athletes-who-got-banned-for-doping.

Gerbing, D. (2016). *Time Series Components.* Portland State University.

Getis, A. (2010). Spatial autocorrelation. In *Handbook of applied spatial analysis* (pp. 255-278). Springer, Berlin, Heidelberg.

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis* (pp. 127-145). Springer, Berlin, Heidelberg.

Griffith, D. A. (2013). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer Science & Business Media.

Haile, M. T., & Andargachew, M. Y. (2018). Knowledge of doping among Ethiopian professional middle and long distance runners. *Indian Federation of Computer Science in sports, 644*(58), 102.

Héas, S., & Régnier, P. (2016). Sociology of Sport: France. In *Sociology of Sport: A Global Subdiscipline in Review* (pp. 169-185). Emerald Group Publishing Limited.

Hervey, G.R. (1974). Hormones and energy balance. *Nutrition & Food Science*, 74(2), 17-19.

Heuberger, J. A., van Dijkman, S. C., & Cohen, A. F. (2018). Futility of current urine salbutamol doping control.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Jones, S. C. (2010). When does alcohol sponsorship of sport become sports sponsorship of alcohol? A case study of developments in sport in Australia. *International Journal of Sports Marketing and Sponsorship*, *11*(3), 67-78.

Kang, E. (2017). *Series: ARIMA Model.* Retrived June 20, 2019 from https://medium.com/@kangeugine/time-series-arima-model-11140bc08c6

Kasdan, D. O. (2013). Seven in a row? The Tour de France and the likelihood of Lance. *Significance*, *10*(1), 20-22.

Kayser, B., & Broers, B. (2012). The Olympics and harm reduction? *Harm reduction journal*, *9*(1), 33.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664-2675.

Kirui, J., Simotwo, S., & Kipkoech, L. C. (2013). An analysis of the extent of students' involvement in sports activities in Secondary Schools in Uasin Gishu County,

Kenya. *Journal of Emerging Trends in Educational Research and Policy Studies*, *4*(2), 392-398.

Lee, C. M., & Ko, C. N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, *38*(5), 5902-5911.

Lee, K. H. (2017). The Olympics, Soft Power, and Democratization: The Role of Bureaucracy. In *The Experience of Democracy and Bureaucracy in South Korea* (pp. 181-215). Emerald Publishing Limited.

Lenskyj, H. (2018). Sport exceptionalism and the Court of Arbitration for Sport. *Journal of Criminological Research, Policy and Practice*, *4*(1), 5-17.

Lenskyj, H.J. (2018). Doping, Genes, and Gender. In *Gender, Athletes' Rights, and the Court of Arbitration for Sport* (pp.115 – 158). Emerald Publishing Limited.

Lenskyj, H.J. (2018). The War on Doping. In *Gender, Athletes' Rights, and the Court of Arbitration for Sport* (pp.71 - 114). Emerald Publishing Limited.

Lentillon-Kaestner, V., Hagger, M. S., & Hardcastle, S. (2012). Health and doping in elite-level cycling. *Scandinavian journal of medicine & science in sports*, *22*(5), 596-606.

Lentillon-Kaestner, V., & Carstairs, C. (2010). Doping use among young elite cyclists: a qualitative psychosociological approach. *Scandinavian journal of medicine & science in sports*, *20*(2), 336-345.

Ling, Y., & Mahadevan, S. (2012). Integration of structural health monitoring and fatigue damage prognosis. *Mechanical Systems and Signal Processing*, *28*, 89-104.

Liu, Q., Liu, X., Jiang, B., & Yang, W. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC infectious diseases*, *11*(1), 218.

Ljung, G. and Box, G. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297-303.

Malcourant, E., Vas A., & Zintz, T. (2015). World Anti-Doping Agency: a meta-organizational perspective. *Sport, Business and Management: An International Journal*, Vol. 5(5), 451-471.

Manners, J. (1997). Kenya's running tribe. *The Sports Historian* 117(2), 14-27.

Mazanov, J., Backhouse, S., Connor, J., Hemphill, D., & Quirk, F. (2014). Athlete support personnel and anti-doping: Knowledge, attitudes, and ethical stance. *Scandinavian journal of medicine & science in sports*, *24*(5), 846-856.

Neuberger, E. W., Jurkiewicz, M., Moser, D. A., & Simon, P. (2012). Detection of EPO gene doping in blood. *Drug testing and analysis*, *4*(11), 859-869.

Omulo, I. (2018). Kalalei is third Kenyan to fail doping test in 10 days. Retrieved October 10, 2018 from https://af.reuters.com/article/sportsNews/idAFKBN1KS1MB- OZASP**.**

Pape, M. (2017). The fairest of them all: Gender-determining institutions and the science of sex testing. In *Gender Panic, Gender Policy* (pp. 177-200). Emerald Publishing Limited.

Pfister, G., & Gems, G. (2015). Fairy tales? Marion Jones, CJ Hunter and the framing of doping in American newspapers. *Sport in Society*, *18*(2), 136-154.

Pluim, B. M., de Hon, O., Staal, J. B., Limpens, J., Kuipers, H., Overbeek, S. E., ... & Scholten, R. J. (2011). β 2-Agonists and Physical Performance. *Sports medicine*, *41*(1), 39-57.

Pokrywka, A., Gorczyca, D., Jarek, A., & Kwiatkowska, D. (2010). In memory of Alfons Bukowski on the centenary of anti-doping research. *Drug testing and analysis*, *2*(11-12), 538-541.

Ratten, V., Ratten, H., Benijts, T., Lagae, W., & Vanclooster, B. (2011). The influence of sport leagues on the business-to-business marketing of teams: the case of professional road cycling. *Journal of Business & Industrial Marketing*.

Ritchie, I., & Henne, K. (2018). Amateurism, scientific control, and crime: historical fluctuations in anti-doping discourses in sport. *Journal of Criminological Research, Policy and Practice.*

Rosca, E. (2011). Stationary and non-stationary time series. *The USV Annals of Economics and Public Administration*, *10*(1), 177-186.

Ruiz, R.R., & Schwirtz, M. (2016, May, 13). Russian Insider Says State-Run Doping Fueled Olympic Gold. *New York Times*. Retrieved June 25, 2018 from http://www.nytimes.com/2016/05/13/sports/russia- doping-sochi-olympics-2014.html.

Sanctis, F.M. (2017). Olympic Games, Football Championships, and Corruption in the Sports Industry. *The Handbook of Business and Corruption*, 423 – 452

Sarpong, S. A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied*, *3*(1), 19-28.

Savulescu, J., & Foddy, B. (2014). Le Tour and Failure of Zero Tolerance. In *Enhancing Human Capacities*.

Seppelt, H. (2015). The secrets of doping: How Russia makes winners, ARD, YouTube, 27 April 2015. https://www.youtube.com/watch?v=iu9B-ty9JCY. Retrieved June 25, 2018.

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience, 35*(8), 3293-3297.

Zaiontz, C. (2019). Granger Causality. *Real Statistics Using Excel*. Retrieved August 15, 2019 from https://www.real-statistics.com/time-series-analysis/time-series-miscellaneous/granger-causality/

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Snyder, D., Fay, T., & DePerno, D. (2005). Interview with Richard Pound, Chair World Anti-Doping Agency, Former Vice-President International Olympic Committee. *International Journal of Sports Marketing and Sponsorship*, *6*(4), 4-10.

Sonksen, P. H. (2001). Hormones and Sport-Insulin, growth hormone and sport. *Journal of Endocrinology*, *170*(1), 13-26.

Sørensen, B.E. (2005). *Granger Causality*. Retrieved August 15, 2019 from http://www.uh.edu/~bsorense/gra_caus.pdf

Souillard, A., Audran, M., BRESSOLLE, F., Gareau, R., Duvallet, A., & Chanal, J. L. (1996). Pharmacokinetics and pharmacodynamics of recombinant human erythropoietin in athletes. Blood sampling and doping control. *British journal of clinical pharmacology*, *42*(3), 355-364.

Sumner, C. (2017). The spirit of sport: the case for criminalisation of doping in the UK. *The International Sports Law Journal*, *16*(3-4), 217-227.

Tak, M., Sam, M. P., & Jackson, S. J. (2018). The problems and causes of match-fixing: are legal sports betting regimes to blame? *Journal of Criminological Research, Policy and Practice*, *4*(1), 73-87.

Theorell, T. (2009). Anabolism and catabolism at work. In *Current Perspectives on Job-Stress Recovery* (pp.249 - 276). Emerald Group Publishing Limited.

Thörngren, J. O., Östervall, F., & Garle, M. (2008). A high-throughput multicomponent screening method for diuretics, masking agents, central nervous system (CNS) stimulants and opiates in human urine by UPLC–MS/MS. *Journal of Mass Spectrometry*, *43*(7), 980-992.

Union Cycliste Internationale (2017). UCI statement on Christopher Froome. Available at http://www.uci.ch/pressreleases/uci-statement-christopher-froome/. Retrieved June 17, 2018.

WADA (2015). WADA Independent Commission Report #1, 9 November 2015. Retrieved June 25, 2018 from https://www.wada-ama.org/en/resources/world-anti-doping-program/independent-commission-report-1.

WADA (2016). McLaren Independent Investigation Report – Part I, 18 July 2016. June 25, 2018 from https://www.wada-ama.org/en/resources/doping-control-process/mclaren-independent-investigation-report-into-sochi-allegations.

WADA (2016). WADA Independent Commission Report #2, 14 January 2016. Retrieved June 25, 2018 from https://www.wada-ama.org/en/resources/world-anti-doping-program/independent-commission-report-2.

Walker, D. M., & Eli Joubert, H. (2011). Attitudes of injecting male anabolic androgenic steroid users to media influence, health messages and gender constructs. *Drugs and Alcohol Today*, *11*(2), 56-70.

Wei, W. W. (2006). Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

Wheatcroft, J. (2016). It isn't all about the money: Do non-monetary awards really have an impact? *Human Resource Management International Digest*, *24*(5), 31-33.

Wilber, D. (1988). Linking sports and sponsors. *Journal of Business Strategy*, *9*(4), 8-10.

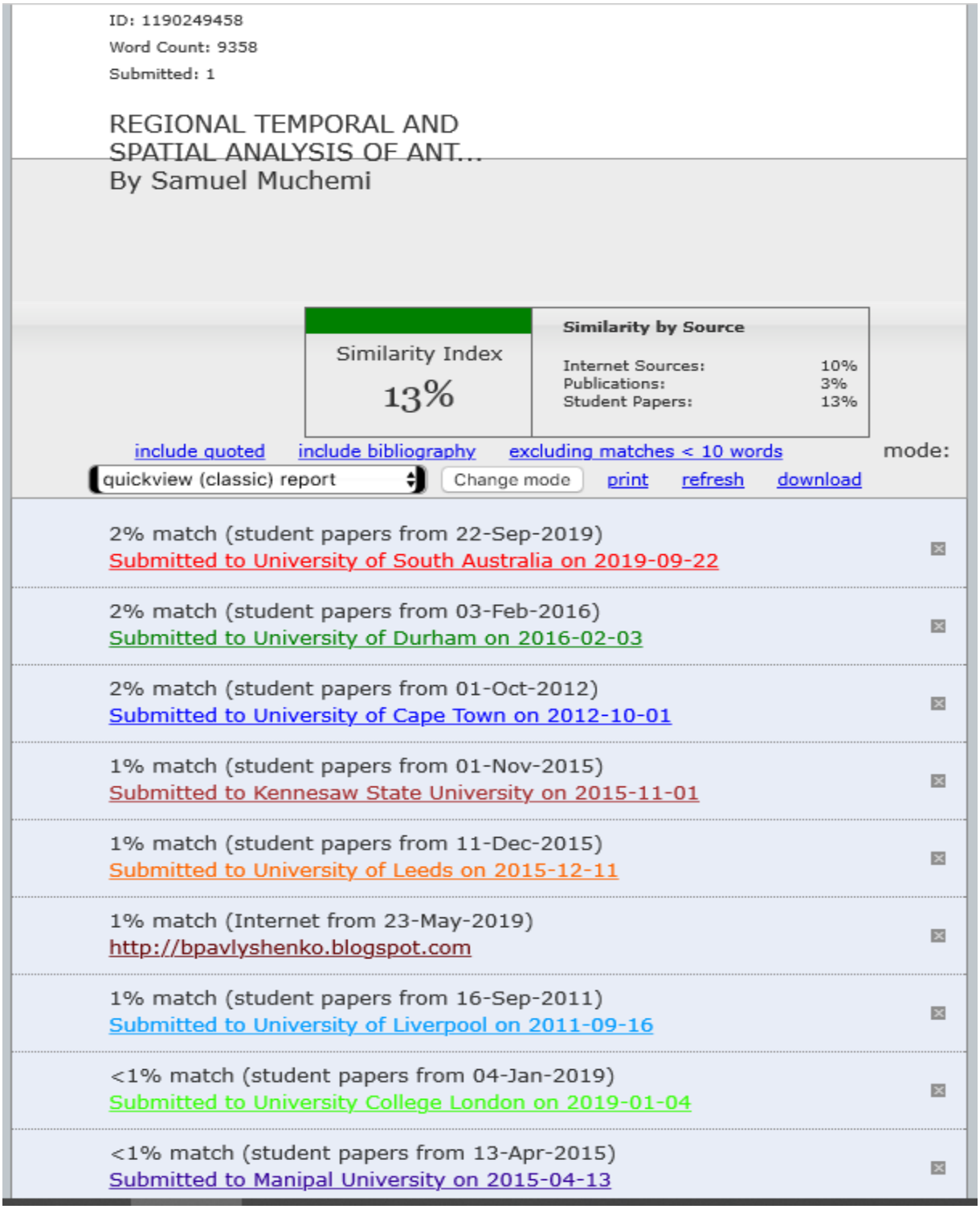
Winand, M., Dolles, H., Malcourant, E., Vas, A., & Zintz, T. (2015). World Anti-Doping Agency: a meta-organizational perspective. *Sport, Business and Management: An International Journal*, 5(5), 451-471.

Wintermantel, J., Wachsmuth, N., & Schmidt, W. (2016). Doping cases among elite athletes from 2000 to 2013. *Dtsch Z Sportmed*, *67*, 263-269.

Zhou, R., Kaplanidou, K., Papadimitriou, D., Theodorakis, N. D., & Alexandris, K. (2018). Understanding the inspiration among active participants in sport events. *International Journal of Event and Festival Management*, *9*(3), 332-348.

APPENDICES

Appendix A: Plagiarism Test Results


ID: 1190249458
Word Count: 9358
Submitted: 1

REGIONAL TEMPORAL AND SPATIAL ANALYSIS OF ANT...
By Samuel Muchemi

Similarity Index
13%

Similarity by Source
Internet Sources: 10%
Publications: 3%
Student Papers: 13%

include quoted include bibliography excluding matches < 10 words mode:
quickview (classic) report Change mode print refresh download

2% match (student papers from 22-Sep-2019)
Submitted to University of South Australia on 2019-09-22

2% match (student papers from 03-Feb-2016)
Submitted to University of Durham on 2016-02-03

2% match (student papers from 01-Oct-2012)
Submitted to University of Cape Town on 2012-10-01

1% match (student papers from 01-Nov-2015)
Submitted to Kennesaw State University on 2015-11-01

1% match (student papers from 11-Dec-2015)
Submitted to University of Leeds on 2015-12-11

1% match (Internet from 23-May-2019)
http://bpavlyshenko.blogspot.com

1% match (student papers from 16-Sep-2011)
Submitted to University of Liverpool on 2011-09-16

<1% match (student papers from 04-Jan-2019)
Submitted to University College London on 2019-01-04

<1% match (student papers from 13-Apr-2015)
Submitted to Manipal University on 2015-04-13

Appendix B: Ethical Consideration

UNIVERSITY OF NAIROBI
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KNH-UON ERC
Email: uonknh_erc@uonbi.ac.ke
Website: http://www.erc.uonbi.ac.ke
Facebook: https://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC https://twitter.com/UONKNH_ERC

KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref: KNH-ERC/A/442

15th November, 2019

APPROVED
15 NOV 2019
KNH/UoN-ERC

Samuel Karugu Muchemi
Reg.No.W62/88410/2016
Institute of Tropical and Infectious Diseases
College of Health Sciences
University of Nairobi

Dear Samuel

RESEARCH PROPOSAL: REGIONAL TEMPORAL AND SPATIAL ANALYSIS OF ANTI-DOPING RULE VIOLATIONS OF ATHLETES (P875/11/2019)

This is to inform you that the KNH- UoN Ethics & Research Committee (KNH- UoN ERC) has reviewed and **approved** your above research proposal. The approval period is 15th November 2019 – 14th November 2020.

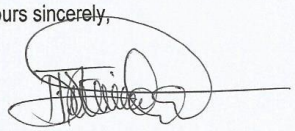
This approval is subject to compliance with the following requirements:

a. Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
b. All changes (amendments, deviations, violations etc.) are submitted for review and approval by KNH-UoN ERC before implementation.
c. Death and life threatening problems and serious adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH-UoN ERC within 72 hours of notification.
d. Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
e. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
f. Submission of an *executive summary* report within 90 days upon completion of the study. This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/ or plagiarism.

Protect to discover

For more details consult the KNH- UoN ERC website**http://www.erc.uonbi.ac.ke**

Yours sincerely,



PROF. M. L. CHINDIA
SECRETARY, KNH-UoN ERC

c.c. The Principal, College of Health Sciences, UoN
The Director, CS, KNH
The Chairperson, KNH- UoN ERC
The Assistant Director, Health Information, KNH
The Director, UNITID, UoN
Supervisors: Dr.Oscar Ngesa, Department of Mathematics, Statistics and Physical Sciences, (Taita Taveta University), Dr. Anne Wang'ombe, UNITID (UoN)

Appendix C: R-Codes

#Loading the required packages

```
library(tseries)
library(forecast)
library(spatial)
library(sp)
library(rgdal)
library(rgeos)
library(raster)
library(ggmap)
library (spatstat)
library(gstat)
library(lmtest)
library(sf)
library(rgdal)
library(tmap)
library(maps)
library(spdep)
```

#The First Objective Analyses

```
#Box-Jenkings Procedure
#Bloem
bloem<-read.csv(file.choose(), header = T)
```

```
bloem<-ts(bloem, start = c(2003), end = c(2016))
plot(bloem, main="Bloemfontein and Tunis ADRVs", ylab="ADRVs", type="o")
abloem<-auto.arima(bloem, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2bloem<- diff(bloem, differences = 2)
acf(d2bloem, main="Bloemfontein and Tunis ADRVs")
pacf(d2bloem, main="Bloemfontein and Tunis ADRVs")
plot(abloem$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(abloem$residuals, lag=5, type ="Ljung")
Box.test(abloem$residuals, lag=10, type ="Ljung")
fbloem<-forecast(abloem, h=5)
fbloem
plot(fbloem, ylab="ADRVs")
#######
#London
lon<-read.csv(file.choose(), header = T)
lon<-ts(lon, start = c(2003), end = c(2016))
plot(lon, main="London and Cambridge", ylab="ADRVs", type="o")
alon<-auto.arima(lon, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2lon<- diff(lon, differences = 2)
acf(d2lon, main="London and Cambridge ADRVs")
pacf(d2lon, main="London and Cambridge ADRVs")
plot(alon$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(alon$residuals, lag=5, type ="Ljung")
```

```
Box.test(alon$residuals, lag=10, type ="Ljung")

flon<-forecast(alon, h=5)

flon

plot(flon, ylab="ADRVs")

#####

#Paris

paris<-read.csv(file.choose(), header = T)

paris<-ts(paris, start = c(2003), end = c(2016))

plot(paris, main="Paris", ylab="ADRVs", type="o")

aparis<-auto.arima(paris, ic="aic", test = "adf", trace = TRUE, lambda = 0)

d2paris<- diff(paris, differences = 2)

acf(d2lon, main="Paris ADRVs")

pacf(d2paris, main="Paris ADRVs")

plot(aparis$residuals,main="Residuals Plot", ylab="Residuals")

Box.test(aparis$residuals, lag=5, type ="Ljung")

Box.test(aparis$residuals, lag=10, type ="Ljung")

fparis<-forecast(aparis, h=5)

fparis

plot(fparis, ylab="ADRVs")

######

#Los Angels and Salt Lake

los<-read.csv(file.choose(), header = T)

los<-ts(los, start = c(2003), end = c(2016))
```

```
plot(los, main="Los Angeles and Salt Lake ", ylab="ADRVs", type="o")
alos<-auto.arima(los, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2los<- diff(los, differences = 2)
acf(d2los, main="Los Angeles and Salt Lake  ADRVs")
pacf(d2los, main="Los Angeles and Salt Lake  ADRVs")
plot(alos$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(alos$residuals, lag=5, type ="Ljung")
Box.test(alos$residuals, lag=10, type ="Ljung")
flos<-forecast(alos, h=5)
flos
plot(flos, ylab="ADRVs")
####
#Montreal
mont<-read.csv(file.choose(), header = T)
mont<-ts(mont, start = c(2003), end = c(2016))
plot(mont, main="Montreal ", ylab="ADRVs", type="o")
amont<-auto.arima(mont, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2mont<- diff(mont, differences = 2)
acf(d2mont, main="Montreal  ADRVs")
pacf(d2mont, main="Montreal ADRVs")
plot(amont$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(amont$residuals, lag=5, type ="Ljung")
Box.test(amont$residuals, lag=10, type ="Ljung")
```

```
fmont<-forecast(amont, h=5)
fmont
plot(fmont, ylab="ADRVs")
####
#Rio
rio<-read.csv(file.choose(), header = T)
rio<-ts(rio, start = c(2003), end = c(2016))
plot(rio, main="Rio ", ylab="ADRVs", type="o")
ario<-auto.arima(rio, ic="aic", test = "adf", trace = TRUE)
d2rio<- diff(rio, differences = 2)
acf(d2rio, main="Rio  ADRVs")
pacf(d2rio, main="Rio ADRVs")
plot(ario$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(ario$residuals, lag=5, type ="Ljung")
Box.test(ario$residuals, lag=10, type ="Ljung")
frio<-forecast(ario, h=5)
frio
plot(frio, ylab="ADRVs")
#####
#Beijing
bei<-read.csv(file.choose(), header = T)
bei<-ts(bei, start = c(2003), end = c(2016))
plot(bei, main="Beijing ", ylab="ADRVs", type="o")
```

```
abei<-auto.arima(bei, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2bei<- diff(bei, differences = 2)
acf(d2bei, main="Beijing  ADRVs")
pacf(d2bei, main="Beijing ADRVs")
plot(abei$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(abei$residuals, lag=5, type ="Ljung")
Box.test(abei$residuals, lag=10, type ="Ljung")
fbei<-forecast(abei, h=5)
fbei
plot(fbei, ylab="ADRVs")
#####
#Sydney
syd<-read.csv(file.choose(), header = T)
syd<-ts(syd, start = c(2003), end = c(2016))
plot(syd, main="Sydney ", ylab="ADRVs", type="o")
asyd<-auto.arima(syd, ic="aic", test = "adf", trace = TRUE)
d2syd<- diff(syd, differences = 2)
acf(d2syd, main="Sydney  ADRVs")
pacf(d2syd, main="Sydney ADRVs")
plot(asyd$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(asyd$residuals, lag=5, type ="Ljung")
Box.test(asyd$residuals, lag=10, type ="Ljung")
fsyd<-forecast(asyd, h=5)
```

```
fsyd
plot(fsyd, ylab="ADRVs")
#####
#Moscow and Sochi
mos<-read.csv(file.choose(), header = T)
mos<-ts(mos, start = c(2003), end = c(2016))
plot(mos, main="Moscow and Sochi", ylab="ADRVs", type="o")
amos<-auto.arima(mos, ic="aic", test = "adf", trace = TRUE, lambda = 0)
d2mos<- diff(mos, differences = 2)
acf(d2mos, main="Moscow and Sochi  ADRVs")
pacf(d2mos, main="Moscow and Sochi ADRVs")
plot(amos$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(amos$residuals, lag=5, type ="Ljung")
Box.test(amos$residuals, lag=10, type ="Ljung")
fmos<-forecast(amos, h=5)
fmos
plot(fmos, ylab="ADRVs")
####
#Prize Money
pay<-read.csv(file.choose(), header = T)
pay<-ts(pay, start = c(2003), end = c(2016))
plot(pay, main="Prize Money in Dollars", ylab="ADRVs", type="o")
apay<-auto.arima(pay, ic="aic", trace = TRUE, lambda = 0)
```

```
d1pay<- diff(pay, differences = 1)
acf(d1pay, main="Prize Money in Dollars")
pacf(d1pay, main="Prize Money in Dollars")
plot(apay$residuals,main="Residuals Plot", ylab="Residuals")
Box.test(apay$residuals, lag=5, type ="Ljung")
Box.test(apay$residuals, lag=10, type ="Ljung")
fpay<-forecast(apay, h=5)
fpay
plot(fpay, ylab="ADRVs")
```

#Objective Two Analysis

```
top<-read.csv(file.choose(), header = T)
summary(top)
```

#Objective Three Analyses

```
#Granger Causality #Order 2
library(lmtest)
payb<-read.csv(file.choose(), header = T)
attach(payb)
grangertest(bloem~pay, order=2, data = payb)
paylon<-read.csv(file.choose(), header = T)
attach(paylon)
grangertest(lon~pay, order=2, data = paylon)
paypar<-read.csv(file.choose(), header = T)
attach(paypar)
```

```
grangertest(paris~pay, order=2, data = paypar)
paylos<-read.csv(file.choose(), header = T)
attach(paylos)
grangertest(los~pay, order=2, data = paylos)
paymont<-read.csv(file.choose(), header = T)
attach(paymont)
grangertest(mont~pay, order=2, data = paymont)
payrio<-read.csv(file.choose(), header = T)
attach(payrio)
grangertest(rio~pay, order=2, data = payrio)
paybei<-read.csv(file.choose(), header = T)
attach(paybei)
grangertest(bei~pay, order=2, data = paybei)
paysyd<-read.csv(file.choose(), header = T)
attach(paysyd)
grangertest(syd~pay, order=2, data = paysyd)
paymos<-read.csv(file.choose(), header = T)
attach(paymos)
grangertest(mos~pay, order=2, data = paymos)
```

#Objective Four Analyses

```
map1<- readOGR("C:/Users/user/Desktop/Data for Doping/E.Africabyyear",
"EasternAfrica_2013")
plot(map1, col="lightgrey")
```

```
map2013<-spplot(map1, "Mean", main="2013 Rado Zone V ADRVs")
map2013
map2<- readOGR("C:/Users/user/Desktop/Data for Doping/E.Africabyyear",
"EasternAfrica_2014")
map2014<-spplot(map2, "Mean", main="2014 Rado Zone V ADRVs")
map2014
map3<- readOGR("C:/Users/user/Desktop/Data for Doping/E.Africabyyear",
"EasternAfrica_2015")
map2015<-spplot(map3, "Mean", main="2015 Rado Zone V ADRVs")
map2015
map4<- readOGR("C:/Users/user/Desktop/Data for Doping/E.Africabyyear",
"EasternAfrica_2016")
map2016<-spplot(map4, "Mean", main="2016 Rado Zone V ADRVs")
map2016
require(gridExtra)
library(gridExtra)
grid.arrange(spplot(map1, "Mean", main="2013 Rado Zone V ADRVs"), spplot(map2, "Mean",
main="2014 Rado Zone V ADRVs"), spplot(map3, "Mean", main="2015 Rado Zone V
ADRVs"), spplot(map4, "Mean", main="2016 Rado Zone V ADRVs"))
#######
queen.nb<-poly2nb(map, row.names = map$Mean)
rook.nb<-poly2nb(map, queen = FALSE)
queen.listw<-nb2listw(queen.nb, style = "B")
```

```
rook.listw<-nb2listw(rook.nb)

listwl<-queen.listw

names(map)

summary(queen.nb)

#wasnt working coz ZERO was replaced by NA in the map$Mean

moran(map$Mean,listwl, length(map$Mean), Szero(listwl))

moran.test(map$Mean, listwl)

map$Mean

#this changed NA to ZERO

map$Mean[which(is.na(map$Mean))]<-0

# Geary C

queenz.listw<-nb2listw(queen.nb, style = "W")

qz<-queenz.listw

geary.test(map$Mean, qz)
```