ARIMA MODELS FOR FORECASTING OWN SOURCE REVENUE FOR THE COUNTY OF MACHAKOS

SUBMITTED BY:

JOHN WAMBUA KIOKO

REG NO.: 19/05485

SUPERVISOR:

DR. MWENDIA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DATA ANALYTICS IN THE FACULTY OF COMPUTING AND INFORMATICS AT KCA UNIVERSITY

OCTOBER 30TH 2021

DECLARATION

I declare that this dissertation is my original work and has not been previous published or submitted elsewhere for the award of a degree. I also declare that this contains no material written or published by other people except where due reference is made and author duly acknowledged.

Student Name: JOHN WAMBUA KIOKO Reg. No: 19/05485

Sign: Date: 30TH OCTOBER 2021

I do hereby confirm that I have examined the master's dissertation of

John Wambua Kioko

And have approved if for examination.

Sign: Date:

[Name of Supervisor]
Dissertation Supervisor

ABSTRACT

Fiscal stresses have created a need for County Governments in Kenya to use accurate Own Source Revenue figures in budgeting for good economic planning. This is because they are faced with unlimited demands from taxpayers coupled with limited tax or revenue resources. Counties have not adopted a formal any quantitative forecasting technique leading to forecasting errors and disruption in government services. Therefore, forecasting is becoming increasingly relevant and essential in the context of county governments in Kenya. Forecasting is not only a legal requirement, but an essential tool for fiscal planning. This study therefore aims at developing an autoregressive integrated moving average model and using it to forecast own source revenue for the County Government of Machakos and then comparing the forecast results with prediction generated using the expert judgment approach which is the current in-house forecasting technique in the county. The findings show that the autoregressive integrated moving average method generated forecasts with a higher level of accuracy than those generate through the expert judgment approach. This study uses own source revenue data from the financial year 2013/2014 to 2019/2020 while the mean absolute percentage error is used as the measure of accuracy. This study recommends adoption the Autoregressive Integrated Moving Average models for forecasting of Own Source Revenue for Machakos County.

ACKNOWLEDGEMENT

I would like to thank my Dissertation Supervisor Dr. Mwendia and the various Dissertation defense panels for helping me through the proposal and dissertation writing process with your advice, expertise and guidance. I would also like to thank my colleagues from the budget department of the County of Government of Machakos for their help in acquiring the data. I hope this forecasting model proves valuable to you.

This dissertation is dedicated to my brilliant and supportive wife for keeping me on track. Also my son, for being my support system and my mother who pushed me to enroll for the course.

ACRONYMS AND ABBREVIATIONS

ARIMA	Auto Regressive Integrated Moving Average
BPS	Budget Policy Statement
CBROP	County Budget Review and Outlook Paper
CEC	County Executive Committee Member
CoB	Controller of Budget
CoK	Constitution of Kenya
CRA	Commission of Revenue Allocation
CRF	Consolidate Revenue Fund
FY	Financial Year
GCP	Gross County Product
GNP	Gross National Product
ICT	Information Communication Technology
IGRTC	Inter-Government Relations Technical Committee
KACC	Kenya Anti-Corruption Commission
KRA	Kenya Revenue Authority
LATF	Local Authority Transfer Fund
OSR	Own Source Revenue
PFM	Public Financial Management
RoK	Republic of Kenya
VAR	Vector Autoregressive Model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1. Background of the study

The 2010 Kenya constitution changed the structure of governance in the Republic of Kenya and introduced the devolved system of government. The system comprises of a central government and 47 county governments, all which function autonomously except as stated in the constitution. Before enactment of the 2010 constitution, local authorities were a representation of the devolved system of government. They were anchored in the Local Authorities Act (Cap 265) as opposed to the constitution where the counties governments are anchored. In total there were 175 local authorities in Kenya which included municipal, county or town and urban councils. They derived their powers to raise revenue from various legislations including, Valuation for Rating Act, Rating Act, Trade Licensing Act and the Local Government Act. Despite these enabling legislations, local authorities were persistently and consistently in financial distress. Reforms saw them lose significant ability to raise their own revenue. The Transfer of Function Act (1969) for example took away the primary health and health services from the local authorities and the ability of municipalities to levy the Graduated Personal Tax (National Treasury, 2019). Local authorities had to borrow in order to settle short term obligations such as salaries (Omamo, 1995). Ntoiti (2013) found that both Nairobi and Mombasa City Councils owed creditors a combined total of 7.2 Billion Kenya shillings in their 2011 budgets. This affected their ability to offer good, quality and efficient services to the residents of the two cities. Omamo (1995) noted that revenue deficits contributed significantly to the financial distress experienced by the local authorities. This was corroborated by Ntoiti (2013) who concluded poor financial management practices and a host of other factors are a cause to the financial mess in local authorities. Waema & Mutullah (2006) were of the view that lack of adoption of ICT contributed to the poor financial management in local authorities. Even though ICT systems were introduced in a few local authorities to enhance the financial management, they were rated to be inadequate (Ntoiti, 2013).

Revenue deficits in the local authorities were likely caused by over ambitious targets or leakages of the revenues as most systems for collection of revenue were not automated (KACC, 2007). In response to the large number of local authorities in financial distress, Local Authority Transfer Fund (LATF) was established as a life line. However, instead of the fund acting as a measure of last resort, some local authorities become entirely dependent on the fund (National Treasury, 2019).

1.1.1. Post devolution

With the advent of the new constitution, counties took over the defunct local authorities. A consolidate report on assets and liabilities of the defunct local authorities by Intergovernmental Technical Relations Committee found that the total liabilities inherited by the County Governments was Kshs 53,756,398,754.25 (IGRTC, 2018). Nairobi County inherited the largest liability accounting for 67.68% of the total inherited liabilities while Mombasa County inherited the second largest liabilities amounting to 8.31% of the total inherited liabilities. Machakos County inherited liabilities worth 409,041.853.00 which was 0.76% of the total inherited liabilities. To balance the needs of the county governments, service delivery to the citizens and clearing debts it is imperative that county budgets and revenue forecasts be based on realistic estimates that reflect the prevailing economic conditions and circumstances. Wrong estimates run the risk of aggravating the financial burden inherited from the defunct local authorities as well as compromise service delivery.

Besides the liabilities, the counties also inherited some of the very traits the bedeviled the local authorities particularly weak financial management procedures and especially poor revenue collection, management and forecasting. According to the Commission on Revenue Allocation (2021), all counties with an exemption of those with game parks were collecting less than 40% of their estimated revenue potential. In their inaugural report on efforts by counties towards revenue mobilization, the Commission on Revenue Allocation recognized that revenue potential differs across counties (CRA, 2021). Counties with large economic sizes and economic diversification collect more revenue compared to counties with narrow economic diversification. Generally, counties with a higher share of agriculture in their economy recorded lower revenue collections whereas those with a more diversified economic base collected high revenue.

The County Governments' sources of funding as defined in the constitution include: Equitable share of at least 15% of most recent audited revenues raised nationally (Article 202(1) and 203(2)), conditional and unconditional grants from the national government (Article 202(2), Equalization fund (Article 204), Local revenues (Own Source Revenue (OSR)) and Loans & grants

This study focuses on Own Source of Revenue (OSR) for the County Governments. OSR constitutes of revenue from property taxes, entertainment taxes, charges levied for the services they provide and any other tax authorized by an Act of Parliament. Own source revenue streams differ markedly across the 47 counties. According to the Controller of Budget Report (2016), Machakos County had 17 own source revenue streams, Makueni County 11, Kitui County 11, Nairobi County 15 and Mombasa County 10. A high number of own source revenue streams may be an indication of a more diversified economic base and vice versa. The naming of the own source revenue streams is also unique in the 47 counties. This confirms that there is no particular standard for the county governments to follow while naming a particular OSR stream. Such disparities cause inconsistency across counties in the reporting of OSR. To address this challenge, the National Treasury issued a financial statement template to be used by County Governments in reporting their OSR but its adoption has not been adhered to (National Treasury, 2019). While some OSR streams cut across all the 47 counties such as business permits, others are unique to particular counties. For example while Machakos has a game park fee and quarry OSR streams, the county of Makueni has no such OSR streams. Also, Makueni County lacked a land rates revenue stream while Machakos, Nairobi and Kitui Counties all have the land rates OSR stream. The uniqueness of each county own sources of revenue makes it impossible to develop a common forecasting formula applicable to all the 47 counties. Also the huge number of different own source revenue streams makes it difficult to have a revenue prediction model that includes all of them with forecasted OSR as the dependent variable.

1.1.2. The budget process

The formulation and preparation of the budget involves development and submission of key documents for approval by the county assembly. The process is guided by the budget calendar which stipulates timelines for a number of key activities to be undertaken in order to finalize the budget and submit it for approval by 30th April of each financial year. The budget process for the County Governments is an elaborate one. It is guided by the Constitution, the Public Finance Management Act, Public Finance Management Regulations Act and the County Governments Act. The county assemblies play a huge role in the public finance management at the county level. They: approve the budget expenditure of the county government, approve borrowing by the county government, receive and approve plans and policies, exercise oversight over the county executive and other county executive organs and make necessary laws for effective performance of the functions and exercise of the powers of the county government. The county assemblies are charged with the role of oversight and in particular revising and amending revenue estimates tabled by the county executive. Between May 1st and June 30th, the budget committee for the county government conducts public hearings on the budget proposal or estimates. When the county executive member for finance tables the county budget estimates to the county assembly, the assembly could approve the estimates with or without amendments. After consideration and approval of the budget estimates, the county assemble authorizes the county executive to withdraw funds from the County Revenue Fund (Article 207) by passing the Appropriations Act. As discussed earlier, sources of county revenue are clearly set out by the constitution of Kenya. All revenue sources with an exemption of own source revenue are not within the purview of the county revenue collection and management. This is to mean that the precise amounts to be allocated from nationally raised revenue are already known. The Commission for Revenue Allocation is charged with the responsibility of designing a basis for sharing of the nationally raised revenues between the national and county governments. Besides considering and approving the budget estimates, the county assembly should satisfy itself that the county budget contains: estimates of revenue and expenditure, differentiating between recurrent and development, proposals for financing if any anticipated deficit for the period to which they apply and proposal regarding borrowing and other forms of public borrowing that will increase public debt.

Considering the above, the county assemblies' role in regards to considering estimated revenues, are left with only the option of considering OSR estimates. The county assemblies are supposed to ensure the OSR estimates are realistic and feasible. The National Policy that supports OSR in the counties notes that counties are unable to meet their revenue targets in part due to unrealistic targets. Most counties do not include detailed revenue forecast in their County Budget Review and Outlook Papers (CBROPs) in line with the PFM Act, 2012 (National Treasury, 2019). The national policy to support enhancement of OSR in counties proposes to strengthen the oversight role of county assemblies through capacity building through the Center for Parliamentary Studies and Training.

1.1.3. The Case for Machakos County

Machakos County took over functions of several local authorities. They are: Mavoko Municipal council, Machakos Municipal council, Tala Municipal council and Matuu town council. The county's average own source revenue collection has averaged more than one billion annually. The CRA classifies Machakos as one of the sixteen counties that had moderate revenue growth for the period 2013/2014 to 2018/2019(CRA, 2021). Since the inception of devolution, the county has consistently collected over a Billion Shillings in OSR revenue. Commission on Revenue Allocation (2021) notes that four factor seem to influence the amount of own source revenue collected by the counties. They are: agriculture, games parks, urban and peri-urban status and improved revenue collection efficiency. Agricultural cess is a major revenue stream in rural counties but accounts for only 3 per cent of the total county revenue. Games parks were found to confer significant advantage to the counties that managed them. Counties that were grouped and either urban or peri-urban had more diverse economies compared to the rural economies which relied on agriculture. Regarding the revenue collection systems, counties that had overhauled and automated their revenue collection and management systems had significantly higher owns source revenue collections than those that had not. Machakos is ranked among the six counties of: Narok, Nakuru, Nairobi, Mombasa and Kiambu with the most diversified economies. In the financial year 2019/2020, Machakos was one of only five counties that exceeded their targeted OSR revenue collection for the period (National Treasury, 2020). In terms of revenue collection as a proportion of revenue size, Machakos County was found to collect own source revenue proportional to it economic size (CRA, 2021). The threshold for financing own budgets is set at 15% implying anything below 15% signifies an over dependence on transfers from the National Government. Machakos county was found to finance slightly above 15% of their budget through own source revenue. For financial year 2019/2020, actual OSR as a percentage of targeted (predicted) OSR was 118.60%. Actual OSR surpassed the targeted OSR by 18.6%. This may be attributed to an increase in the fees and charges as captured in the Machakos County Finance Act 2019. The 2019 Machakos

County Finance Act had two fees and charges schedules, schedule 8 and 5. Schedule 8 specifies fees and charges for urban areas while schedule 5 specifies fee and charges for peri-urban and rural. The fees and charges for schedule 5 are lower than those of the 8th schedule. Despite the schedules being designated as for peri-urban /rural and urban, the billing clerks exercised their discretion on which schedule to use without regard to the whether the area is urban or rural/peri-urban. This grey area created a loophole where fees and charges could be underpaid. To clear this ambiguity, the Finance Act 2020 did away with the 5th schedule and retained the 8th schedule only. As a result, the OSR for the financial year 2020/2021 might reflect an increase.

Table 1 shows actual and targeted annual OSR revenue for the financial years 2013/2014 through to 2019/2020 and the forecast error.

Table 1: Actual and Target OSR collection and Forecast Error

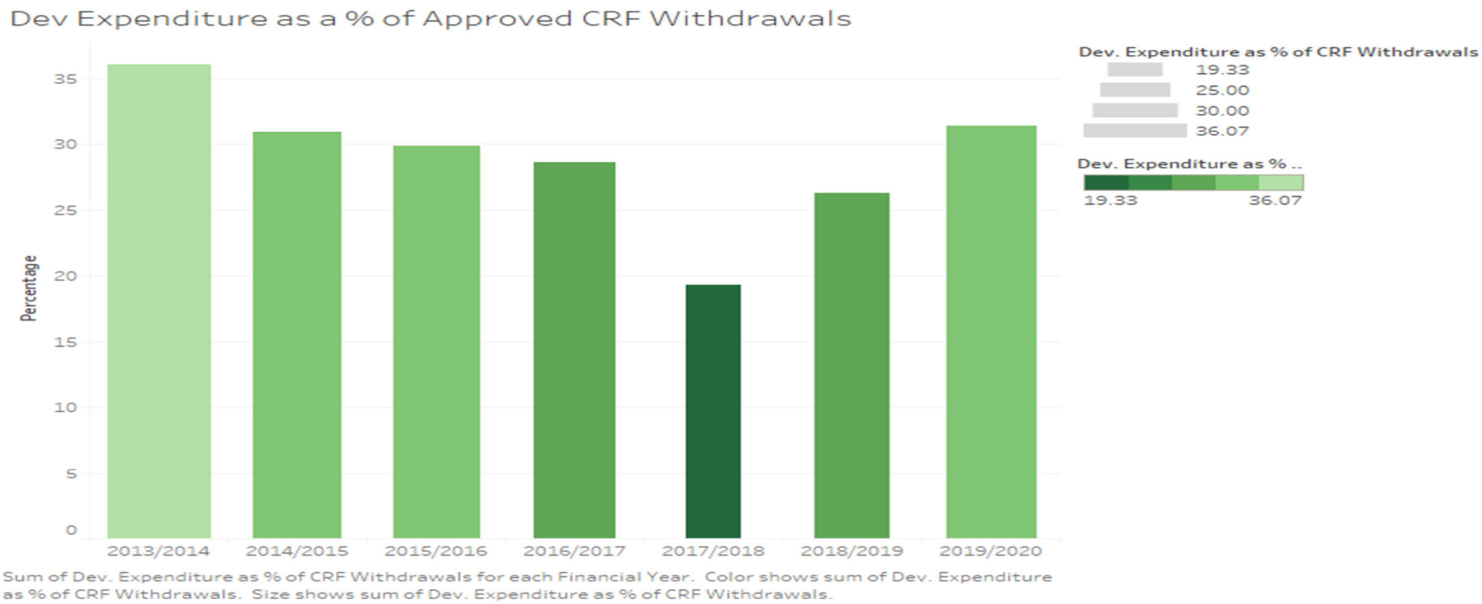
Financial Year	**Actual OSR (Kshs Billions)**	**Predicted OSR(Kshs Billion)**	**% Error in OSR forecast**
2013/2014	1,175,300,000.00	2,511,300,000.00	113.67%
2014/2015	1,356,559,888.00	2,850,000,000.00	110.09%
2015/2016	1,121,680,950.00	2,371,633,578.00	111.44%
2016/2017	1,259,304,944.00	2,861,623,481.00	127.24%
2017/2018	1,063,726,784.00	1,594,386,715.00	49.89%
2018/2019	1,557,229,789.00	1,720,061,674.00	10.46%
2019/2020	1,376,170,000.00	1,160,780,000.00	(18.56%)

Source: CoB, Kenya.

The declining trend of targeted or predicted OSR from financial year 2016/2017 to 2019/2020 may be as a result of the budget officers setting more realistic targets or lowering their expectations over time.

Since the transition from local authorities to county governments, there has been significant progress in the financial management. However inability to make realistic projection of own source revenue continues to persist. This limits the ability of the county to accurately plan as well as hinder effective service delivery to county residents. Inability to make accurate revenue prediction also affects the ability of the counties to comply with public financial management laws. The Public Financial Management Act requires that a minimum of 35% of county budget be allocated to development expenditure. Analysis of the development expenditure of the County of Machakos shows that financial year 2013/2014 was the only year that the county exceeded this threshold. This is shown in figure 1. Budgets based on inaccurate prediction will affect ability of counties to comply with this provision of the law.

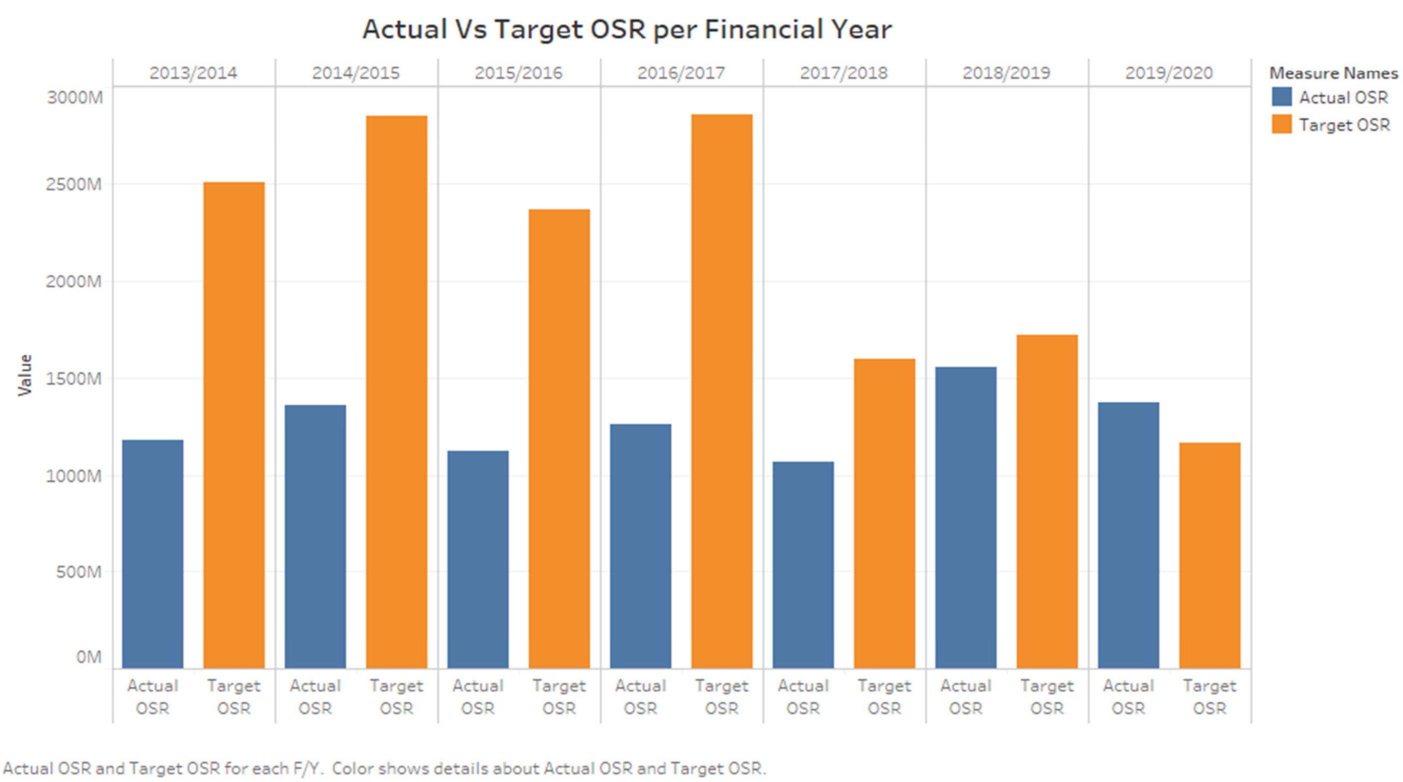
Figure 1: Development expenditure as a percentage of approved CRF withdrawals



Source: Author's Computation

Figure 2 shows actual OSR against targeted OSR for the financial years 2013/2014 through to financial year 2019/2020.

Figure 2: Actual vs. Targeted OSR collection



Source: CoB Reports

1.1.4 ARIMA Models

ARIMA processes are a class of stochastic processes used to analyze time series. It is a multidisciplinary scientific tool used in prediction problems. ARIMA models use information obtained from the variable itself to forecast it. It uses the philosophy "let the variable speak for itself". Saayman and Botha (2015) acknowledged that univariate time series models like the autoregressive integrated moving average and seasonal autoregressive integrated moving average produce accurate forecasts but have the limitations in explaining influential external forces. Box and Tiao (1975) looked in to the question of how many observation were need to generate a forecast with the autoregressive integrated moving average models and recommended at least 50 observations.

ARIMA model generally have a good fit and are easier to interpret than some of the other forecasting models (Du Preez and Witt, 2003). There underlying assumption of ARIMA modeling. The first assumption is stationarity for AR model. A series must be stationary. A series is deemed to be stationary if: it exhibits mean reversions, has a finite and time-invariant variance and has a theoretical correlogram that diminished as the lag length increases. The second assumption is invertibility for MA models. This implies that: the series can be represented by a finite order MA or convergent autoregressive process uses the autocorrelation function (ACF) and partial autocorrelation function (PACF) for identification and implicitly assumes that the series can be approximated by an autoregressive model.

The ARIMA/SARIMA model is fitted on stationary data hence if data is non-stationary; it has to be transformed to be stationary. The Augmented Dicky-Fuller (ADF) test is used to check stationarity. The Box-Jenkins method is an iterative process that involves four stages: identification, estimation, diagnostic checking and forecasting of the time series (Wabomba et al 2016, Larmore, 2016). Farhath et al (2016) eliminated the fourth stage in building of ARIMA models. Its implementation is easy and flexible because it only requires historical observation of the variables of interest. Larmore (2016) found ARIMA models to be the most appropriate in instances where the data pool is small or insufficient. Differencing is used to achieve stationarity. A non-seasonal ARIMA model can be written as "ARIMA (p,d,q)" model where: 'p' is the number of autoregressive terms, 'd' is the number of differences needed for stationary 'q' is the number of lagged forecast errors in the prediction equation (Gujarati, 1995)

Seasonal ARIMA models exist for seasonal time series data which exhibit seasonal components. In most time series data, there are elements of seasonality and it is there important to include them in forecasting. A SARIMA (p,d,q)(P,D,Q) model captures both the non-seasonal and seasonal component (Lim & McAleer, 1999). The parameters P indicate the seasonal AR component, D, the seasonal order of integration and Q specifies the seasonal MA component.

1.2.Statement of the Problem

The relationship between income and expenditure of government is the core issue of the budget. Budget deficit and budget surpluses are a common feature in budgeting. They are caused by under and over prediction of revenue forecasts respectively. A budget deficit occurs as a result of expenditure surpassing revenues which can as a result from poor revenue forecasting. On the other hand, a budget surplus occurs as a result of revenue exceeding expenditure. Deficits necessarily result in rising scaling back of development expenditure in favor of recurrent where borrowing is restricted. A challenge facing the county government of Machakos is the persistent delays in the disbursement of funds from the National Government. This causes deficits that negatively affect the operations in the county. Consequently, focus has then shifted to local revenues also known as Own Source Revenue.

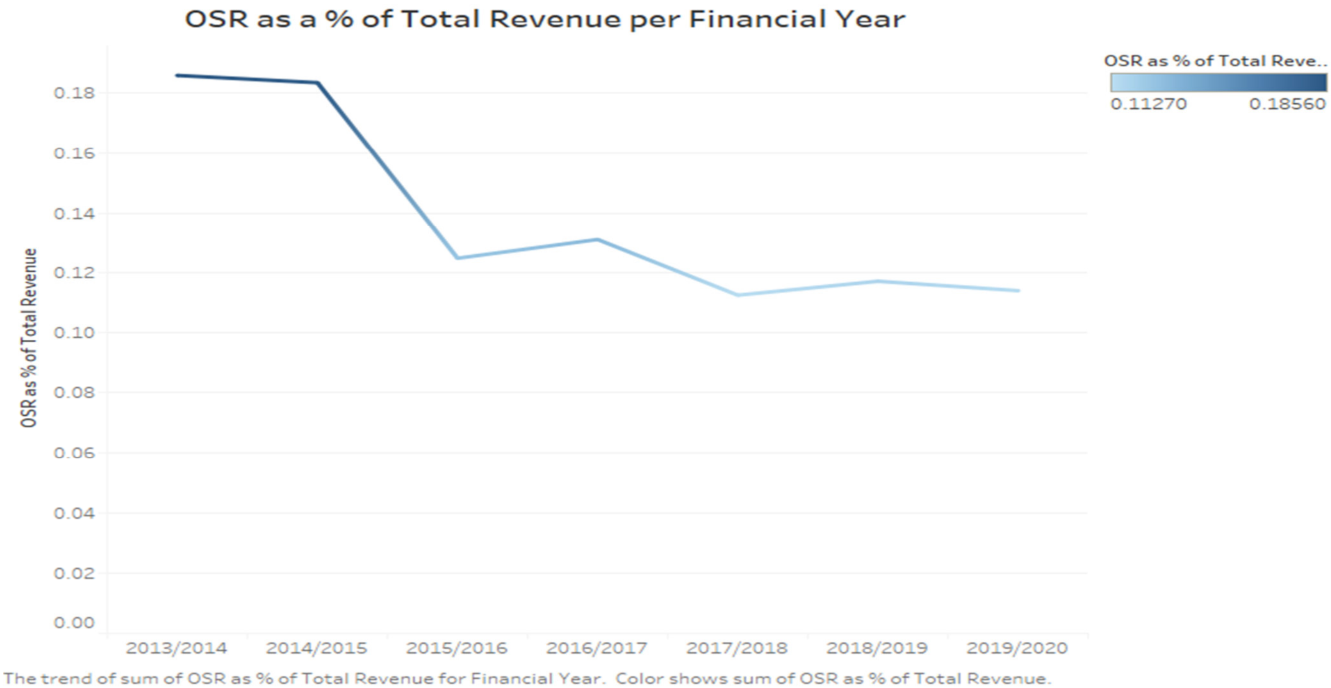
The problem is Machakos County's failure to accurately forecast own source revenue resulting to over ambitious or grossly understated revenue targets. The County does not currently employ an empirical forecasting model and its forecasts are based on expert judgment. This has resulted in high forecasting errors as shown in figure 4. As a consequence it has led to overdependence on the Equitable Share of Revenues from the National Government resulting in service disruption when there are delays in disbursement of funds. Figure 3 shows the county's OSR as a share of total revenue per year steadily declining over the years. At the onset of devolution, it was at 18.56% but has reduced to stand at 11.27% in the 2019/2020 financial year. In 2020 for example county governments shut down services and sent staff home over delay in disbursement of funds occasioned by Senates' delay in resolving a revenue sharing stalemate. In order to reduce the overdependence on Equitable Share of Revenues from the National Government and improve service levels at county level, realistic targets need to be set based on accurate forecasting of the OSR revenues.

Figure 4 shows the percentage forecasting error for Machakos County. The percentage error reduced marginally FY 2014/2015, increased thereafter reaching peak in FY 2016/2017 and declined sharply thereafter to record a surpass of the target revenue by 18.56% in the FY 2019/2020. Unmet revenue targets have resulted in the county failing to meet the legal requirement to allocate at least 35% of the budget to development.

In Ayakeme et al (2021), they used an ARIMA model to forecast total revenue for Bayesla state. This approach lumped together all revenue streams to generate a single model. In this study, the total revenue is disaggregated and the leading revenue sources individually forecasted. This improves the accuracy and also factors the dynamics of each source in the model. Larmore (2016), acknowledged that it was difficult to create a single prediction model for total revenue due to multiple revenue sources and resulting variables. In his study, he focused on only two revenues sources that contributed over 50% of the revenues. Although this study uses the same approach, four revenue streams are considered instead of two. In Ahmed et al (2020), no accuracy measure are given. This makes it difficult to determine the validity of the model. In this study the mean absolute percentage error and Akaike Information Criterion are used to determine the validity and accuracy of the model. Li et al (2005) found that the length of the forecasting horizon affects the accuracy of the results. This is because of the possibility of unforeseen events like changes is tax rates, tax waivers and economic shocks such as that caused by the covid 19 pandemic. Ahmed et al (2020) used data covering a 20 year period. Pelinescu et al (2010) used data covering a period of 10 years. In Jayesekara and Passty (2009) the data initially covered the period from 1970 to 2009 but was readjusted to cover the period from 1989 to 2009 due to changes in tax rates. This study covers the period from the financial year 2013/2017 to financial year 2018/2019. This period was relatively stable without any tax rate changes or tax waivers. Again, the in the period no economic shocks were experienced. Wong (1995) reviewed the use of regression and econometric methods in forecasting of revenue. He noted that the methods may range from simple regression to multivariate econometric models. In econometric models, several independent variables are used to forecast an independent variable. This approach differs from the ARIMA approach in that, the ARIMA model does not require any other data other than the historical data of the variable to be forecast. Wong also noted the response between the dependent variable and the independent variable is likely to change over time. Stewart et al (1989) in analyzing the use of expert judgment in a hail forecast acknowledged that the human information processing system is the least understood yet probably the most important component of forecasting accuracy. They concluded that the intuitive processes that weather forecasters use to aggregate information into a forecast can be analyzed and descried in quantitative terms. Bolger and Wright (1994) concluded that judgment performance is a function of interaction between the two dimensions of ecological validity and learnability. If both were high then performance will be good but if one is low then performance will be poor. Neural networks are mathematical models inspired by the functioning of biological neurons. Several studies (Connor 1988, Hornik et al. 1989, Wassermna 1989, White 1992) have argued that neural networks overcome the problem of outliers and the need for expertise knowledge of traditional time series methods making them superior (Hill et al. 1996). Hill et al (1996) neural network forecasts with exponential smoothing method, the Box-Jenkins method and the expert judgment method. The study concluded that neural network models were significantly better than the other three models. However the conclusion was based on quarterly and monthly data. With annual data, all four models were found to be sufficient. Sharda and Patil (1990) compared the prediction accuracy of neural networks and automatic Box-Jenkins methods. Both methods were found to perform well.

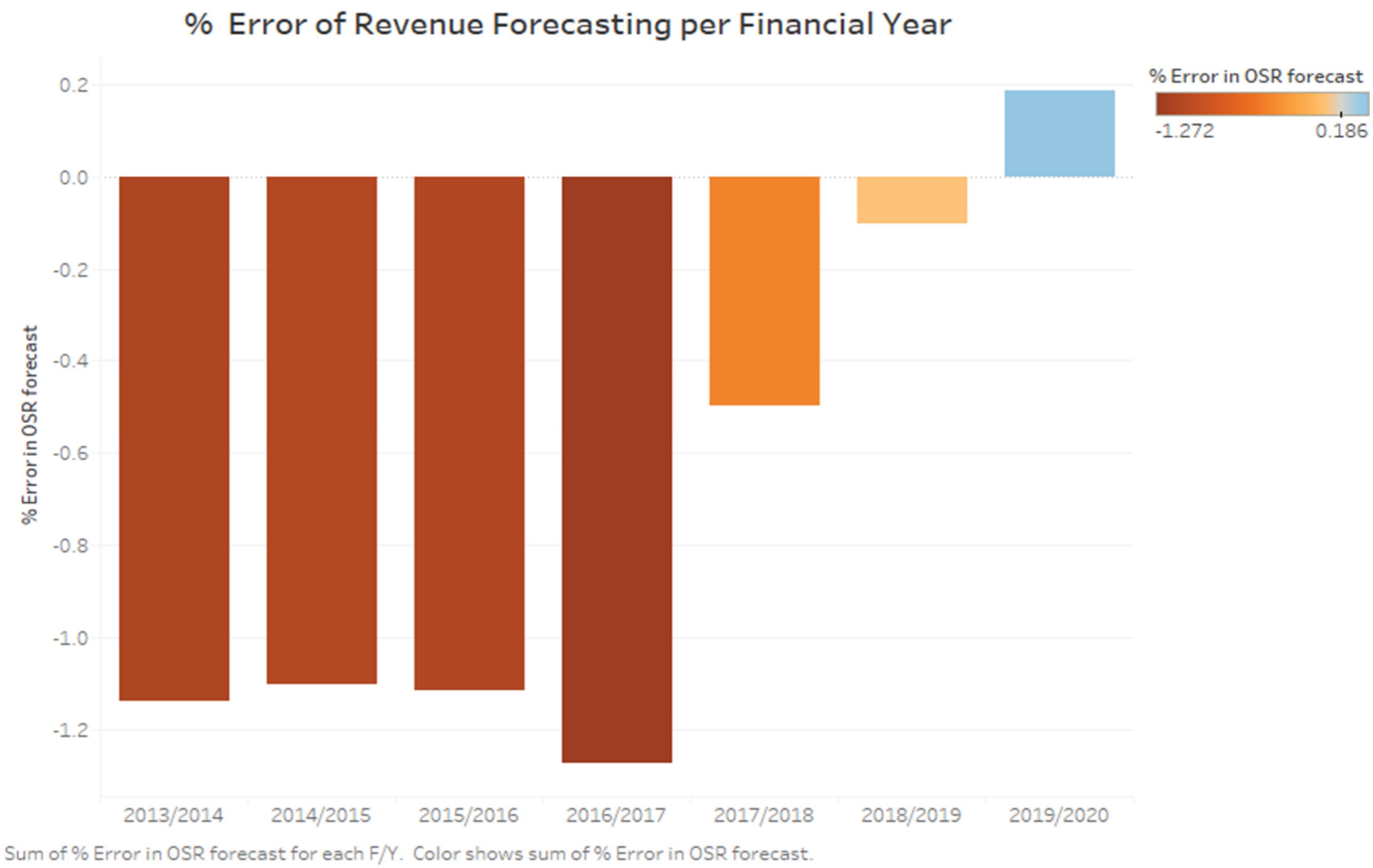
This ARIMA models have not been used in the County context in Kenya to forecast Own Source Revenue. This study therefore aims at developing an ARIMA own source revenue forecasting model for Machakos County and compare the accuracy of the forecasts against those of the current model in use by the county of Machakos. If the resultant ARIMA model proves superior with better forecasting accuracy, then it will help in setting realistic budgets estimates, properly plan service provision and ensure uninterrupted service provision.

Figure 3: OSR as a % of Total Revenue per Financial Year.



Source: CoB Reports

Figure 4: Percentage Error of Revenue Forecasting Per Financial Year.



Source: Author's Computation

1.3.Main Objective

The main objective of this study is to develop ARIMA models that can accurately predict OSR for the County of Machakos.

1.4. Specific Objectives

The specific objectives of this study are:

i. To investigate and identify own source revenue streams that greatly influence own source revenue
ii. To develop ARIMA models for the identified streams to predict own source revenue
iii. To evaluate the developed models.

1.5. Research Questions

i. Which OSR streams contribute percentage of total own source revenue?
ii. Which are the most suited ARIMA models for forecasting?
iii. How effective are the developed model?

1.6.Significance of the study

This study not only adds to the existing rich literature on revenue forecasting, but it also adds literature to the novel niche that is empirical forecasting of OSR in the Counties in the Republic of Kenya. It is hoped that the Machakos County will adopt the model for future forecasting of OSR due to the models ability to give accurate revenues forecasts. Other counties can adopt the model too but modify it to suit their own unique own source revenue dynamics. It is hoped that this study will also inform future policy action by the National Treasury and Commission on Revenue Allocation as regards revenue forecasting in the Counties and revenue allocation to the counties and address the challenge that is forecasting of revenue in counties.

1.7. Motivation of the study

The expert judgment approach to forecasting has results in forecasting errors resulting to constant disruption of critical and social services such as healthcare, and emergency responses to the detriment of the county residents. Additionally, instances of employee go slow and delayed salaries are also prevalent. Inaccurate forecasting has also led to decline development expenditure and over reliance on national transfers. These instances have served as a motivation for my study. This study intent is to enhance the forecasting accuracy of own source revenue and subsequently help the County of Machakos meet its obligations to the residents of the county and enhance its fiscal planning ability. The impulse to use the ARIMA model is derived from the fact that it is relatively simple and does not require additional variables to use.

CHAPTER TWO
LITERATURE REVIEW

2.1. Introduction

The objective of this chapter is to conceptualize the foundation upon which this study has been built on. It looks at the theories and previous works on revenue forecasting and forecasting in general. Literature on revenue forecasting in Kenya particularly in small jurisdictions such as local authorities and counties could not be found. Therefore literature on the same from similar jurisdictions was used. The literature gives insight on forecasting approaches and models, their accuracy and how the same can be applied in the context of Machakos County. Both qualitative and quantitative approaches to revenue forecasting methods are analyzed.

2.2. Theoretical review

Revenue forecasting helps in setting performance targets for revenue departments. Buetter and Kauder (2010) looked at the revenue forecasting practices: differences across countries and consequences for forecasting performance. They found that while the mean forecast errors were small, the precision of the forecasts measured by the standard deviation of the forecasts errors differed substantially across countries. Tax structure and timing differences contributed to differences in forecasting practices. Also, the independence of revenue forecasting and use macroeconomic models was associated with a lower standard deviation of the forecast error. In the book 'Tax Analysis and Revenue Forecasting – Issues and Techniques', it is noted that there are common methodologies and assumptions that are a prerequisite to forecasting. These assumptions are made relying on the most recently available information, and on what is assumed at the time for such economic variables as growth in the national income, the rate of inflations, interest rates and others (Jenkis et al, 2000). Jenkins and his peers noted that there are several steps involved in the preparation forecasts of tax revenues. They are: evaluation of tax elasticity, evaluation of changes in economic conditions and evaluation of the effects of inflation and prices changes.

There is no one single method of revenue forecasting is consistently more accurate than others (Mc Nichol, 2014). In his research, MC Nichol instead identified other important advantages for forecasting revenue in a professional, transparent and inclusive manner. He theorized that the basic features of a well-designed revenue forecasting process for States in America to include: the Governor and the legislature to jointly produce revenue estimates, forecasting body to include outside experts, the revenue estimating process be open to all interested parties and estimates be revised during budget sessions.

Revenue forecasting is as much an art as it is a science (Sun, 2005). Human relationships and both individual and group judgment influence the estimation process and therefore final forecast. Sun and lynch (2008) saw generating accurate forecast as not only being dependent on institutional processes and structures but also on human relationships, judgment and numerous analytical applications and assumptions. Historically, the cities of Sparks and Reno and Washoe County projected revenues based on staff experience with historical revenues and their relationships to economic conditions (Larmore, 2016). Larmore found the process to be remarkable in its level of expertise and understanding of the economy and the tax systems involved with econometric models meant to augment and not eliminate them.

Quantitative and qualitative approaches are complementary and are used simultaneously to improve estimates and forecasts. Quantitative approaches include: time series forecasting, regression modeling, ARIMA and complex techniques such as neural networks. Expert judgment forecasting or Delphi and Consensus forecasting are two methods of qualitative forecasting (Sun and Lynch, 2008). In the Delphi method, a panel of experts is chosen. Information relevant to the forecast is sent the panelists and they asked to respond with a forecast. The experts never meet. In the consensus method, the panel of experts is brought together with the purpose of generating the requested forecasts.

Sun and Lynch (2008) concluded that highly complex analytical methods of forecasting do not necessarily results in greater accuracy rather a mix of methods both quantitative and qualitative support improved accuracy. Botric and Vizek (2012) also expressed similar sentiments. They noted that the cost-benefit analysis of applying advanced versus simple econometric methods for forecasting does not always rule in favor of advanced models. Quoting the works of Beckett-Camarata (2006), Bretschneider et al (1989) and Grizzle and Klay (1994), they suggested that a combination of expert judgment and simple time series models yields better results in term of forecasting accuracy when compared to more complex time series models.

Bretschneider et al (1989) compared the accuracy and precision of the expert judgment approach to with both simple and complex time series methods. They also looked at looked at institutional, political and methodological factors that affect forecasting. Expert judgment and simple regression models were found to give more accurate results. Grizzle and Klay (1994) compered the official forecasts of sales tax in twenty eight states in the United States with forecasts obtained from seven time series methods. The study concluded that the official forecasters provided more accurate predictions than time series models. The cost of econometric methods outweighed the benefits. Nonetheless, a combination of the two approaches produced a forecasts with higher accuracy measures. In analyzing the accuracy of forecast between expert judgment and econometric forecasting techniques, Beckett-Camarata (2006) used data from a county in Ohio that used the expert judgment approach and a city in Ohio that used the econometric forecasting models. The study though finds forecasting models more accurate as compared to judgment approach; it concludes that in general the accuracy of forecasting procedures varies by the level of revenue aggregation and the source of revenue.

Bolger & Wright (1994) in the works 'Assessing the quality of expert judgment' found it difficult to assess expert judgment as the definition of expertise was lacking in literature. They concluded that judgmental performance is largely a function of the interactions between the two dimensions of ecological validity and learnability. If both were high then good performance will be manifest but if one or both are low than performance will be poor. Although computer models and algorithms help aggregate weather information for operational forecasters, the human forecaster remains the primary information processor (Stewart et al 1989). The study acknowledged the importance of the supremacy of expert judgment despite the availability of conventional forecasting techniques.

Golosov and Kind (2002) and Kyobe and Danninger (2005), found one of the major sources of error or bias in revenue forecasting was the methods or models adopted in the forecasting and in addition a variety of political and institutional factors.

Counties and in particular the county of Machakos does not collect county specific data or statistics such as the County Gross National product, county population, unemployment rates etc. This can be attributed to lack of capacity and resources at the county level. For county specific data, one has to rely on national bodies such as the Kenya National Bureau of Statistics and others. In majority of the cases, data is presented as national aggregates. This limits its use in the county context. A resultant effect of this is to limit the use of econometric models or regression models that require the use of other independent variables to forecast a dependent variable.

2.2.1. Overview of Own Source Revenue

Constitutionally, counties can levy property rates, entertainment taxes, and service charge and fees among others. This pool of revenue sources form what is referred to OSR streams. OSR contributes less than 15% of total revenues. CRA sets 15% as the threshold for financing own budgets implying anything below 15% signifies an over dependence on transfers from the National Government. The focus on the ability of counties to raise and accurately forecast OSR has been necessitated by this overreliance and overdependence on national transfers. It is on this backdrop that the Kenya Government through the National Treasury and other stakeholders formulated the National Policy to Support Enhancement of County Governments' Own-Source Revenue. The policy which was first muted in 2015 by the Intergovernmental Budget and Economic Council in early 2015 was considered and approved by Cabinet on 14th of August 2018. The policy proposes a standardized framework for County Governments' own source revenue measures as well as compliance enforcement.

The significance of particular own source revenue stream also differs across the 47 counties. A particular OSR stream may be a significant contributor to overall OSR of a county while in another county the stream does not meet that criterion. In the counties of Mombasa and Nairobi, the land rates revenue stream contributed the largest percentage of the overall OSR for the financial year 2015/2016 followed by business permits while in the counties of Machakos and Kitui, business permits own source revenue stream was the largest contributor to total OSR of the two counties for financial year 2015/2016 followed by land rates.

With an exception of local revenues, all other sources of funding are external to the county meaning that the county does not in any way participate in raising them directly but receives them from the National Treasury based on some agreed formula or conditions as in the case of conditional grants. While the disbursements from the equitable share of nationally raised revenues should offer counties with some semblance of stability and guarantee of revenue, competing priorities, debt, declining revenue and other commitments by the National government have seen inordinate delays in disbursement of funds. These delays have negatively affected operations of the County Governments and service delivery. Focus has therefore shifted to own source revenue with the intent to strengthen its collection mechanism and increase it to wean counties off their dependence on equitable share transfers.

Across the 47 County Governments own source revenue collection systems differ. Only four structures of revenue collection and management at the county level are legally recognized and permitted. Counties can: establish internal revenue administration departments, establish autonomous county revenue authorities, contract Kenya Revenue Authority or contract private firms and other agents. With an exception of Nairobi, Nakuru and Laikipia counties, all other counties use internal revenue departments or a hybrid system incorporating two or more structures. The County of Laikipia has an autonomous revenue board tasked with collecting and receiving all revenue, administration and enforcement, assessment and accounting, provision of advice to the finance County Executive Committee Member on revenue matters, preparation of annual reports and payment of all revenue into the CRF. Nairobi County used a hybrid structure where its internal revenue department was tasked with OSR collection with an exception of some OSR streams including parking which had been contracted to an agent. On 25th February 2020, however, KRA took over the revenue collection function of the Nairobi County after the county transferred some of its functions to the National Government. Nakuru County through the Nakuru County Revenue Act of 2020 set up an autonomous revenue agency to collect revenue in the county. The Act provides a legal framework for establishing the Nakuru County Revenue Authority. Due to the duplication of efforts and wastage of resources, the President of the Republic of Kenya in 2019, directed a multiagency team to setup an integrated revenue collection systems for both national and county governments. A draft report with various possible systems has been presented for consideration (National Treasury, 2020)

The law requires that the citizen be involved throughout the budget process. To facilitate this involvement key budget documents should be easily accessible to the public. A review on budget documents available on county websites (Biegon and Wambui, 2019) found that while there was improvement in the availability of county budget information documents online, only 32% of the total expected documents were published online. The review was based on seven budget documents that should be published and publicized on the county websites between July 1st, 2018 and January 31st, 2019. The documents are: Approved Program Based Budgets 2018/2019, Citizen Budget (Enacted Budget) 2018/2019, Annual Development Plan 2019/2020, County Budget Review and Outlook Paper (CBROP) 2018, Quarterly Budget Implementation Report for the First Quarter of 2018/2019, Quarterly Budget Implementation Report for the Second Quarter of 2018/2019 and Finance Act 2018. 28 of the 47 County Governments published fewer than three of the seven total documents. Only the Elgeyo Marakwet County published all the 7 documents online. Machakos County only published one document online

Counties have maintained an upward trajectory in OSR growth achieved by local authorities albeit at a slower pace. However whether the growth is as a result of expansion of the tax base and improvement of operational efficiency or the growth is caused by an increase in rates and levies and introduction of new others is out of the scope of this study.

The composition of OSR for each county is unique. The national treasury (2019) identifies 59 different OSR streams reported by the county governments. Machakos County for example in the financial year 2015/2016 reported 17 OSR streams while neighboring counties such as Makueni and Kitui reported 11 each. The Commission of Revenue Allocation through financial year 2017/2018 and 2018/2019 condenses the OSR streams to 18 OSR streams. They are: hospital fees and public health services, single business permit, property rates, receipts from administrative fees, parking, game parks, others, cess, technical services, market trade Centre fees, environment and conservancy administration, housing rent, externals services, advertising and sign board, natural resources exploitation, liquor licenses, building plan approval and fines, penalties and forfeitures.

Even with such a huge tax base, 70% of total OSR collections were attributed to only 10 OSR streams (National Treasury, 2019). In FY 2016/2017, 40 percent of the aggregated OSR for the 47 counties was generated from three streams: business licenses, property related income (poll rates and plot rent) and vehicle parking. In the financial years 2017/2018, the three streams contributed to 43.2% and 40.1% respectively.

2.2.2. Factor that influence OSR

The National Treasury (2019) identified and estimated the potential of only 6 OSR streams that met the following four criteria: had adequate policy rationale, clear legal basis, high contribution to overall OSR and a revenue-raising potential. The six OSR stream are: property rates, building permits, business licenses, liquor licenses, vehicle parking fees and outdoor advertising. Of the six identified OSR streams, only three have consistently contributed over 10% to total OSR for Machakos County. They are: property rates, building plan approval and business licenses or single business permits. In FY 2015/2016 for example, they contributed

17.32%, 14.66% and 17.65% respectively to total OSR. Quarry OSR stream has also consistently contributed to over 10% of total OSR for Machakos County. It contributed 14.62 per cent of total OSR revenue for the FY 2015/2016. In designing a forecasting model for the Local Governments revenues in Nevada- City of Reno, City of Sparks and Washoe County, Larmore (2016) noted that it was difficult to include all revenue streams in the model. As a result, he only focused on two streams that contributed over 50% of the revenue for the local governments. Due to the many OSR streams the County of Machakos has, it is difficult develop a single model that includes them all. As a result this study will focus on the four revenue streams: property rates, building plan approval, business permits and quarry.

a) Property rates

Property rates OSR stream has consistently ranked among the top three revenue streams in Machakos County. In the year 2015, property rates accounted for 18.25% of total OSR. Despite its huge potential, revenues realized from property taxes are still low (National Treasury, 2019). The Draft policy on enhancement of county OSR attributes this to: outdated valuation rolls, computation of rates based on unimproved site value, operation of multiple rolls with different tax rates (one for each former local authority) and most land is communally owned or unregistered. Machakos County valuation roll was last updated in 1983. National level legislative legal reforms have been initiated to improve the administration of land rates in Kenya. The National Treasury in collaboration with the Ministry of Lands and Physical Planning is in the process of developing a National Rating Legislation to replace the outdated Valuation for Rating Act (Cap. 266) and Rating Act (Cap 267).

b) Single Business licenses

Single business licenses were introduced in 1998. The objective of its introduction was to: simplify the local regulatory environment, reduce compliance and administration cost, enhance local government revenue, generate a consistent business related data and establish a relation between the business community and local government. The regulatory framework for SBP is contained in the Local Government (Single Business Permit) Rules, 2008. A number of challenges bedevil the administration of single business permits. A significant number of counties have failed to enact trade licensing laws that should strengthen the single business permit. In addition, to boost their revenue, most counties have resorted to raising the SBP fee. These costs are then passed on directly to the consumer impacting negatively on the country’s ease of doing business. A lack of general information and complexities within the fee structure negatively impact the SBP stream. Professional bodies such as the Institute of Certified Public Accountants of Kenya, Law Society of Kenya and

Engineers Board of Kenya, have filed cases in court and have been granted restraining SBP on their members. These litigations have negatively affected revenue raising potential for the county governments.

c) Building plan approval

The power to control development within area of their jurisdiction is derived from The Physical and Land Use Planning Act, of 2019. The objectives of development control are: protect and conserve the environment, ensure orderly and planed building development, ensure optimal land use, safeguarding national security among others. Machakos County being one of the counties in the Nairobi metropolitan area has witnessed an upsurge in development and consequently development approval requests. Most of the development in concentrated in Mavoko, Machakos, and Tala sub counties as they border Nairobi. Availability of relatively cheap land, good infrastructure and proximity to the City has made Machakos prime for development. Though there is an increase, in compliance, enforcement challenges still continue to persist. Previously to get approval, an investor had to manually carry their plan to every concerned office for relevant approvals. This took time and most times all necessary approvals were never acquired. To ease the approval process the County of Machakos brought together all concerned departments in the approval process under one roof. This ensured that development plans are approved in the quickest time possible.

d) Quarry

This OSR stream is mainly concerned with all type of extraction of mineral resources. It includes sand, blast and building stones. Significant mining of blast activity takes place in the sub county of Mavoko and in the biggest source of revenue for the sub county. The Machakos County Management of Quarrying Activities Act of 2016 sets out the application process, granting of permit, cancellation and prohibited activities. It also sets the time that quarrying can be done and which roads can be used. Enforcement challenges still continue to persist. This is occasioned by use of undesignated roads, deliberate failure to pay required fees and licenses and lack luster enforcement.

The National Treasury (2019) identified the following factors has having a major influence on OSR collection:

- Illegal granting of waivers – to encourage compliance, several county governments have issued waivers without legal basis. From 2009 to 2021 the County of Machakos has issued several waivers ranging from rates to single business permits. It is a requirement of the law that when waiver are issued a report be made to the auditor general and there be a public record of each waiver together with the reasons. Most of the waivers issued do not meet this legal threshold. Furthermore it is not known if the waivers do encourage compliance and to what extent.

- Poor Human capacity and non-enforcement – policy on enhancement of own sour revenue identifies human capacity and enforcement weaknesses as factors influencing own source revenue. Most revenue administrators across the counties lack basic skills for the function. Efforts to collect revenue and enforce provisions of Finance Acts are more often than not combative involving physical force, arrests, street chases and riots (National Treasury, 2019). Though effort has been made to mitigate this weakness, the results indicate nothing has changed substantially.
- Low automation of revenue collection systems- slow rate of automation of the revenue collection process has seen counties continue using manual revenue collection systems. According to a 2015 county revenue baseline study, approximately 80% of revenue collectors are paid in hard cash on a daily basis. This model of revenue collection presents an inherent risk. Current automated revenue collection and management systems in use can be clustered into four categories. First, some counties use the Integrated Financial Management Information Systems (IFMIS) system. This is mainly used for reporting and is yet to be customized to incorporate own source revenue collection and management needs. The second system is the Local Authorities Integrated Financial and Operation Management System which was inherited from the local authorities. The third type of systems in use is customized revenue management systems developed by third parties. They are customized to particular revenues need of each county. These systems are not based on the Standard Chart of Accounts (SCoA) and are incompatible with IFMIS (National Treasury, 2019). The final system is a standalone receipting systems. For the benefit of automation the adoption of ICT systems to be felt there is need to standardize the ICT systems across all counties.
- Weak capacity for revenue forecasting and analysis - This affects own source revenue collection in part because of unrealistic revenue targets. Failure to meet revenue projections lead to budget deficits. Weak forecasting and revenue analysis is the focus of this study. This study aims to improve capacity to prepare realistic revenue projections.

2.2.3. ARIMA Model

This study uses an ARIMA model using the Box Jenkins methodology to forecast OSR in terms of current and past values. In instances where data is small or insufficient, ARIMA models are the most appropriate (Larmore, 2016). ARIMA models use information obtained from the variable itself to forecast it. It uses the philosophy "let the variable speak for itself". Botric and Vizek (2012) noted that comparison of different forecasting methods frequently found that ARIMA outperforms other models. Saayman and Botha (2015) acknowledged that univariate time series models like the autoregressive integrated moving average and seasonal ARIMA produce accurate forecasts but have the limitations in explaining influential external forces. Box and Tiao (1975) looked in to the question of how many observation were need to generate a forecast with the autoregressive integrated moving average models and recommended at least 50 observations.

ARIMA model generally have a good fit and are easier to interpret than some of the other forecasting models (Du Preez and Witt, 2003). There underlying assumption of ARIMA modeling. The first assumption is stationarity for AR model. A series must be stationary. A series is deemed to be stationary if: it exhibits mean reversions, has a finite and time-invariant variance and has a theoretical correlogram that diminished as the lag length increases. The second assumption is invertibility for MA models. This implies that: the series can be represented by a finite order MA or convergent autoregressive process uses the autocorrelation function (ACF) and particle autocorrelation function PACF for identification and implicitly assumes that the series can be approximated by an autoregressive model.

The autocorrelation functions (ACF) and partial auto-correlation functions (PACF) are used in the identification of the appropriate ARIMA/SARIMA models (Cuhadar 2014). Adhikari and Agrawal (2013) stressed the importance of analyzing the partial auto-correlation and auto-correlation functions inoder to come up with an appropriate time series model. The ACF determines the order of the MA term, while the PACF determines the order of the AR term (Box and Jenkins, 1970). An indication of an MA(q) is seen if the ACF cuts off after lag q and the AR(p) is identified if the PACF cuts off after lag p (Box and Jenkins, 1970). The best SARIMA or ARIMA model is one with statistically significant coefficients, uncorrelated residuals and smallest Akaike Information Criterion (AIC) and Schwarz Bayesian information Criterion (BIC) (Makoni & Chikobvu, 2018). In forecasting electricity prices for the following day, Contreras et al (2003) found that an ARIMA model had smaller error compared to artificial neural networks models.

An ARIMA model is classified as "ARIMA (p,d,q)" model where: p is the number of autoregressive terms, d is the number of differences needed for stationary, q is the number of lagged forecast errors in the prediction equation (Gujarati, 1995). In creation of the ARIMA model, four general steps will be considered: identification, estimation, diagnostics and forecasting (Larmore, 2016).

a) Identification – this involves finding the appropriate vales of p, d and q. This is normally done through trial and error. Three models with the lowest Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are selected. AIC and BIC measure the relative quality of statistical model for a given set of p, d, and q.
b) Estimation – here the parameters of the ARIMA models selected in the identification stage are estimated. The parameters include significance statistics, errors and model coefficients.

c) Diagnostics – the forecasting ability of the selected models is tested here using in and out of sample predictions. The Mean Absolute Percentage Error (MAPE) for the three models is then estimated and the model with the least MAPE is chosen.
d) Forecasting – the final model is then used in forecasting.

ARIMA models have proven to be reliable and accurate in time series prediction, forecasting and analysis (Farhath et al, 2016).

2.3. Empirical review

For smooth operation of government, accurate revenue forecast is essential. During times of unforeseen economic shocks such as those occasioned by the Covid_19 pandemic the ability to accurately predict revenue is vital. According to the Kenya's Auditor General half year report of 2018/2019 FY, counties lacked capacity to set, review and report realistic revenue targets with most basing their revenue projections on the deficit gaps based on what they receive from the exchequer. Wong (1995) made three general conclusion in his study on local government revenue forecasting. One, the systematic revenue forecasting and long-range planning are necessities and not luxuries, two, risk aversion to technical revenue forecasting can be overcome and three, the implementation of a systematic revenue forecasting systems does not require a battery of "rocket scientists."

A survey of the local governments of Ohio found 75% of the 290 finance officers did not use any formal revenue forecasting methods (Beckett-Camarata, 2006). Larmore (2016) noted that for smaller jurisdictions formal or quantitative forecasting model are much more difficult to come up with. This, he attributed to two reasons. First, he acknowledges the challenge of getting accurate and verifiable data for the small jurisdictions. Past data for many variables is often unavailable at the local government level (Weller and Kurre 1999). The second reason was the markedly different economic and business cycles for the economies at the micro-economic and macro-economic level. Weller noted that at the micro level, economies tend to be more specialized than at the macro level leading to different business cycles.

There are several time series models used in forecasting in different fields such as finance, revenue and budgeting and tourism. By far the most common are autoregressive integrated moving average models, vector autoregressive models, simple exponential smoothing and Holt Winters models. These models have become useful tools in modeling univariate data due to their accuracy in predicting in-sample and out of sample values. Some researchers believe that econometric regression models are more precise than univariate time series models (Makananisa, 2015). Unlike time series models which attempt to forecast variables based on the past values of the variable, econometric forecasts are based on regression models that relates one or more dependent variables to a number of independent variables. Corvalao et al (2010) compared the ability of a regression model and ARIMA model to predict VAT collection in Santa Catharina, Brazil. The mean absolute percentage error was used to compare the two models. The MAPE were 2.51% and 4.63% for the regression and ARIMA model respectively. However, the study did not convincingly show the leading econometric regression model against the ARIMA as the MAPE was calculated for a one year period and not the entire sample. (Makananisa, 2015). Ramanathan (2002), found time series approaches to be generally more superior to econometric approaches when making short-term predictions.

The ARIMA and Holt Winters Additive and Multiplicative models were used in separate studies in Romania to predict future total revenue and own revenue of local authorities. Pelinescu et al (2010) used the Holt Winter additive and multiplicative approaches. They used data from the first quarter of 2000 to the last quarter of 2010. The study used the E-views software to build and run the Holt Winter equations. The model with the least Root Mean Squared Error was the most preferred. The study recommended use of Holt Winters model because they were user-friendly and provided stable forecasts. Brojba et al (2010) used the data from 2007 to 2008 to develop an ARIMA model to forecast total budget revenue. The study found that the fitted values were close to the actuals and concluded that ARIMA models could be used to predict future revenues. However the study noted the parameters of the model were sensitive to sample selection and therefore the model was best suited for short-term forecasts (Brojba et al, 2010). Jayesekara and Passty (2009) used ARIMA models to predict the net income tax revenue for Cincinnati Ohio. Data used was from 1970 to April 2009 but was readjusted to start from 1989 due to tax rates changes. The developed ARIMA model fitted the data well, capturing seasonality in the data throughout the sample period. The in-sample forecasts were close to the actual values for the period 2006 to 2009. The study recommended the use of ARIMA model for short term forecasts of net income tax. Slobodnitsky and Drucker (2008) evaluated the ability of two ARIMA models: monthly and quarterly to forecast value added tax revenue for the state of Israel. The quarterly ARIMA model included explanatory variables such as tax rate, sector GDP and consumption. Using the quadratic loss function as a measure of accuracy, the quarterly net VAT ARIMA model was found to be the most suitable forecast model. Urrutia et al (2015) developed an ARIMA model to forecast income tax for the Philippines for 7 years. They found ARIMA (0,1,0) or random walk model as the best suited model. The paired t-test was used to test the forecasting accuracy of the model and showed no significant difference between actual and predicted values.

To forecast tourism growth trends in Indonesia, Malaysia and Thai, Suwanvijit (2014) used monthly data from January 2002 to December 2011. The Mean Absolute Percentage Error was used as the measure for accuracy. The study concluded that average annual increase of the overall arrivals would be approximately seventeen per cent on the out-of-sample forecasts.

In their study, Silvestrini et al (2008) aimed to build a statistical univariate model to detected the narrowing and widening of the government deficit of the government of France in advance which would we used to advise the government on policy and decision making and control of the deficit. They used data dating from January 1996 to December 2004. In the study the actual sample used data up to December 2001. It was used to meld all revenue and expenditure variables using seasonal auto regressive integrated moving average models. Data from 2004 to 2005 was reserved for model validation. The generated monthly forecast updated and summed to form the monthly cumulative forecasts for the tow validation years and form the temporary aggregated annual ARIMA models constructed for one step ahead yearly forecasts. The monthly cumulative forecasts, aggregated annually forecast and the French official forecasts were compared with the 400 and 2005 yearly actual figures. The study observed that the temporary aggregated annual ARIMA forecast were closed to the actual when compared to the traditional forecasts.

Fuel tax, License and registration fees were a major source of revenue for North Dakota's transport department (Berwick & Malchose, 2012). The auto regressive integrated moving average model was used to forecast fuel tax revenue, license and registration fees. Data dating from 1951 to 2010 was used to predict the revenue for the period 2008 to 2013, with forecast from 2008 to 2010 used to test the efficacy of the model. ARIMA(1,1,1) and ARIMA(1,2,1) were fitted for fuel revenue and license and registration fees respectively. The fitted auto regressive integrated moving average models captured the movement of the two revenue sources with minimal errors against the actual and the forecast generated from the models.

Jain and Kumar (2007) used a hybrid time series neural network model to forecast the monthly stream flow of the Colorado River at lees Feery, United States. The hybrid model is a combination of traditional time series approaches and artificial neural networks. It capitalizes on the unique strengths of both forecasting approaches to build a more superior forecasting model. Trend and seasonality is first removed from the data before presenting it the artificial neural network model. The study however raises a pertinent question: whether artificial neural networks need the times series to be stationary like in conventional or traditional time series modeling. The study answers this question to the affirmative but notes however that more studies need to be done to give strength to the findings.

Chimmula and Lei (2020) used the Long short-term memory (LSTM) networks to forecast the future of COVID-19 cases. Chimmula and lei noted that most approaches sued in previous studies were linear and often neglected the temporal components in data. Statistical models such as ARIMA overwhelmingly depend on assumptions and such are difficult for forecasting real-time transmission rates. The root mean squared error of the network was 34.83 while the accuracy was 93.4% for the short term prediction in Canada. Based on the validation dataset, the RMSE error was about 45.70% with an accuracy of 92.67%. Loannidis et al (2020) noted that epidemic forecasting has a dubious track record and its failures are manifested in the covid-19 pandemic. They attributed this to poor data, wrong modeling assumptions, high sensitivity of estimates, lack of incorporations of epidemiological features, poor past evidence on effects of available interventions, lack of transparency, errors, lack of determinacy, consideration of only one or a few dimensions of the problem at hand, lack of expertise in crucial disciplines group thinking, bandwagon effects and selective reporting.

Etuk and Igbuda (2013) used the seasonal autoregressive integrated moving average (SARIMA) model to explain the Nigerian Naira – British Pound exchange rate. Monthly exchange rates were used from 2004 to 2011. The final model derived was SARIMA (0,1,0)(2,1,1)(12). The model explained 61% of the exchange rate variations in the data and was thus found to be adequate.

The seasonal autoregressive integrated moving average model was used to forecast inflation rates in Nigeria (Otu et al. 2014). A total of 120 data point were used from November 2003 to October 2013. SARIMA (1,1,1)(0,0,1)(12) was found to be the best fitting model. The study recommended the forecast results be used to assist policy makers in decision making.

In Koirala (2011), five forecasting models were used to forecast Nepal's total revenue for the financial year 2012/2013 to 2013/2014. 192 monthly observations from 1997 to 2012 were used. The five methods were: Holt method, the Decomposition methods, Seasonal Autoregressive Integrated Moving Average method (SARIMA), Winters method and Growth Rate method. An evaluation of the results from the five methods found that the Winters method with a seasonal component and Seasonal autoregressive integrated moving average to best fit Nepal's total revenue. However, when the least mean percentage error (MPE) and mean absolute percentage error (MAPE) were considered, the Seasonal Autoregressive Integrated Moving Average model was found to be albeit superior then the Winter method as it had the least errors.

Ahmed et al (2020) used an ARIMA model for the short-term forecasting of tourism revenues in Bangladesh for the period for 45 years from 1973 to 2017. They found out that an ARIMA (0,1,1,) with drift model and a confidence interval for 80% and 95% is the most fitted method to describe the trend of tourism receipts over time. The general form of the model was:

$(1-\phi_1 B-\cdots-\phi_p B^p)(y'_t-\mu)=c + (1+\theta_1 B+\cdots+\theta_q B^q)\epsilon_t$

Where:

$\phi_1 \ldots\ldots, \phi_p$ are the regular autoregressive parameters

$\theta_{1,\ldots\ldots,}\ \theta_q$ are the regular moving average parameters,

$(1-B)^d y_t = y'_t$ is the first difference of y_t

ϵ_t is the white noise term

$c=\mu\ (1-\phi_1-\cdots-\phi_p)$ and

μ= mean of $(1-B)\ y_t$ Therefore μ can be interpreted as a drift coefficient.

In explaining their choice of ARIMA over SARIMA, they acknowledge that lack of monthly or quarterly datasets was the main reason. A major limitation of this study is it never gave the accuracy measures of the

ARIMA model. It only provided the forecasted values. This makes it hard to determine the suitability of the model and its performance in forecasting future revenues from tourism. In predicting the number of tourist arrivals in South Africa, Saayman & Saayman (2008) used ARIMA, Holt-Winters exponential smoothing and SARIMA models. They found out that the SARIMA model outperformed all the other models. While evaluating non-linear approaches in forecasting South Africa tourist arrivals, Saayman and Botha (2015) acknowledges that the univariate time series models like the autoregressive integrated moving average and seasonal ARIMA gave better forecasts when compared to seasonal naïve forecasts. In Tanzania Ndiege (2015) used quarterly data in modeling international tourism demand and fitted ARIMA and SARIMA model that estimated a low growth of the Country's tourist rate per year. His study strengthened the belief of SARIMA models superiority in terms of accuracy.

Rahmat et al (2020) used the autoregressive integrated moving average to predict regional revenue and expenditure budget for the North Sumatra regional government in Indonesia. The regional revenue and expenditure budget is viewed as the main determinant of regional development. The higher the revenue, the higher the amounts spent on infrastructure and other regional development activities. The study concluded that the auto regressive integrated moving average produced the best results with the least standard errors. In this study prediction results were based on the ARIMA model with the least AIC.

Liu & Wang (2015) forecasted the development of income and financial expenditure in China Using autoregressive integrated moving average method. The ARIMA model processed data from 1950 to 2013 (64 years). ARIMA model (2,1,6) was found to be the best with results indicating that there would be an increase in allocation from 9.93% to 9.97% on China's state finances the following year. Almasarweh & Alwadi (2018) used an ARIMA model to forecast banking stock data in Jordan with data from the Ammam stock exchange from 1993 to 2017. The study concluded that ARIMA (1,1,2) with a standard error of 1.4 was the best forecasting model.

Ekims et al (2016) used a hybrid model of the ARIMA and autoregressive feed forward artificial neural network to predict the revenue of 130 stores of a leading fashion retail chain. The models accuracy could reach 80% to 85% at the end of the day and forecasting accuracy for each hour could reach 90% t0 95%. The accuracy was higher than the accuracy generated by either model individually. The hybrid model was preferred because of the time series with different lengths (Ekms et al, 2016). The feed forward autoregressive artificial neural networks are models types obtained by adding sequential values of the series as independent variables to lagged structure (Yt-k).

$$\widehat{Y_t} = b_0 + w_j f_j(b_j + \sum_{i=1}^{p} w_{i,j}\, x_i),$$

In this;

p : Autoregressive lag constant (ro),

$b_j,\ w_{i,j},\ w_j$: Constant variables,

x_i : Independent variable,

$f(.)$: Sigmoid function.

Cross validation tests of the results from the ARIMA, and neural network showed that the neural network model forecasted with a higher accuracy by length of time series which are more than 2 years.

Makananisa (2015) used ARIMA/SARIMA models and Holt-Winters to forecast annual tax revenue of the South African taxes. The study focused on personal income tax, corporate income tax, values added tax and total tax revenue. The data set was from January 1995 to March 2010 for the in sample data. The forecast period was from 2010/11 to 2012/13. Due to the high volatility of the corporate income tax and negative values, it was converted to quarterly data from the first quarter of 1995 to the first quarter of 2010. The SARIMA and holt-winters model performed well in modeling and forecasting personal income tax and value added tax while the holt-winters model outperformed the SARIMA model in modeling and forecasting the more volatile corporate income tax and total tax revenue. In the study it was noted that in some years forecast error against the actual was higher than 5% dues to the economic recession. The study concluded that selected models are expected to perform better when forecasting future values assuming that there will be no shocks such as economic recession. Both models were to be sued for short term forecasts and it was expected that if the tax recovery approaches did not change, then the models will be precise with limited bias in forecasting tax revenues with minimal errors and fewer model revisions being necessary.

Wang et al (2015) used the autoregressive integrated moving average model to forecast revenue for telecommunication companies. They considered ARIMA model to be advantageous as economic rules or social rules of income indicators needed not be factored or considered. In comparison to the ARIMA models, regression models require the variance of the sample data are equal and independent while the neural networks though acknowledge to have a higher level of accuracy was complex. In the study they used data from China Telecom business analysis systems from year 2011 to 2012 as the in sample data while data from the third month of 2013 to the five month of 2013 was used as the out of sample data. ARIMA(1,2,1) was found to have the least forecast error. In their conclusions, they noted that, the study used only 24 months and hence seasonal analysis was not considered. They advocated for further research using a more comprehensive ARIMA model which included trend, seasonal and cyclical factors.

In forecasting fiscal revenues in Croatia, Botric and Vizek (2012) used several models: trend model, random walk, ARIMA, regression and error correction models. The forecasts of the various models were compared with official forecasts which were obtained mostly using expert judgment with a view of determining whether policy makers would profit from adopting a formal forecasting approach. Botric and Vizek took a disaggregated approach to revenue forecasting. They disaggregated the fiscal revenue in to various revenues streams; revenue from income tax, corporation tax, value added tax, property tax, import duties, excises and social contributions and then used the forecasting models to predict the individual revenue streams. Thiels U and mean absolute percentage error (MAPE) were used to compare the accuracy of the models. The study concluded that time series methods demonstrated the ability to produce relatively accurate forecasts than the expert judgment approach in a challenging environment. The study also concluded that forecasting accuracy increased with the complexity of the method or model. In this sense, ARIMA, error correction and regression models were perceived to be more complex in the study and hence would probably give much better results as opposed to random walk or linear trend models which were perceived to be relatively simple.

In Ghana, Ofori et al (2021) forecast value added tax using the ARIMA with intervention model and Holt linear trend method. The ARIMA model with intervention was found to outperform the Holt linear model in both accuracy and precision. The ARIMA with intervention model had the least root mean squared error, mean absolute percentage error and mean absolute deviation for the predicted value of value added tax. The study also found the ARIMA with intervention model to be more accurate than the in-house model adopted by the Ghana Revenue Authority in forecasting monthly value added tax revenues. The model was thus recommended for adoption by the Ghana Revenue Authority on grounds of precision and effective fiscal planning.

Nikolov (2002) used monthly tax revenue data from January 1998 to July 2002 to forecast tax revenues for the Republic of Macedonia using the Box-Jenkis model with intervention. The study concluded that the model was good for predicting the tax revenues but the variance of the forecasts in time series models became larger over time. The model was therefore effective in forecasting only a few time units ahead as opposed to longer time periods.

Regression and econometric revenue forecasting approaches were found to be effective in predicting revenue for a medium-sized city (Wong, 1995). The models were particularly useful in obtaining multiyear forecasts for both stable and erratic revenue sources.

Hill et al (1996), suggested that traditional statistical time series methods suffered from four limitations, the first one, was without expertise it was possible to incorrectly specify the functional form relating the independent and dependent variables and fail to make the necessary data transformations. Secondly, traditional time series methods do not respond well to outliers and can lead to biased estimates of model parameters (Iman and Conover, 1988). The third limitation was human interaction and evaluation was required when many kinds of traditional statistical time series models were estimated. The forth limitation was that traditional statistical methods do not learn incrementally as new data is added and instead they must be re-estimated periodically. To overcome these limitations, Hill et al(1996) suggested the use of neural networks as an alternative to traditional statistical forecasting methods.

In various studies neural networks and traditional time series techniques have been compared using data from the 'M-competition' (Makridakis et al. 1982) which used 1001 real time series. In the competition, various groups of forecasters considered to be experts in a particular technique were given all but the most recent data points in each series. Each group was free to use any technique in their domain of expertise to forecast the many time series. After all the forecasts were made, Makridakis compared the forecasts to the actual values in the hold put data set and reported in Makridakis et al (1982). Using data from the M-competition, neural networks were found to be inferior to Holt's, Brown's and the least square statistical models for time series of yearly data but comparable with quarterly data (Foster et al, 1992). For time series data with lengthy histories, the Box-Jenkis models and neural networks were found to produce similar results (Sharda & Pati, 1992). Hills et al (1996) using the M-competition data concluded that the neural network model was not significantly better than traditional statistical and human judgment methods in forecasting monthly and quarterly data. However the traditional models and neural network model were comparable on the annual data.

Weatherford and Kimes (2003) compared different forecasting methods for predicting hotel revenue. Data from Choice Hotels and Marriott Hotels was used. For the Choice Hotels data, pickup methods and regression produced the lowest error while for the Marriott's Hotel data set, exponential smoothing, pickup and moving average produced the best results.

Williams and Kavanagh (2016) compared the forecast accuracy of fifty five revenue data series (thirteen property tax, thirteen sales tax, ten total general fund and nineteen other revenue types) from eighteen local governments with the last 18 months evaluated for accuracy. The resuts showed that simple exponential smoothing methods performed best with monthly and quarterly data and use of quarterly or monthly data was marginally better than annual data.

Ayakeme et al (2021) compared the ARIMA and Winter's additive and multiplicative model to predict internally generated revenue for the State of Bayesla in Nigeria. With a confidence interval of 95%, they found that the ARIMA model had the least Mean Absolute Error and Mean Square Error compared to the two Winter model. They used the general model for ARIMA (p, d, q) given by;

$$\Phi(B)(1 - B)^d X_t = \theta(B)e_t$$

The Winters multiplicative model was given by:

$$\hat{Y} = (L_{t-1} + T_{t-1}) S_{t-p}$$

Where;

$$Tt = y [L t - L t\text{-}1] + (1\text{-}y) T t\text{-}1$$

$$St = \alpha (Y t / Lt) + (1 - \alpha) St\text{-}p$$

$Lt = \beta (Yt / St\text{-}p) + (1 - \beta) [L\ t\text{-}1 + T\ t\text{-}1]$

Lt is the level at time t, α is the weight for the level.

Tt is the trend at time t, y is the weight for the trend.

St is the seasonal component at time t, β is the weight for the seasonal component.

P the seasonal period.

Yt is the data value at time t.

Ŷt is the fitted value, or one-period ahead forecast at time t.

Model for the winters additive model was given by:

$$\hat{Y} = L\ t\text{-}1 + T\ t\text{-}1 + S\ t\text{-}p$$

Where;

$Tt = [L\ t - L\ t\text{-}1] + (1 - y)\ T\ t\text{-}1$

$St = \alpha (Yt - Lt) + (1 - \alpha)\ St\text{-}p$

$Lt = \beta (Yt - S\ t\text{-}p) + (1 - \beta) [L\ t\text{-}1 + T\ t\text{-}1]$

L, is the level at time t, α is the weight for the level.

Tt, is the trend at time t, y is the weight for the trend.

St is the seasonal component at time t, β is the weight for the seasonal component.

p is the seasonal period.

Yt is the data value at time t.

Yt is the fitted value, or one-period ahead forecast at time t

In concluding that the ARIMA model was the best suited model, the Mean absolute percentage error and mean squared error were considered. Table 2 shows that accuracy measure for the ARIMA model. And the Winters additive and multiplicative models.

Table 2: MAPE and MSE for ARIMA and Winters multiplicative and additive models

Model	**Mean Absolute Percentage Error**	**Means Squared Error**
ARIMA model	0.915	93.873
Winters Additive model	4.92707E + 01	9.83475E + 16
Winters Multiplicative model	4.68776E + 01	9.46733E + 16

From the above, the ARIMA model performed better in both measures than the Winters additive and multiplicative models.

Larmor(2016) in his quest to create a revenue prediction model for all Local governments revenues in Nevada- City of Reno, City of Sparks and Washoe County. He concluded that it was difficult to create due to multiple revenue sources and resulting variables. As a result the paper focused on two streams: property tax and sales tax that made up over 50% of the revenues. To forecast the two streams, three models were used: single equation regression, ARIMA and vector auto-regression models.

Table 3: Comparison of taxable sales (Washoe County) in-sample and out-sample MAPE for ARIMA, single equation regressed model and vector auto-regression models

	In-sample	**Out-sample**
Single equation regression model	6.7013%	3.3093%
ARIMA model	2.86%	2.85%
Vector auto-regression model	162.95%	169.67%

From table 3, ARIMA model is the best suited model for forecasting taxable sales in Washoe County as it has the least MAPE for both in sample and out-of sample. However, it was noted that unless used by a person familiar with the ARIMA model it should not be used at all.

Table 4: Comparison of assessed value (property tax) in-sample and out-sample MAPE for ARIMA, single equation regressed model and vector auto-regression models

	City of Reno		**City of Sparks**		**City of Washoe**	
	In-sample	Out-sample	In-sample	Out-sample	In-sample	Out-sample
Single equation regression model	4.87%	4.58%	-	-	-	-
ARIMA model	3.82%	5.12%	4.87%	8.68%	3.62%	7.25%
Vector auto-regression model	138.9%	534.8%	137.7%	6000%	109%	463%

Table 4 shows the MAPE for forecasted assessed value. Larmore predicted assessed values as opposed to property tax to avoid the impact of changes in tax rates. The single equation model and ARIMA models resulted in the lowest MAPE. The vector auto-regression models as in forecasted taxable sales for Washoe County resulted in high MAPE errors and are therefore not recommended for use.

Temür et al (2019) compared the ARIMA, long short-term memory (LSTM) and a hybrid of the two models to predict house sales in Turkey. Table 5 shows the MAPE results of the three models. The ARIMA model has the lower MAPE compared to the LSTM model but an hybrid of both model yields the lowest MAPE. The hybrid model was found to have the least mean absolute percentage error and the least mean squared error.

Table 5: MAPE for ARIMA, LSTM and Hybrid models

Model	MAPE
ARIMA	0.121
LSTM	0.150
Hybrid (LSTM 1500 epoch-ARIMA (1,1,1))	0.072

Most ARIMA forecasting researchers use the size of the forecasting errors to examine the performance of the developed forecasting models majority have used the mean absolute percentage error (MAPE), root mean square error (RMSE), and root mean square percentage error (RMSPE) as a means of examining the forecasting performance of their predictive models.

Forecasting error values are often turned into absolute values in order to prevent negative and positive forecasting errors from cancelling out. The RMSE is a quadratic function whereas the MAPE represents a linear loss function. The MAPE measures the mean absolute percentage difference between the forecast revenues and actual revenue. The RMSE evaluates the extent to which forecast errors deviate from the mean actual revenues. The RMSPE measure the exact same deviation as the RMSE but s expressed as a percentage.

Hyndman and Koehler (2006) noted that the MAPE has the advantage that it is not dependent on scale and is therefore very suitable when forecasting results between different data series are compared. But it also has the disadvantage it is more sensitive to positive forecast errors than negative forecast errors (Hyndman and Koehler 2006)

To determine the best model to forecast income tax revenue, Jayesakara and Passty (2009) chose the model with the least Root Mean Squared Error, Akaikae Information Criterion and with the highest R-Squared. Makananisa (2015) employed several measures of accuracy such as Mean error, root mean squared error, mean absolute error, mean percentage error, mean absolute percentage error and mean absolute squared error. Other accuracy measures used included R2, Akaiake information criterion and Bayesian information criterion.

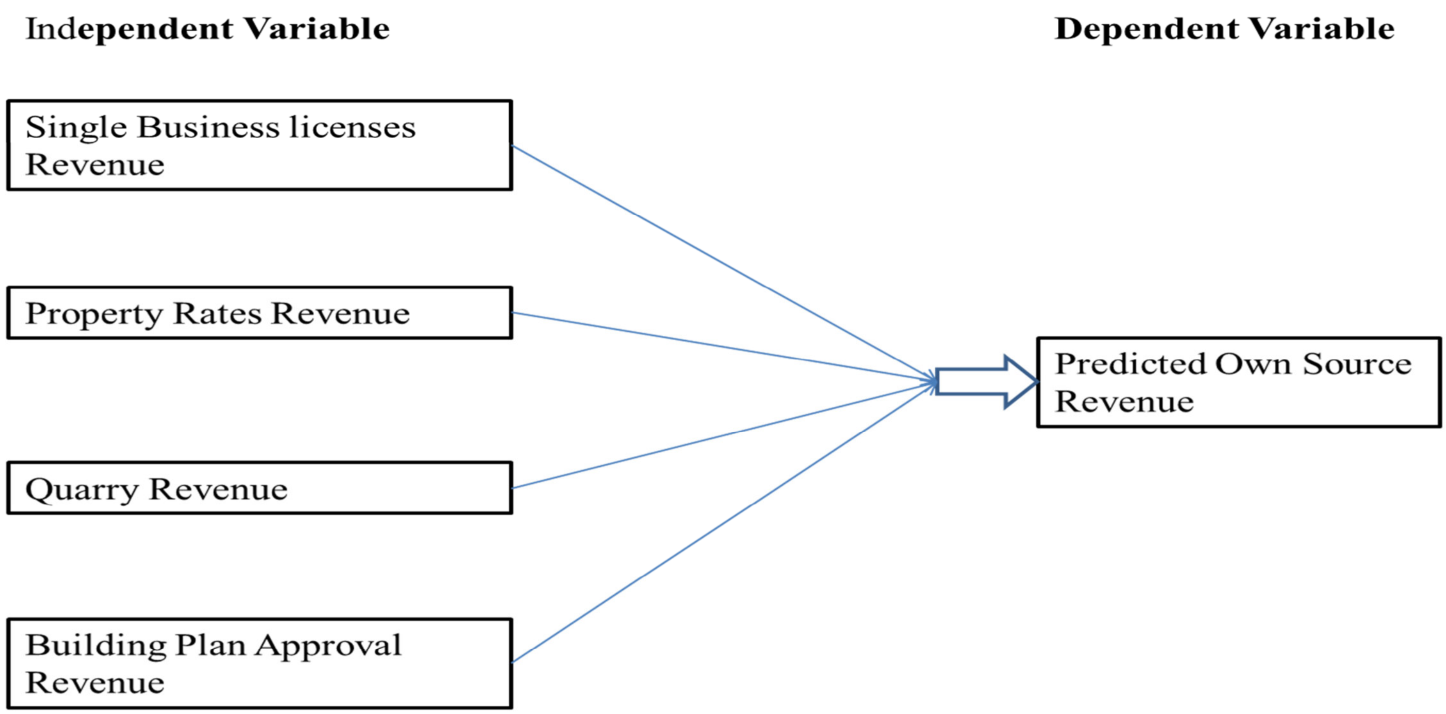
According Hsu (2010), a forecasting accuracy result with a forecasting error of 10 per cent and less than is an indication of highly accurate forecasting results, whereas a percentage number of between 10 and 20 indicates a good forecasting accuracy. A percentage number between 20 and 50 shows a reasonable forecasting accuracy. Researchers have also found forecasting accuracy to vary between forecasting horizons. A longer forecasting horizon often leads to less accurate forecasting results dues to the possibility of several unforeseen events taking place (Li et al, 2005). The type of data used may also have an influence on forecasting accuracy. Li et al (2005) noted that advanced causal methods often outperform time series techniques when annual data is used whereas time-series techniques using monthly or quarterly data often lead to better forecasting accuracy results than casual methods sing monthly or quarterly data.

In this research the MAPE will be the preferred measure of accuracy due to its advantage in that it is not dependent on scale.

2.4. Conceptual Framework

The conceptual framework of the study is presented in figure below:

Figure 5: Conceptual Framework



Forecasted OSR base is the response variable while the predicted OSR revenue streams are the explanatory variables.

2.5. Operationalization of Variables

Table 6: Operationalization of variables

Variables	Indicators	Data to be collected
Property rates	Annual Budgeted revenue estimate in Kshs Annual Actual revenue collected in Kshs	Annual budgeted property rates revenue estimate Annual actual property rates revenue
Single business permits	Annual Budgeted revenue estimate in Kshs Annual Actual revenue collected in Kshs	Annual budgeted single business permit revenue estimate Annual actual single business permit revenue
Building plan approval	Annual Budgeted revenue estimate in Kshs Annual Actual revenue collected in Kshs	Annual budgeted building plan approval revenue estimate Annual actual building plan approval revenue
Quarry	Annual Budgeted revenue estimate in Kshs Annual Actual revenue collected in Kshs	Annual budgeted quarry revenue estimates Annual actual quarry revenue

2.6. Summary

In this chapter, we have looked at previous studies touching on revenue forecasting, various techniques and there challenges. The chapter builds the foundation of the next chapter which is methodology. As explained the focus will be only two OSR streams due to constraints of time and data.

CHAPTER THREE
METHODOLOGY

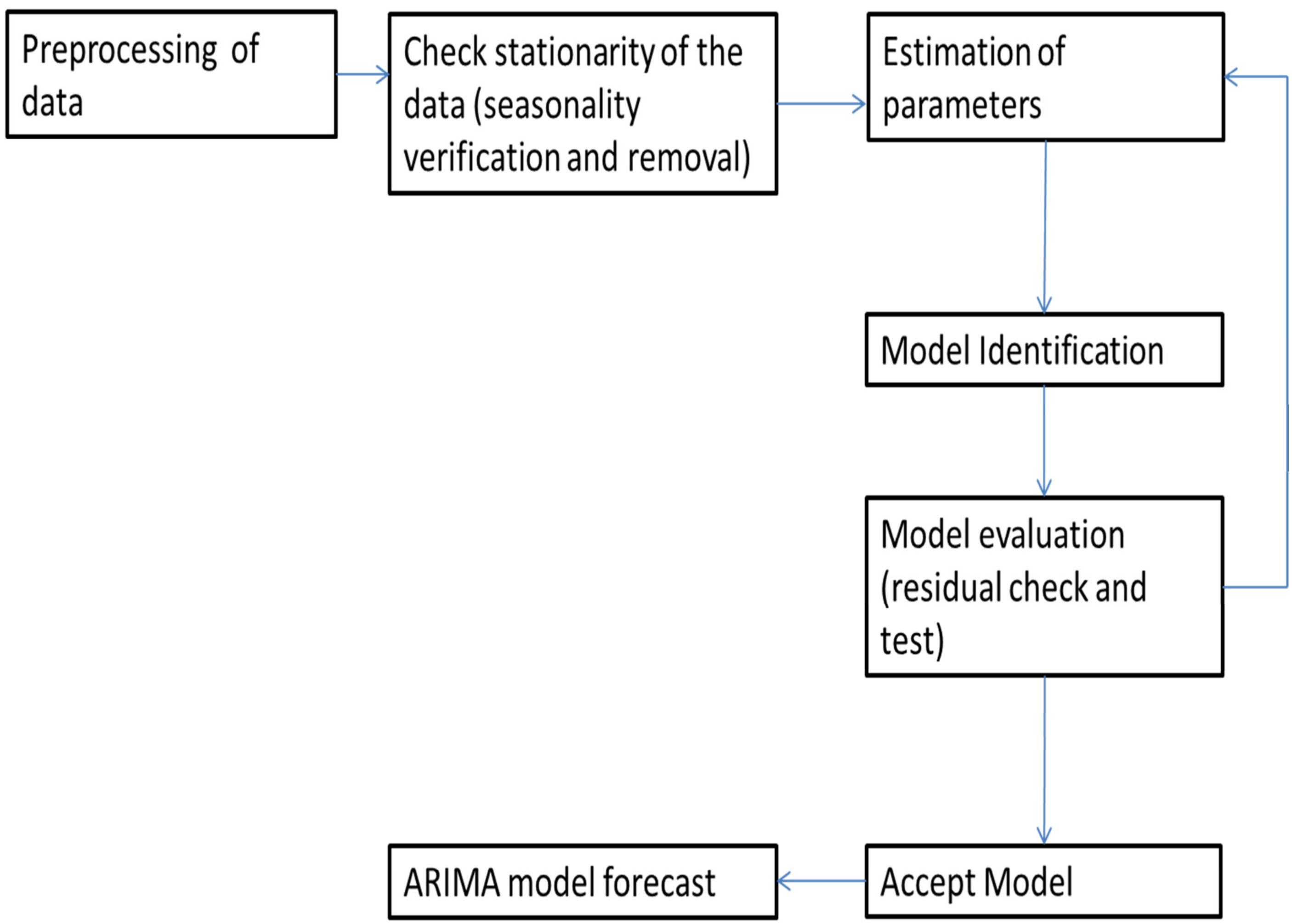
3.1 Introduction

In this chapter, the research design framework, data sources, data collection and analysis techniques are dealt with. It also presets the data analysis techniques used. The chapter justifies the choice of research methodology adopted in the quest to achieve the outlined objectives and answers the research questions posed.

3.2 Research Design

Figure 6 illustrate the steps to be followed to achieve the main objective of this study.

Figure 6: Research Design



Source: *International Journal of Intelligence Science*, *10*(03), 65.

This study adopted a mix of quantitative research design and simulation research design. The study used secondary data available from various government agencies such as: Commission of Revenue Allocation, Controller of Budget and records from the County. The Legal requirements that counties submit their detailed budget estimates make secondary data the most appropriate for this study. Data obtained was first preprocessed. This ensured that data obtained is uniform and there are no missing values. The first objective is to determine which OSR streams contribute the highest to overall OSR. To do this a correlation matrix is used to highlight the correlation of the first nine highest grossing and total. The second objective was to develop ARIMA forecasting models for the identified streams. The study confined itself to those OSR streams identified in the first objective as being high contributors to overall OSR. The next step was to get the times series trend of the identified OSR streams. If trend and seasonality was detected, it was removed through differencing to achieve stationarity. The Augmented Dickey-Fuller test was used to check for stationarity. After stationarity was achieved, the next step was to estimate the parameters of the model. Akaike's Information Criterion (AIC) will be used to determine which model will be selected. The chosen model was then be evaluated on its ability to generate accurate forecasts vis-à-vis the prediction made in the past by the county government. This was done through simulation. This was the third objective of this study.

3.3 Data Collection Techniques

Only secondary data was used. The data was collected for the period from financial year 2013/2014 to the financial year 2018/2019. Secondary sources mainly comprised of statutory filing by the county executive with the county assembly and records from the county treasury as well as from the office of the Controller of Budget and Commission on Revenue Allocation. The data was used in developing the forecasting models.

3.4 Data analysis techniques

The analysis of data was done using python. Predictive ARIMA models were developed for each of the identified OSR models. The data first underwent preprocessing to eliminate any redundant values and fill in missing variables. The data was then be tested for stationarity using the Augmented Dickey-Fuller test. The data was made stationary where it failed the stationary test. There after the four steps used in the development of ARIMA models: identification, estimation, diagnostics and forecasting were used. The mean absolute percentage error (MAPE) was the main diagnostic tool. It was used to evaluate the forecasting ability of each model. A lower MAPE means a better forecasting accuracy of the developed models.

CHAPTER FOUR
DATA ANALYSIS, FINDING AND DISCUSSIONS

4.1 Introduction

This chapter presents the findings of this study. It starts with descriptive statistics showing the relevance of the chosen OSR streams, then goes on to present and evaluates the findings of the three objectives of this research. The findings of the objectives are presented in this order: identification of the attributes or factors that influence own source revenue, development of an ARIMA model that can be used to forecast OSR and an evaluation of the developed model. A brief summary is given at the end of the chapter.

4.2 Descriptive statistics

This paper seeks to build an ARIMA model for forecasting own source revenue for the county of Machakos. To achieve this, monthly own source revenue data was collected for the county of Machakos from the financial year 2013/2014 to financial 2018/2019. The revenue data obtained is disaggregated into the various different revenue streams. The revenue data is the used to develop ARIMA forecasting models for the building approval, quarry, business permits and property rates. Figure 7 shows the descriptive statistics of the four revenue streams. A high standard deviation meaning that it's spread over a large range of values while a low standard deviation means that the values are generally clustered close to the mean. A high standard deviation may be an indication of seasonality. This is because the revenue varies with season creating widely dispersed revenue figures. A low standard deviation on the other hand may be an indication of a consistent revenue stream not subject to seasonal variations.

As discussed from the literature review only four revenue streams: building approval, quarry, business permits and property rates, are used to develop the forecasting model. The four revenue streams continually and consistently makeup more than 50% of total OSR collect since financial year 2013/2014. In addition, it would be impossible to include all the revenue streams in one forecasting model. Figure 7 shows the four revenue streams total collection from the FY 2013/2014 to FY 2018/2019 as a percentage of total OSR collected and the combined percentage of the four streams in that particular financial year.

Figure 7: Descriptive Statistics

	Land Rates	Quarry	Sand Gravel	Building Approvals
count	7.200000e+01	7.200000e+01	7.200000e+01	7.200000e+01
mean	1.619806e+07	2.032865e+07	5.463079e+06	1.274310e+07
std	1.562218e+07	2.370206e+07	6.938338e+06	9.902673e+06
min	9.682000e+05	1.198965e+06	1.013000e+05	2.959500e+05
20%	6.418568e+06	8.722544e+06	2.121944e+06	4.541949e+06
40%	9.302783e+06	1.466101e+07	3.213122e+06	1.024979e+07
50%	1.071798e+07	1.524056e+07	4.104797e+06	1.106143e+07
60%	1.173603e+07	1.594351e+07	4.807617e+06	1.198027e+07
80%	2.467311e+07	2.153157e+07	6.934026e+06	1.751845e+07
max	8.436698e+07	1.398867e+08	5.549406e+07	4.962898e+07

Table 7: Percentage total of OSR Streams to Total OSR

OSR Stream	% contribution to total OSR for FY 2013/2014	% contribution to total OSR for FY 2014/2015	% contribution to total OSR for FY 2015/2016	% contribution to total OSR for FY 2016/2017	% contribution to total OSR for FY 2017/2018	% contributio n to total OSR for FY 2018/2019
Property rates	27.13	23.14	15.77	12.69	17.69	17.96
Business permits	33.09	25.49	17.66	13.57	17.19	11.95
Sand gravel & quarry	34.16	36.64	17.35	23.13	26.00	33.15
Building plan approvals	4.96	9.52	14.66	15.88	17.19	11.74
Total	**99.34**	**94.79**	**65.44**	**65.27**	**78.07**	**74.80**

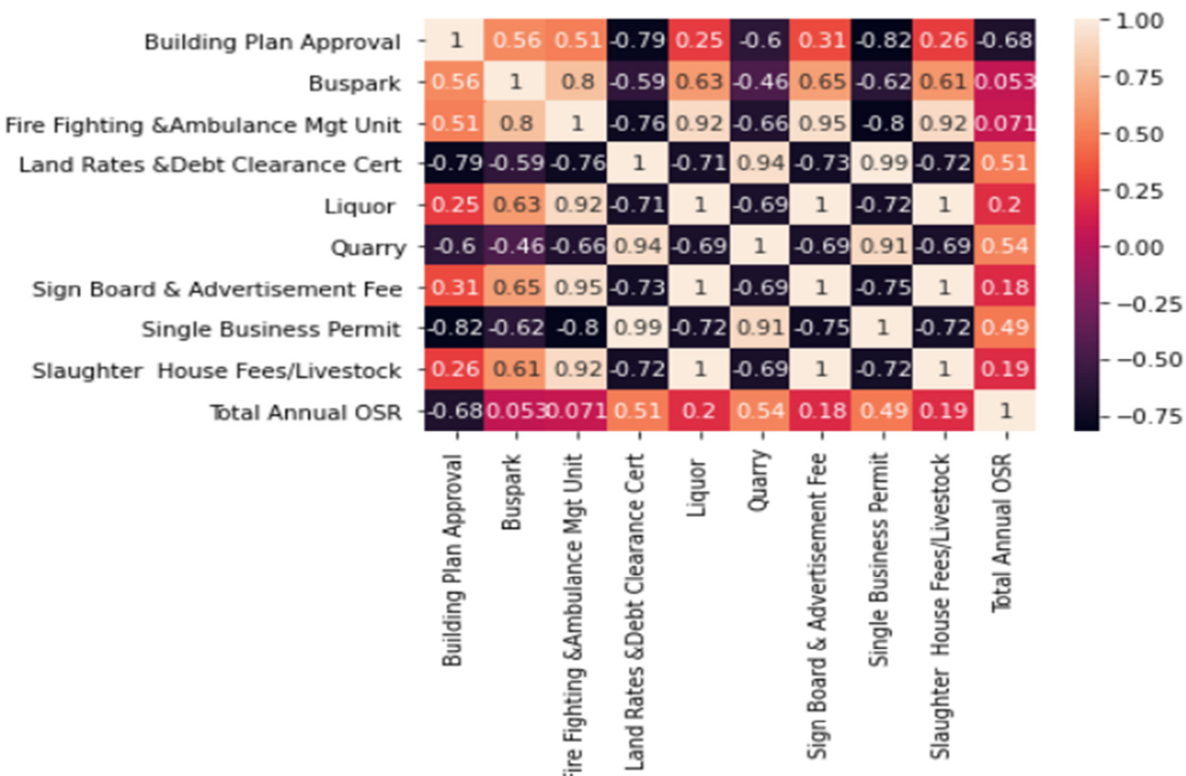
The four streams contribution to total OSR have well been over 60% for the period under review with the four accounting for almost 100% of the revenue collected in financial year 2013/2014.

4.3 Research Findings

4.3.1. Objective one Results

The first objective was to investigate and identify streams that influence own source revenue. Figure 7 shows a heatmap of the correlations between the top 9 revenue streams and total own source revenue. There is a strong positive correlation between land rates, single business permit, quarry and total own source revenue. The correlations were +0.51, +0.49 and +0.54 respectively. The three streams where therefore considered for this study. While the other revenue stream exhibited weak positive correlation, building approvals had a high negative correlation with total OSR. This is because of the strong negative correlation with lands rates; quarry and single business permits which contribute the highest to the total own source revenue. Regardless of this, the stream was chosen for this study because it also contributes a high proportion to own source revenue collected annually as shown in table 7. Of the four revenue streams identified, from the literature review: property rates, building plan approval and business licenses were identified as having adequate rationale and clear legal basis. This maybe an indication that revenue streams supported by adequate laws and clear policies, contribute the most to total OSR.

Figure 8: Correlation Matrix



4.3.2 Objective Two Results

The second objective is to develop a ARIMA models. This research will focus on the four identified OSR streams that contribute the bulk of local revenues. An ARIMA model has two parts: the autoregressive model (AR) and the Moving average model (MA).

The general model for the Auto Regressive model of order p {AR(p)} is given by:

$$X_t = Ø_1 X_{t-1} + Ø_2 X_{t-2} + \ldots.. + Ø_p X_{t-p} + e_t$$

Where; X_t is the monthly own source revenue for a particular stream collected monthly,

$Ø_p$ is the autoregressive model parameters,

X_{t-I} is the prior observations

e_t is the pure random process

The general model for the Moving Average model of order q {MA(q)} is given by:

$$X_t = e_t + \Theta_1 e_{t-1} + \Theta_2 e_{t-2} + \ldots.. + \Theta_q e_{t-q}$$

Where: Xt = is the monthly own source revenue for a particular stream collected monthly

Θ_i is the moving average parameters

e_{t-i} is the white noise error term

The general ARIMA model is given by:

$$X_t = Ø_1 X_{t-1} + Ø_2 X_{t-2} + \ldots.. + Ø_p X_{t-p} + e_t + \Theta_1 e_{t-1} + \Theta_2 e_{t-2} + \ldots.. + \Theta_q e_{t-q}$$

4.3.2.1 Building Approval OSR stream

Building approval revenues are generated by the county through its power to control development within its jurisdiction. The objectives of development control are: protect and conserve the environment, ensure orderly and planed building development, ensure optimal land use, safeguarding national security among others. Machakos County being one of the counties in the Nairobi metropolitan area has witnessed an upsurge in development and consequently development approval requests. The figure below show monthly revenue collected since the financial year 2013/2014 to 2018/2019.

Business approval revenues do not exhibit any apparent seasonality. However subjecting the data to the ADF test reveals the data is not stationary. After the first differential the p-values=0.000. Since the data achieves stationarity after the first differential, d=1, and the order of the ARIMA model is (p, 1, q). Using the information criterion for selecting the most suitable order of the ARIMA model, ARIMA (2,1,4) had the least AIC and was therefore selected as the best fitting model for forecasting building approval revenues.

Figure 9: Stepwise Search to minimize AIC for Building Approval Revenue ARIMA model

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=2485.303, Time=0.08 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=2489.794, Time=0.02 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=2488.858, Time=0.01 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=2484.710, Time=0.03 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=2487.906, Time=0.02 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=2482.302, Time=0.03 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=2484.240, Time=0.07 sec
 ARIMA(0,1,3)(0,0,0)[0] intercept   : AIC=2483.003, Time=0.07 sec
 ARIMA(1,1,3)(0,0,0)[0] intercept   : AIC=2476.021, Time=0.08 sec
 ARIMA(2,1,3)(0,0,0)[0] intercept   : AIC=2478.000, Time=0.10 sec
 ARIMA(1,1,4)(0,0,0)[0] intercept   : AIC=2477.266, Time=0.12 sec
 ARIMA(0,1,4)(0,0,0)[0] intercept   : AIC=2478.500, Time=0.06 sec
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=inf, Time=0.20 sec
 ARIMA(2,1,4)(0,0,0)[0] intercept   : AIC=2472.319, Time=0.33 sec
 ARIMA(3,1,4)(0,0,0)[0] intercept   : AIC=2478.640, Time=0.19 sec
 ARIMA(2,1,5)(0,0,0)[0] intercept   : AIC=inf, Time=0.47 sec
 ARIMA(1,1,5)(0,0,0)[0] intercept   : AIC=2480.131, Time=0.14 sec
 ARIMA(3,1,3)(0,0,0)[0] intercept   : AIC=2478.214, Time=0.15 sec
 ARIMA(3,1,5)(0,0,0)[0] intercept   : AIC=inf, Time=0.54 sec
 ARIMA(2,1,4)(0,0,0)[0]             : AIC=2474.768, Time=0.26 sec

Best model:  ARIMA(2,1,4)(0,0,0)[0] intercept
Total fit time: 3.003 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   72
Model:               SARIMAX(2, 1, 4)   Log Likelihood               -1228.160
Date:                Mon, 13 Sep 2021   AIC                           2472.319
Time:                        11:22:10   BIC                           2490.421
Sample:                             0   HQIC                          2479.518
                                 - 72
Covariance Type:                  opg
==============================================================================
```

4.3.2.2 Quarry

This OSR stream is mainly concerned with all revenue generated from extraction of mineral resources. This is sand, gravel, building stones. The Machakos County Management of Quarrying Activities Act of 2016 sets out the application process, granting of permit, cancellation and prohibited activities. It also sets the time that quarrying can be done and which roads can be used. Figure 12 shows monthly revenue collected from FY 2013/2014 to FY 2018/2019.

The ADF test on stationarity on quarry revenues gives p-value = 0.000209 which is less than 0.05 hence the data is stationary. Differencing will not be required. Since no differencing is required, the value of d=0. Performing a stepwise search of the best order of the model reveals that ARIMA (1,0,0) is the model with the least AIC hence most suitable for forecasting quarry OSR.

Figure 10: Stepwise search to minimize AIC for the Quarry ARIMA model

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=2666.461, Time=0.06 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=2672.879, Time=0.02 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=2664.589, Time=0.02 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=2665.538, Time=0.02 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=2717.318, Time=0.02 sec
 ARIMA(2,0,0)(0,0,0)[0] intercept   : AIC=2666.335, Time=0.03 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=2668.509, Time=0.18 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=2677.630, Time=0.00 sec

Best model:  ARIMA(1,0,0)(0,0,0)[0] intercept
Total fit time: 0.364 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   72
Model:               SARIMAX(1, 0, 0)   Log Likelihood               -1329.294
Date:                Mon, 13 Sep 2021   AIC                           2664.589
Time:                        13:55:59   BIC                           2671.419
Sample:                             0   HQIC                          2667.308
                                 - 72
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept   1.662e+07   4.66e-10   3.57e+16      0.000    1.66e+07    1.66e+07
ar.L1          0.3640      0.051      7.069      0.000       0.263       0.465
sigma2       6.41e+14   1.79e-17   3.58e+31      0.000    6.41e+14    6.41e+14
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):               981.90
Prob(Q):                              0.81   Prob(JB):                         0.00
Heteroskedasticity (H):               1.44   Skew:                             3.84
Prob(H) (two-sided):                  0.38   Kurtosis:                        19.38
===================================================================================
```

4.3.2.3 Property Rates

The time series plot for property rates revenue does exhibit some seasonality. The ADF test for stationarIty gives p-values = 0.018867 which is less than 0.05 hence the data is stationary. Using the information criterion, a stepwise search of the most suitable model give ARIMA (3,0,0) and is therefore chosen as the forecasting model.

Figure 11: Stepwise search to minimize AIC for the Property Rates ARIMA model

```
Performing stepwise search to minimize aic
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=2643.649, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=2593.308, Time=0.02 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=2624.566, Time=0.02 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=2586.915, Time=0.05 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=2588.744, Time=0.12 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=2588.671, Time=0.08 sec
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=2588.496, Time=0.10 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=2590.641, Time=0.13 sec
 ARIMA(2,0,0)(0,0,0)[0] intercept   : AIC=2582.254, Time=0.04 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=2582.672, Time=0.03 sec
 ARIMA(3,0,0)(0,0,0)[0] intercept   : AIC=2579.721, Time=0.08 sec
 ARIMA(4,0,0)(0,0,0)[0] intercept   : AIC=2581.007, Time=0.12 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=2581.028, Time=0.22 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=2582.333, Time=0.16 sec
 ARIMA(4,0,1)(0,0,0)[0] intercept   : AIC=2582.956, Time=0.18 sec

Best model:  ARIMA(3,0,0)(0,0,0)[0] intercept
Total fit time: 1.404 seconds
```

4.3.2.4. Single Business Permits

Single business permits exhibit a very a strong seasonal component. The developed model therefore must incorporate a seasonal component. When the single business permits OSR data is subjected to the ADF test, it is found to be stationary. To remove seasonality and make the data stationary, the data is differenced seasonally. The data becomes stationary after the first diffrential. The final model will therefore take the form SARIMA (p,0,q) (P,1,Q)(12) model. Using the information critertio, a stepwise search reveal SARIMA (0,0,0) (0,1,1)(12) as the model with least AIC and therefore the most suited to forecast single business permits.

4.3.3. Objective Three

The third objective was to evaluate the effectiveness of the developed models. Their effectiveness is gauged by their ability to forecasts revenue as close as possible to the actual revenues with the least amount of error. The graphical display of the models predictive ability is displayed in figure 16. The forecasted values lie within the 95% confidence interval. The developed models were used in the forecast of an out of sample forecast for the last 5 months of the financial year 2018/2019. Figures 12 to 15 show the MAPE of the respective ARIMA models. The mean absolute percentage error (MAPE) is relatively low in all the models.

From table 8, the average forecast error for prediction made with an ARIMA model for building revenue is 1.94% which is lower than that of forecasts made without any model which stand at 3.06%. the same case applies for property rates, quarry and single business permits which show the percentage forecast error as 0.94%, 1.84% and 1.95% respectively against forecasts made without an ARIMA model which stand at 1.40%, 3% and 2.03 %. On the other when the MAPE results are considered, it means that the forecasts are on average 18.95%, 13.29%, 20.71% and 17.87% off from the actual values of building approvals revenues, property rates revenue, quarry revenue and single business permits revenue respectively.

Using the developed ARIMA models to forecast the four revenue streams improve the overall forecast of total own source revenue. Figure 16 shows that the ARIMA own source revenue forecasts closely simulated the actual own source revenue for the month of February to June 2019. Also shown are the comparison between the actual, ARIMA forecasts and expert judgment forecast for the four revenue streams. The ARIMA models display better forecasts.

Figure 12: Building Approval ARIMA model Accuracy Measure

```
# Building Approval Accuracy Metrics
def forecast_accuracy(forecast, actual):
    mape = np.mean(np.abs(forecast - actual)/np.abs(actual))  # MAPE
    return({'mape':mape})

forecast_accuracy(fc, test.values)
```

```
{'mape': 0.1895639515986218}
```

Figure 13: Property Rates ARIMA Model Accuracy Measure

```
# Property Rates Accuracy metrics
def forecast_accuracy(forecast, actual):
    mape = np.mean(np.abs(forecast - actual)/np.abs(actual))  # MAPE
    return({'mape':mape})

forecast_accuracy(fc, test.values)
```

```
{'mape': 0.1329423105295208}
```

Figure 14: Quarry ARIMA Model Accuracy Measure

```
# Quarry (1,0,0) Accuracy metrics
def forecast_accuracy(forecast, actual):
    mape = np.mean(np.abs(forecast - actual)/np.abs(actual))  # MAPE
    return({'mape':mape})

forecast_accuracy(fc, test.values)
```

```
{'mape': 0.2071478891755844}
```

Figure 15: Single Business Permits ARIMA Model Accuracy Measure

```
# Single Business Permits Accuracy Metrics
def forecast_accuracy(forecast, actual):
    mape = np.mean(np.abs(forecast - actual)/np.abs(actual))  # MAPE
    return({'mape':mape})

forecast_accuracy(fc, test.values)
```

```
{'mape': 0.17872613074622595}
```

Table 8 Predicted and Actual Revenues figures and errors

	Actual	Ordinary Forecasts	ARIMA Forecasts	% Error of ordinary Forecasts	% error of ARIMA forecasts
Building Approvals					
Feb-19	1, 564,473.00	12, 545,000.00	7, 255,242.00	8.02	4.64
Mar-19	2, 025,331.00	10, 000,000.00	6, 304,554.00	4.94	3.11
Apr-19	46, 219,402.00	10, 540,300.00	5, 590,626.00	0.23	0.12
May-19	49, 628,983.00	30, 376,000.00	35, 482,800.00	0.61	0.71
Jun-19	33, 492,496.00	50, 245,460.00	37, 240,820.00	1.50	1.11
			Average	**3.06**	**1.94**
Singles Business Permits					
Feb-19	9, 908,718.00	20, 500,000.00	16, 423,950.00	2.07	1.66
Mar-19	4, 558,074.00	30, 050,400.00	24, 824,060.00	6.59	5.45
Apr-19	67, 489,442.00	25, 167,342.00	43, 578,770.00	0.37	0.65
May-19	38, 468,557.00	21, 609,245.00	31, 593,220.00	0.56	0.82
Jun-19	27, 882,655.00	16, 048,800.00	32, 574,070.00	0.58	1.17
			Average	**2.03**	**1.95**
Quarry					
Feb-19	4, 655,445.00	31, 760,000.00	20, 677,300.00	6.82	4.44
Mar-19	3, 528,411.00	23, 102,900.00	10, 341,900.00	6.55	2.93
Apr-19	149, 821,506.00	16, 000,500.00	18, 291,760.00	0.11	0.12
May-19	139, 807,741.00	50, 120,400.00	97, 939,380.00	0.36	0.70
Jun-19	62, 547,017.00	73, 100,345.00	63, 605,370.00	1.17	1.02
			Average	**3.00**	**1.84**
Property Rates					
Feb-19	6, 017,582.00	15, 000,000.00	7, 990,360.00	2.49	1.33
Mar-19	5, 672,393.00	18, 425,000.00	8, 294,860.00	3.25	1.46
Apr-19	77, 527,776.00	30, 681,055.00	62, 183,459.00	0.40	0.80
May-19	44, 213,825.00	26, 300,000.00	34, 597,449.00	0.59	0.78
Jun-19	84, 366,976.00	21, 950,000.00	27, 630,250.00	0.26	0.33

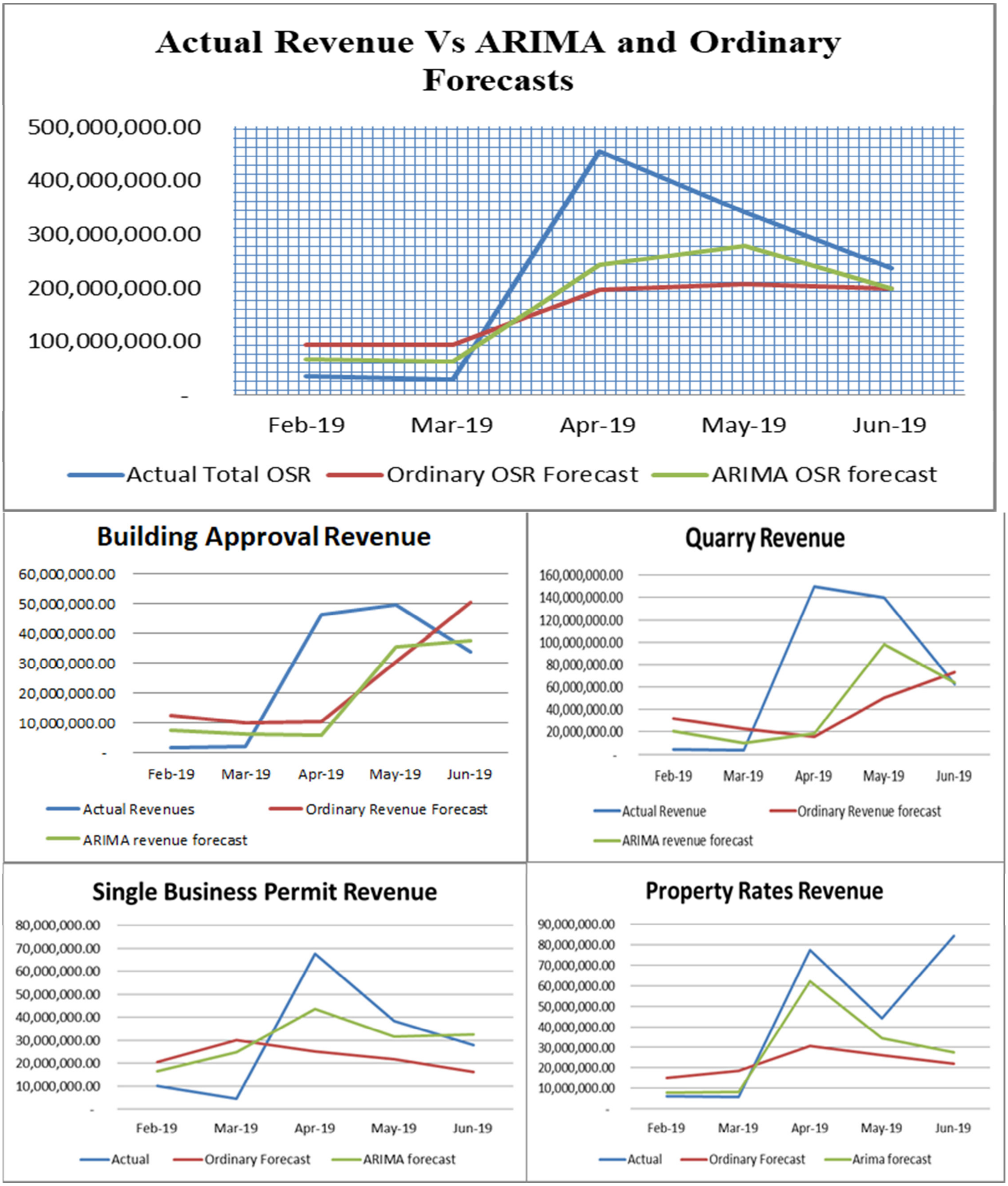
4.4 Discussion of Results

The four revenue streams make up the bulk of the revenue from own revenue sources. This study was based on the premise that if the four own source revenue streams can be forecasted accurately, then accuracy of the overall own source revenue would improve. Single Business permit revenue stream was found to have a seasonal component which repeated itself after every 12 months. Quarry, property rates and building approval did not exhibit any seasonal variations. From the correlation matrix, business approval was found to have a high negative correlation with the other three revenue streams.

Corvalao et al (2010) compared the prediction ability of a regression model and ARIMA and came to the conclusion that ARIMA models were superior. This study compared expert judgment and the ARIMA model and concluded that the ARIMA models had greater predictive ability. Brojba et al (2010) arrived at a similar conclusion but on condition that they are suitable for short-term forecasts. The Ahmed et al (2020) shared the same view that ARIMA models were suitable for short-term forecasts. This study arrives at the same conclusion. Long term forecast are problematic and are bound to have errors are other exogenous variables such as a different tax regime, unplanned waivers and economic shocks may influence the variable of interest. To improve on forecast accuracy, this study developed forecasting models for the top four revenue streams as opposed to a single model for total revenue. this method was also employed by larmore(2016) to forecast revenue. This study concludes that concludes that ARIMA models can be used to forecast OSR revenues with a higher degree of accuracy in the counties.

Figure 16: Graphical display of model outcomes

Figure 17: Graphs Displays



4.5 Summary

The results shows that using the ARIMA models improves forecasting of own source revenue. Developed ARIMA models for each of the four revenue streams are shown in Tables 10. The single business permit and property rates were found to have seasonal components while quarry and Building approval did not have any seasonal element.

Table 9: ARIMA Model for the four identified stream

OSR Stream	Model	MAPE
Building Approval	ARIMA (2,1,4)	18.95%
Quarry	ARIMA (1,0,0)	13.29%
Property Rates	ARIMA (3,0,0)	20.71%
Single Business Permit	SARIMA (0,0,0)(0,1,1)(12)	17.87%

By looking at the average forecasting errors the ARIMA forecasting models outperformed the existing forecasting techniques used in the county of Machakos in all the four revenue streams.

CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Forecasting of total own source revenue improved when ARIMA models were used to forecast the four highest revenue streams contributing to total own source revenue. When compared to the expert judgment method currently in use by the county, the ARIMA models proved to be more superior. The ARIMA models also generated lower errors compared to the expert judgment approach used by the county of Machakos. This study therefore shows that the overall accuracy of the forecasted own source revenue can be improved by using ARIMA models to forecast individual revenue streams then summing up them up to give the total own source revenue.

From objective one results, it can be concluded that, revenue streams that have adequate policy rationale and clear legal basis have contribute the highest to the overall OSR. For counties to boost their own source revenue collections therefore, it is imperative that all revenue streams be properly anchored in the law and have clear policy rationale.

From the second objective, four ARIMA models were developed to predict the four own source revenue streams. While it is possible to model one ARIMA model using the total OSR figures, this approach would have ignored the unique nature of each of the constituent own source revenue streams. The four model developed were found to be the best because they had the least Akaike Information Criterion (AIC). Additionally, the simulation results provided the least errors compared to the existing forecasting methods.

From the third objective, the developed models were evaluated on their ability to make accurate forecasts. The mean absolute percentage error (MAPE) measure was used as an accuracy measure and all model had a MAPE of 20% and below. Also the forecast error of the two forecasts: ARIMA forecast and other methods forecast and the ARIMA model forecasts were found to have the least forecast error. This proved true for all the four streams.

5.3 Contribution of the study

This study contributes to the overall debate on counties need to increase and accurately forecast their own source revenue and cut over reliance on the equitable national share of revenues disbursed by the national government. It shows that the ARIMA models are superior to the expert judgment forecasting approach currently in use. Empirical models such as the ARIMA models when used in forecasting do result in improved accuracy. Through accurate forecasting, counties are able to set realistic and sustainable budgets. Own source revenue streams present a great source of county revenues and ability to accurately forecast them ensure sound fiscal management as well as sound policy formulation at the micro and macro-economic levels.

5.4 Recommendations

This study only considered the ARIMA forecasting models. However attributes of the data itself may determine which accuracy measure are most suitable. In the case of own source revenue persistent fees and levies waivers and increases in fees and charges levied by the counties cause would lead to errors in the forecasting models or their inability to produce accurate forecasts. Therefore, in evaluating which model and the accuracy metrics to use, the data itself must be interrogated further. Future research should address the application and modeling of other models such as hybrid models, artificial neural networks, conditional volatility models, VAR model among others that may provide forecasts that are close to the actual time series and of greater accuracy.

References

Republic of Kenya. The National Treasury and Planning. (2019). National Policy to Support Enhancement of County Governments' Own-Source Revenue. *Government Printer, Nairobi.*

Omamo, W. O. (1995). The Commission of Inquiry on Local Authorities in Kenya. *Government Printer, Nairobi*.

Ntoiti, J., Gakure, R. W., Waititu, A., & Gekara, M. (2013). The Contribution Of Financial Management Practices to Financial Challenges Facing Local Authorities In Service Delivery In Kenya. *CLEAR International Journal of Research in Commerce & Management*, *4*(3).

Mitullah, W. & Waema, (2007). ICTs and Financial Management in Local Authorities in Kenya: Case Study of Mavoko and Nyeri Municipal Councils, A Project of LOG-IN Africa Initiative Technical Report

Kenya Anti-Corruption Commission. Directorate of Preventive Services. (2007), Examination Report on the Systems, Polices, Procedures and Practices of the City Council of Nairobi, Directorate of Preventive Services.

Inter-Governmental Relations Technical Committee. (2018). Report on the Identification, Verification, Validation and Transfer of Assets and Liabilities of the Defunct Local Authorities.

Republic of Kenya. The National Treasury and Planning. (2020). 2020/2021 Budget Policy Statement.

Controller of Budget. (2016). Annual County Government Budget Implementation Review Report for the Financial Year 2015/16.

Sun, J. (2005). The dynamics of government revenue forecasting from an organizational perspective: A review of the literature. *Journal of Public Budgeting, Accounting & Financial Management*.

Sun, J., & Lynch, T. D. (Eds.). (2008). *Government budget forecasting: Theory and practice*. CRC Press.

Beckett-Camarata, J. (2006). Revenue forecasting accuracy in Ohio local governments. *Journal of Public Budgeting, Accounting & Financial Management*.

Weller, B. R., & Kurre, J. A. (1999, June). A Leading Index for Small Metropolitan Areas. In *19th International Symposium on Forecasting, Washington, DC*.

Ahmed, A., Amin, S. B., & Khan, A. M. (2020). Forecasting Tourism Revenue in Bangladesh Using ARIMA Approach: The Case of Bangladesh. *International Review of Business Research Papers*, *16*(1).

Ayakeme, T. I., Biu, O. E., Enegesele, D., & Wonu, N. (2021). Forecasting of bayelsa state internally generated revenue using ARIMA model and winter methods.

Temur, A. S., Akgün, M., & Temur, G. (2019). Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models.

Commission on Revenue Allocation. (2021). Counties' Efforts Towards Revenue Mobilization: A Stock of the Last Six Years.

Larmore, E. (2016). *Methodologies for forecast modeling for small areas with limited data availability and unique tax structures*. University of Nevada, Reno.

Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, *18*(3), 1014-1020.

Farhath, Z. A., Arputhamary, B., & Arockiam, L. (2016). A survey on ARIMA forecasting using time series model. *Int. J. Comput. Sci. Mobile Comput*, *5*, 104-109.

Gujarati, D. N., & Porter, D. C. (1999). *Essentials of econometrics* (Vol. 2). Singapore: Irwin/McGraw-Hill.

International Budget Partnership kenya. (2019, March). *Are Kenya Counties Making Budget Documents Available to the Public? A Review of County Websites*. https://www.internationalbudget.org

Saayman, A., & Botha, I. (2015). Evaluating non-linear approaches in forecasting tourist arrivals. *Economic Research Southern Africa. Tourism Economics*, *23*(3), 594-613.

Box, G. E., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, *70*(349), 70-79.

Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, *19*(3), 435-451.

Wabomba, M. S., Mutwiri, M. P., & Mungai, F. (2016). Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models. *Science Journal of Applied Mathematics and Statistics*, *4*(2), 64-73.

Lim, C., & McAleer, M. (1999). A seasonal analysis of Malaysian tourist arrivals to Australia. *Mathematics and computers in simulation*, *48*(4-6), 573-583.

Cuhadar, M. (2014). Modelling and forecasting inbound tourism demand to istanbul: A comparative analysis. *European Journal of Business and Social Sciences*, *2*(12), 101-119.

Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.

Box, G. E. P., & Jenkins, G. M. (1970). Control. *Halden-Day, San Francisco*.

Makoni, T., & Chikobvu, D. (2018). Modelling and forecasting Zimbabwe's tourist arrivals using time series method: a case study of Victoria Falls rainforest. *Southern African Business Review*, *22*(1).

Makananisa, M. P. (2015). Forecasting annual tax revenue of the South African taxes using time series Holt-Winters and ARIMA/SARIMA Models. *South Africa: University of South. Africa*.

Corvalão, E. D., Samoryl, R. W., & Brasil, G. H. (2010). Forecasting the collection of the state value added tax (ICMS) in Santa Catarina: the general to specific approach in regression analysis. *Brazilian Journal of Operations & Production Management*, *7*(1), 105-121.

Pelinescu, E., Anton, L. V., Ionescu, R., & Tasca, R. (2010). The analysis of local budgets and their importance in the fight against the economic crisis effects. *Romanian Journal of Economic Forecasting*, 17-32.

Brojba, L. C., Dumitru, C. G., & Belciug, A. V. (2010). On the use of some optimal strategies of fiscal administration during economic crisis. *Romanian Journal of Economic Forecasting*, *1*, 151-164.

Jayesekara, N., & Passty, B. W. (2009). City of Cincinnati income tax revenue: A forecast based on economic data and causal inference. *Cincinnati, Ohio: University of Cincinnati*.

Slobodnitsky, T., & Drucker, L. (2008). VAT revenue forecasting in Israel. *Ministry of Finance, State Revenue Administration, The Maurice Falk Institute for Economic Research in Israel Ltd., Digest*, *10*.

Saayman, A., & Saayman, M. (2008). Determinants of inbound tourism to South Africa. *Tourism economics*, *14*(1), 81-96.

Saayman, A., & Botha, I. (2015). Evaluating non-linear approaches in forecasting tourist arrivals. *Economic Research Southern Africa. Tourism Economics*, *23*(3), 594-613.

Ndiege, B. O. (2015). An Analysis of the Performance of International Tourism Demand in Tanzania. *ARA: Journal of Tourism Research*, *5*(1), 53-62.

Rahmat, R. F., Syaputra, A. H., Faza, S., & Arisandi, D. (2020, May). Prediction of Regional Revenue and Expenditure Budget using Autoregressive Integrated Moving Average (ARIMA). In *IOP Conference Series: Materials Science and Engineering* (Vol. 851, No. 1, p. 012064). IOP Publishing.

Liu, Q., & Wang, H. (2015). Research on the forecast and development of China's public fiscal revenue based on ARIMA model. *Theoretical Economics Letters*, *5*(04), 482.

Almasarweh, M., & Alwadi, S. (2018). ARIMA model in predicting banking stock market data. *Modern Applied Science*, *12*(11), 4.

Ekmiş, M. A., Hekimoğlu, M., & Atak Bülbül, B. (2017). Revenue Forecasting Using a Feed-Forward Neural Network and ARIMA Mod

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.

Hsu, L. C. (2010). A genetic algorithm based nonlinear grey Bernoulli model for output forecasting in integrated circuit industry. *Expert systems with Applications*, *37*(6), 4318-4323.

Li, G., Wong, K. K., Song, H., & Witt, S. F. (2006). Tourism demand forecasting: A time varying parameter error correction model. *Journal of Travel Research*, *45*(2), 175-185.

Golosov, M. (2002). Tax revenue forecasts in IMF-supported programs.

Kyobe, A. J., & Danninger, S. (2005). Revenue forecasting-How is it done? Results from a Survey of Low-Income Countries.

Urrutia, J. D., Mingo, F. L. T., & Balmaceda, C. N. M. (2015). Forecasting Income Tax Revenue of the Philippines Using Autoregressive Integrated Moving Average (Arima) Modeling: a Time Series Analysis. *Am. Res. Thoughts*, *1*(9), 1938-1992.

Etuk, E. H., & Igbudu, R. C. (2013). A Sarima Fit to Monthly Nigerian Naira-British Pound Exchange Rates. *Journal of Computations & Modelling*, *3*(1), 133-144.

Otu, O. A., Osuji, G. A., Opara, J., Mbachu, H. I., & Iheagwara, A. I. (2014). Application of Sarima models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, *2*(1), 16-28.

Koirala, T. P. (2011). Government revenue forecasting in Nepal. *NBR Economic Review*, *4*(2).

Ramanathan R. 2002. "Introductory Econometrics with Applications." 5th Edition, Thomson Asia Pte Ltd., Singapore

Wong, J. D. (1995). Local government revenue forecasting: Using regression and econometric revenue forecasting in a medium-sized city. *Journal of Public Budgeting, Accounting & Financial Management*.

Stewart, T. R., Moninger, W. R., Brady, R. H., Merrem, F. H., Stewart, T. R., & Grassia, J. (1989). Analysis of expert judgment in a hail forecasting experiment. *Weather and forecasting*, *4*(1), 24-34.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision support systems*, *11*(1), 1-24.

Hill, T., O'Connor, M., & Remus, W. (1996). Neural network models for time series forecasts. *Management science*, *42*(7), 1082-1092.

Wang, M., Wang, Y., Wang, X., & Wei, Z. (2015). Forecast and analyze the telecom income based on arima model. *The Open Cybernetics & Systemics Journal*, *9*(1).

Botrić, V., & Vizek, M. (2012). Forecasting fiscal revenues in a transition country: the case of Croatia. *Zagreb International Review of Economics & Business*, *15*(1), 23-36.

Wong, J. D. (1995). Local government revenue forecasting: Using regression and econometric revenue forecasting in a medium-sized city. *Journal of Public Budgeting, Accounting & Financial Management*.

OFORI, M. S., FUMEY, A., & NKETIAH-AMPONSAH, E. (2021). Forecasting Value Added Tax Revenue in Ghana. *Journal of Economics and Financial Analysis*, *4*(2), 63-99.

Beckett-Camarata, J. (2006). Revenue forecasting accuracy in Ohio local governments. *Journal of Public Budgeting, Accounting & Financial Management*.

Bretschneider, S. I., Gorr, W. L., Grizzle, G., & Klay, E. (1989). Political and organizational influences on the accuracy of forecasting state government revenues. *International Journal of Forecasting*, *5*(3), 307-319.

Grizzle, G. A., & Klay, W. E. (1994). Forecasting state sales tax revenues: comparing the accuracy of different methods. *State & Local Government Review*, 142-152.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision support systems*, *11*(1), 1-24.

Stewart, T. R., Moninger, W. R., Brady, R. H., Merrem, F. H., Stewart, T. R., & Grassia, J. (1989). Analysis of expert judgment in a hail forecasting experiment. *Weather and forecasting*, *4*(1), 24-34.

Suwanvijit, W. (2014, November). Forecasting Tourist Arrivals in IMT-GT using Lee-Carter Method. In *AFBE 2014 CONFERENCE PAPERS (Thaksin University)* (p. 240).

Silvestrini, A., Salto, M., Moulin, L., & Veredas, D. (2008). Monitoring and forecasting annual public deficit every month: the case of France. *Empirical Economics*, *34*(3), 493-524.

Berwick, M. D., & Malchose, D. C. (2012). Forecasting North Dakota fuel tax revenue and license and registration fee revenue.

Jain, A., & Kumar, A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, *7*(2), 585-592.

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, *135*, 109864.

Ioannidis, J. P., Cripps, S., & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International journal of forecasting*.

Hill, T., O'Connor, M., & Remus, W. (1996). Neural network models for time series forecasts. *Management science*, *42*(7), 1082-1092.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., ... & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of forecasting*, *1*(2), 111-153.

Foster, W. R., Collopy, F., & Ungar, L. H. (1992). Neural network forecasting of short, noisy time series. *Computers & chemical engineering*, *16*(4), 293-297.

Iman, R. L., & Conover, W. J. (1988). *Modern business statistics*. John Wiley & Sons, Inc..

Sharda, R., & Patil, R. B. (1992). Connectionist approach to time series prediction: an empirical test. *Journal of Intelligent Manufacturing*, *3*(5), 317-323.

Buettner, T., & Kauder, B. (2010). Revenue forecasting practices: Differences across countries and consequences for forecasting performance. *Fiscal Studies*, *31*(3), 313-340.

Jenkins, G. P., Kuo, C. Y., & Shukla, G. (2000). Tax Analysis and Revenue Forecasting: Issues and Techniques.[on-line], Boston. *Harvard Institute for International Development, c2000,[cited 8 th August, 2013],< http://jdintl. econ. queensu. ca/publications/qed_dp_169. pdf.*

Weatherford, L. R., & Kimes, S. E. (2003). A comparison of forecasting methods for hotel revenue management. *International journal of forecasting*, *19*(3), 401-415.

McNichol, E. C. (2014). Improving state revenue forecasting: best practices for a more trusted and reliable revenue estimate. *Center on Budget and Policy Priorities-CBPP, set*.

Williams, D. W., & Kavanagh, S. C. (2016). Local government revenue forecasting methods: Competition and comparison. *Journal of Public Budgeting, Accounting & Financial Management*.