

Vehicle ownership for South Africa: Developing a forecasting model and assessing household vehicle ownership

By

Trishia Thokozile Mtembu



Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce (Transport Economics) in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Prof S. Krygsman

March 2020

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated) , that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

ABSTRACT

This study provides details on the findings of an analysis of data on the South African vehicle population and data from the National Household Travel Survey conducted in 2013; an analysis to forecast vehicle population and household vehicle ownership.

The study analysed historical data from two data bases, namely eNaTiS and the 2013 National Household Travel Survey. The aim for both data sets was to forecast vehicle numbers and household vehicle ownership respectively. Vehicle population is important for economic development, policy and planning concerning road infrastructure, therefore the study analysed the real GDP as an indicator of economic growth and development, and the South African population as an influence driving vehicle demand.

The historical data used shows an annual average compound growth rate of 3.07% for vehicle population and a predicted annual average compound growth rate of 2.22%. Though a decrease in growth rate, the vehicle population is predicted to grow to 17 637 672 vehicles in 2038 compared to a population of 71 452 500 people, thus an estimated ratio of 247 vehicles per 1 000 people in 2038.

The study applies Multinomial Logistics Regression, an essential method for categorical data, to the National Household Travel Survey 2013 data. All the variables within the model are categorical, and thus this model is evidently a significant fit to the data. The dependent variable in the model is the number of vehicles owned by a household in 3 categories (0, 1 and 2 or more vehicles) and the independent variables consist of: main dwelling, income quintiles, geographical location and total household expenditure.

The analysis shows that the probability of households in the lowest income quintile owning 1 vehicle compared to owning no vehicles was 0.161 times lower than the odds of households in the highest income quintile. Households with total household expenditure of R0-R300 were respectively 0.056 and 0.011 times less likely to influence household ownership of 1 and 2 or more vehicles. However, households residing within metropolitan areas are 1.182 times more likely to influence household ownership of 1 vehicle.

ACKNOWLEDGEMENTS

Foremost, I would like to acknowledge God, for His continuous presence in my life and grace in all things. Through the difficult moments in the process of this master's degree His presence was my footstool: 2 Corinthians 1v20 "Whatever God has promised gets stamped with the Yes of Jesus. In Him, this is what we preach and pray, the great amen." Indeed this journey was stamped with a Yes.

My sincere gratitude of appreciation is extended to my supervisor, Prof Krygsman, for his continuous support in my writing of this thesis, his patience in helping me and the knowledge he shared throughout. His input assisted me in doing research and writing this thesis.

Last and most important, a huge thank you to my family and friends, for the emotional and spiritual support that you continuously so freely gave on this journey. Your words of encouragement kept me going even in the toughest moments; you are all greatly appreciated.

ABBREVIATIONS AND TERMINOLOGY

OECD	Organisation for Economic Cooperation and Development
NHTS	National Household Travel Survey
StatsSA	Statistics South Africa
EAs	Enumeration areas
PSUs	Primary Sampling Units
DUs	Dwelling Units
NMT	Non-Mobilised Transport
eNaTIS	The National Traffic Information System
RTMC	Road Traffic Management Corporation
GDP	Gross domestic product
SPSS	Statistics Package for Social Sciences
ARIMA	Auto Regressive Integrated Moving Average
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
NRTF	National Road Traffic Forecasts
MAPE	Mean Absolute Percentage Error
GDPPC	Gross Domestic Production Per Product
GNPPC	Gross National Product Per Capita
SSET	Sum Square of Error Terms
FEM	Fixed Effects Model

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
ABBREVIATIONS AND TERMINOLOGY	v
LIST OF FIGURES	viii
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Significance of the study	2
1.3 Aim and objectives	3
1.4 Structure of the study	3
CHAPTER 2: LITERATURE REVIEW	5
2.1 Introduction	5
2.2 Vehicle ownership forecasting models or techniques	5
2.3 Models of vehicle ownership	6
2.3.1 Aggregate Time Series Models	6
2.3.2 Aggregate Cohort Models	7
2.3.3 Pseudo-panel methods	7
2.4 Factors influencing vehicle ownership	12
2.5 Review of South African literature on vehicle ownership forecasting	14
2.6 Review of International literature on vehicle ownership forecasting	19
2.7 Vehicle ownership forecasting in low-income countries	29
2.8 Summary and conclusion	30
CHAPTER 3: DATA	31
3.1 Introduction	31
3.2 Historical vehicle population	31
3.3 2013 South African Household Travel Survey (NHTS)	37
3.3.1 Introduction to 2013 National Household Travel Survey	37
3.3.2 Limitations of the National Household Travel Survey	39
3.3.3 Key findings in the National Household Travel Survey 2013	40
3.4 Descriptive statistics	41
3.5 Summary and conclusion	51
CHAPTER 4: METHODOLOGY	52
4.1 Introduction	52

4.2	Research questions	52
4.3	Research design	52
CHAPTER 5: MODEL SPECIFICATION.....		55
5.1	Introduction	55
5.2	Vehicle forecast	55
5.2.1	Holt’s method	56
5.2.2	Forecasting sheet.....	57
5.2.3	SPSS forecasting.....	58
5.3	Vehicle ownership.....	59
5.3.1	Multinomial logit model	59
5.4	Summary and conclusion	60
CHAPTER 6: RESULTS ANALYSIS AND DISCUSSION: VEHICLE FORECAST		62
6.1	Introduction	62
6.2	Vehicle forecast	62
6.2.1	Holt’s method results	62
6.2.2	Forecast sheet results.....	68
6.2.3	SPSS Results	75
6.2.4	Comparison and conclusion of forecast models and methods for vehicle forecasting ...	81
CHAPTER 7: VEHICLE OWNERSHIP MODEL		87
7.1.	Vehicle ownership.....	87
7.1.1.	Evaluating the usefulness of the model	90
7.1.2.	Relationship between the independent variables and dependent variable	90
7.1.3.	Model estimation	91
7.2.	Summary and conclusion	97
CHAPTER 8: CONCLUSIONS		99
8.1.	Future research.....	103
REFERENCES		105
APPENDIX A: Questions extracted from NHTS 2013		110

LIST OF FIGURES

Figure 2.1: Quasi-logistic relationship for vehicle ownership.	22
Figure 3.1: Scatterplot of countries comparing vehicles per 1 000 of the population in 2014.	32
Figure 3.2: Scatterplot of countries comparing total kilometres of paved and unpaved road	33
Figure 3.3: Annual vehicle population.	34
Figure 3.4: Comparing vehicle population, population and real GDP 1986-2016	35
Figure 3.5: Comparison of vehicle population, human population and real GDP growth rate	36
Figure 3.6: Scatterplot representing real GDP with trend line	37
Figure 3.7: Three phases of data collection.	38
Figure 3.8: Sampled travel analysis zones.	39
Figure 3.9: Percentage of households by province	42
Figure 3.10: Household percentage by geographical location	43
Figure 3.11: Household percentage by main dwelling	44
Figure 3.12: Household percentage by income quintiles	45
Figure 3.13: Household percentage by total household expenditure	47
Figure 3.14: Household percentage by income category	46
Figure 3.15: Household percentage by population group	48
Figure 3.16: Household percentage by education institution	49
Figure 3.17: Household percentage by highest grade achieved	50
Figure 6.1: Holt's forecasted and actual vehicle population	64
Figure 6.2: Holt's forecasted and actual real GDP	66
Figure 6.3: Holt's forecasted and actual South African population	68
Figure 6.4: Forecast sheet forecasted and actual vehicle population, with confidence bounds	70
Figure 6.5: Forecast sheet forecasted and actual real GDP, with confidence bounds	72
Figure 6.6: Forecast sheet forecasted and actual South African population, with confidence bounds	74
Figure 6.7: SPSS forecasted and actual South African population, with confidence level and fit	77
Figure 6.8: SPSS forecasted and actual real GDP, with confidence levels and fit	79
Figure 6.9: SPSS forecasted and actual South African population, with confidence levels and fit	81
Figure 6.10: Holt's method vs forecast sheet vs SPSS forecasted vehicle population	83
Figure 6.11: Holt's method vs forecast sheet vs SPSS forecasted real GDP	84
Figure 6.12: Holt's method vs forecast sheet vs SPSS forecasted South African population	86
Figure 8.1: Vehicles per 1000 of the population including South Africa 2038	100

LIST OF TABLES

Table 2.1: Comparison of types of vehicle ownership models	9
Table 3.1: Summary of the number of primary sampling units.....	38
Table 3.2: Survey response rates	40
Table 6.1: Forecasted vehicle population for period 2018-2018 using Holt's method.	63
Table 6.2: Vehicle population Holt's method parameters.....	64
Table 6.3: Real GDP forecasted values for 2018-2038 using Holt's method	65
Table 6.4: Real GDP Holt's parameters	66
Table 6.5: South African population forecasted values for 2018-2038 using Holt's method ..	67
Table 6.6: South African population Holt's parameters	67
Table 6.7: Vehicle population forecasted values and confidence bounds for 2018-2038 using Forecast sheet	68
Table 6.8: Vehicle population forecast sheet parameters	69
Table 6.9: Real GDP forecasted values and confidence bounds for 2018-2038 using forecast sheet	71
Table 6.10: Real GDP forecast sheet parameters	71
Table 6.11: South African population forecasted values and confidence bounds for 2018-2038 using forecast sheet	73
Table 6.12: South African population forecast sheet parameters	74
Table 6.13: Vehicle population forecasted values and confidence levels for 2018-2038 using SPSS.....	75
Table 6.14: Vehicle population SPSS parameters	76
Table 6.15: Real GDP forecasted values and confidence levels for 2018-2038 using SPSS	78
Table 6.16: Real GDP SPSS parameters	79
Table 6.17: South African population forecasted values and confidence levels for 2018-2038 using SPSS	80
Table 6.18: South African population SPSS parameters	81
Table 6.19: Holt's method vs forecast sheet vs SPSS forecasted vehicle population	82
Table 6.20: Holt's method vs forecast sheet vs SPSS forecasted real GDP	84
Table 6.21: Holt's method vs forecast sheet vs SPSS forecasted South African population.	85
Table 6.22: Vehicle population, real GDP and South African population RMSE for all three forecast methods.....	86
Table 6.23: National Household Travel Survey 2013 household vehicle access summary...	87
Table 6.24: Model fitting information	88
Table 6.25: Case processing summary.....	88
Table 6.26: Pseudo R-Square.....	89
Table 6.27: Classification table	90
Table 6.28: Likelihood ratio tests	91
Table 6.29: Parameter Estimates.....	92

CHAPTER 1: INTRODUCTION

1.1 Introduction

Worldwide there is an increasing demand for vehicles, causing unsustainable levels of congestion and increases in resource use. With the constant growth in the number of private vehicles on the national roads of South Africa, vehicle ownership forecasting has become a field of study that for many transport-planning companies is of great interest, and indeed critically important. The ability to approximate the future number of vehicles used in a country can assist governments in timeous planning and policy formulation and ensure adequate infrastructure provision. With a growing population, the need to travel is ever increasing and at the same time and as a result, the anticipation of an increase in vehicle ownership is inevitable. To the best of the researcher's knowledge not much research on this topic has been done in the South African context.

This is somewhat surprising given the importance of the vehicle fleet for various policy questions including the need to construct roads, public transport provision, and planning for congestion. Furthermore, various Government revenue sources are tied to vehicle ownership and use, most importantly, the fuel levy, Value Added Tax on vehicle purchase and the Road Accident Fund Levy. These revenue streams are particularly important to the Government in executing a mandate to maintain, upgrade and expand the transport network for the country.

The increase in vehicle ownership is not an unfamiliar trend for a country, rather an already existing pattern within the economies of developing countries. The increase in ownership reflects the desires of a growing population within a country to become mobile and gain access to more economic activities as levels of real income increase and the economy develops.

Car manufacturers and related industries are concerned with future sales of manufactured vehicles, to acquire knowledge on the size of the future market and to predict factors driving the market size. The standard technique used by this group is for each company to use econometric and statistical procedures to forecast the future demand for new cars in the country of interest. Following this, the manufacturers will determine their market share within the future market for new vehicles.

Service industries selling fuel, tyres and mechanics are an interest group to vehicle fleet growth. These industries will secure an increase in market size with an increase in vehicle fleet, therefore, vehicle forecasts are an important indicator for potential fuel increase import numbers. The government and state companies, such as SANRAL and the provincial departments of transport are another interest group in such research. The interest is first in the

income from ownership and use of vehicles (fuel, RAF, tax, VAT, etc.) and the need to plan for future infrastructure requirements.

Vehicle ownership models have been around since the 1930s and several international authors such as Wolff (1938), Rudd (1951) and Tanner (1958), to name a few have explored the topic. Since the 1980s there was a huge concern about continuously growing traffic congestion and fuel efficiency. This led to an accelerated interest in vehicle forecast and ownership. Various forecasting techniques and data sources were explored resulting in the development of new methodologies.

There are two types of models introduced in the study, namely; a vehicle forecast model which examines the future vehicle fleet of South Africa, to predict the number of vehicles in 5, 10 or even 20 years. The second model the study introduces is a vehicle ownership model, this model predicts what factors influence ownership of vehicles in a household.

Vehicle ownership and use evolved with the growth of the economy. For example, the more real income workers receive, the more likely they are to own private vehicles. Several other factors also influence vehicle ownership, such as urbanisation, household size, level of education, etc. As economies develop, it is therefore clear that the forecast of vehicle fleets will be influenced by trends in these variables. This pattern is observed not only in South Africa but also in the rest of the world. Since the growth in vehicle ownership goes hand in hand with rapid urbanisation, the strains are severe. The increase in vehicle fleets puts a strain on vehicle maintenance facilities and the administrative structures and infrastructure that regulate the road system.

In South Africa, few studies of vehicle ownership have been undertaken over the last three decades. Some of the noteworthy work includes research by A.P. Marks and R.J Brown (1979, cited in Marks & Brown, 1981), Mokonyama and Venter (2007), Venter and Mohammed (2013) and Letshwiti, Stanway, and Mokonyama (2003). A.P. Marks and R.J Brown (1979) research introduces the power growth forecasting model and examines vehicle population growth according to racial groups. Most of these studies were all undertaken before the 1994 election and the introduction of the new constitution. It can be assumed that the democratic election brought about significant socio-economic and spatial transformation. The impact of these developments on mode ownerships and the vehicle fleet has not yet been quantified.

1.2 Significance of the study

The study investigates the development of a model to help forecast the extent of the vehicle fleet in South Africa. This current study will contribute to the existing knowledge of vehicle ownership, and the future vehicle fleet, in South Africa.

The study contributes to knowledge concerned with factors that influence household vehicle ownership. This is done by examining social-economic demographics and spatial factors influencing household decision of owning a vehicle. The second part of the study, provides an indication of the future number of vehicles, with assumption that the past (1986-2017) will continue. Therefore continuing with the trend, taking into consideration the past trends. Finally, the combination of these results provide a significant contribution to knowledge of vehicle forecast and ownership probability.

It is critical to develop a vehicle ownership model well suited for the South African economy, as it is important to transport planning in South Africa and aids in preventing under providing or over providing for the number of vehicles on the country's roads in terms of accurate and necessary road infrastructure. The model outputs will help in determining how much investment is needed for the vehicles that will be predicted to be on the roads.

Developing a model for vehicle forecast in South Africa will assist with transport planning, policymaking and the manufacturing of vehicles in South Africa. With no existing model in South Africa, this study will benefit the government in great lengths by helping them to prepare for the future number of vehicles in terms of road infrastructure, road space and policies and regulations involving road vehicles.

1.3 Aim and objectives

This research aims to develop an aggregate vehicle ownership and forecast model for South Africa using nationally collected data. The Multinomial Logistics Regression function form will be used for vehicle ownership.

The objectives of the vehicle ownership model is to determine the main factors that influence vehicle ownership in South Africa, including socio-economic, demographic and spatial factors. The focus will be on the strength and direction of influence. The second objective will be to project the future vehicle fleet for South Africa, using past vehicle fleet data. Lastly, to evaluate the growth over the next 20 years (2018-2038) and determine the extent of the South African vehicle fleet in 2038.

1.4 Structure of the study

Chapter 2 presents the literature review. This chapter defines vehicle ownership forecasting and what the forecast entails. An elaboration is given on vehicle forecasting models and techniques that are available and used in the study. A review of various factors influencing vehicle ownership is also presented in this chapter. Finally, a review of some literature available on vehicle ownership forecasting in South Africa and internationally is provided.

Chapter 3 discusses the data used in the ownership and forecasting process of the study and offers graphical representations of the historical patterns of the South Africa vehicle population. It also provides some descriptive statistics of the household percentage shares of each variable affecting the number of vehicles owned by households.

Chapter 4 describes the methodology applied in each of the models. This chapter details the research design and sampling processes for vehicle forecasting and vehicle ownership. It briefly introduces the forecasting model implemented in the study.

A more expanded explanation of the model is presented in Chapter 5. This chapter describes all models and methods implemented in the study, giving a detailed description of the process of each model or method. Vehicle forecasting focuses on historical data, which is given for a specified period using Holt's method, forecast sheet, and SPSS forecast. Vehicle ownership determines the factors influencing the number of vehicles owned within a household making use of categorical data, and the Multinomial Logistics Regression is detailed.

In Chapter 6, the study results are analysed and discussed and a brief remark on the implications of these results on South African infrastructure is given.

Chapter 7 concludes the paper by evaluating whether the objectives of the study have been achieved. The chapter reflects on the research questions to determine whether they have been answered. The closing chapter makes recommendations for further studies that are needed to understand the demands for household vehicle ownership using the household travel survey.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this section, a review of the existing body of literature is presented to broaden knowledge regarding the research topic. This section investigates existing knowledge relating to the topic and research findings. The literature review contributes to the process of identifying a gap in knowledge of vehicle ownership forecasting, specifically for South Africa, using methods and models applied in other countries and improving on identified failures.

2.2 Vehicle ownership forecasting models or techniques

Vehicle forecasting models and techniques have been investigated from as early as the 1980s, with many models having been developed in the United Kingdom (Ingram & Liu, 1997). The earliest vehicle-ownership model was developed to analyse household purchase behaviour from the forecasted household vehicle holdings.

Different models and techniques for forecasting vehicle ownership are developed for many reasons in different circumstances, depending on what purpose the forecast needs to serve. Forecasting vehicle ownership is performed for different outcomes and thus different models exist to attend to various purposes.

Models can be developed for the following purposes by different industries:

Vehicle manufacturers may apply models to predict future vehicle demand by consumers. This will assist manufacturers in producing vehicles to match the consumer demand, thus, preventing unnecessary loss of profit.

Oil companies and their products are necessary for vehicles to continue to operate, therefore these companies need to know the predicted future demand for vehicles to know how much of their product will be needed. With the latest introduction of electric vehicles into the market, oil companies might be affected drastically when the demand for fuel decreases as the market for electric vehicles increases. The market for electric vehicles will have an impact on fuel stations, as there will be a minimised need for them.

International organisation (World Bank) uses specific forecasting models for vehicle ownership in a country to analyse and assess the country's investment legibility. These models assist the organisations in deciding whether to invest in a country.

National governments find vehicle ownership forecasting models useful in making forecasts for tax revenues and the governing influences of changes in taxation levels. Governments, including local and regional authorities, furthermore make use of these models to forecast

transport demands, energy consumption and CO₂ emission levels, and the policy impacts of the forecasted vehicle numbers.

The research restricts focus on vehicle ownership models for the investigation of road infrastructure investments. Some of these models are of interest to car manufacturers or oil companies. Nevertheless, models developed for private and public firms are different in their exogenous versus endogenous variables.

Models focusing on investment and transport planning will be reviewed and discussed in this section of the research.

2.3 Models of vehicle ownership

As stated before, it is crucial to know that different vehicle ownership models exist to serve a diversity of purposes. Below a review of a few models is given. Most of these models are still in use in current years. Due to the extensive variety of vehicle ownership models, this section will present only a few of these models.

De Jong, Fox, Daly, Pieters, and Smit (2004) presented the following various models in their research. Their research focused on models developed for public sector planning, such as manufacturers and oil companies. The requirements for these public-sector models differ from those of the private sector.

2.3.1 *Aggregate Time Series Models*

Aggregate time series models are defined by a sigmoid-shape function for the development of vehicle ownership over a certain period, which at the beginning displays a slow increase and thereafter begins to increase abruptly and ends off at a capacity level known as the “saturation level”.

Numerous examples of this model in operation are seen in the period between the late 1950s to the early 1980s in literature focusing on the United Kingdom. Button, Ngoe, and Hine (1993) make use of aggregate time series in research by applying a logistic function in vehicle ownership forecasts. Ingram and Liu (1997) make significant use of this model by applying a double logarithmic specification in the explanation of vehicle and vehicle ownership (in their paper the use of car and vehicle are defined clearly). An aggregate model in National Road Traffic Forecasts is used in the United Kingdom by the application of a logistic curve for saturation, and this was stretched by including saturation levels (focusing on household types) to the disaggregate tree logit calibration.

Developing countries are attracted to these Aggregated Time Series Models due to their requirement for the lowest data, while income is usually considered as the main driving factor

behind vehicle ownership growth. However, Gakenheimer (1999) had a few remarks in the article 'Urban mobility in the developing world', stating that with low-income developing countries having the income of the top 20% of the population as the explanatory variable might be better than overall income. Gakenheimer (1999) added a second remark stating that a quadratic function outperformed a sigmoidal curve.

Romilly, Song, and Liu (1998) had an opposing opinion to that of Gakenheimer regarding the saturation curve method by estimating a time series model without assuming saturation levels.

2.3.2 Aggregate Cohort Models

Aggregate cohort models are popularly used when segmenting the population into groups of for example matching birth years, then used to predict the future of these groups in describing how growth in age within these groups influences the groups in acquiring, keeping and disposing of vehicles.

A popular reason for the increase in vehicle ownership within countries is amongst other things the demographics within the country of study. De Jong et al. (2004) call this the "cohort effect". The generation born earlier than the Second World War grew up in times when the vehicle ownership lifestyle was not yet established, and today this generation is still seen to have a low vehicle ownership rate and low interest in owning vehicles.

The generation thereafter has grown up in the car era and it is evident as this generation has accumulated more vehicles now and is expected to increase the number of vehicles owned as far and long as the economy allows. Aggregated Cohort Models' growth in vehicle ownership is driven by the demographic force factor (De Jong et al., 2004).

In a paper authored by Fridstrøm, Østli and Johansen (2016) the stock-flow cohort model is introduced to keep track of how fast technological developments and other changes in the qualities of new vehicles entering the vehicle fleet. This model depends on available records from the government.

Within the Stock-Flow Cohort Model (Fridstrøm, Østli, & Johansen, 2016) car stock is presented in segments and age, using 2012 as the base year of the research. Car stock is split into 22 segments and 31 age classes; there are nine segments for petrol-driven cars and nine for diesel-driven cars, and each vehicle class is then subdivided into weight classes. Segmentation within this model is based only on objective criteria.

2.3.3 Pseudo-panel methods

A pseudo-panel is an articulated panel based on averages of repeated cross-sections. Restrictions are imposed on the data of a pseudo-panel before it can be treated as actual panel

data. The most vital restriction is that cohorts should be based on time-invariant characteristics of the households, for example, the birth year of the breadwinner of the household. By defining the cohorts, homogeneity is pursued within the cohorts and heterogeneity between the cohorts.

The most critical feature of the pseudo-panel data is that averaging over cohorts' changes disaggregate (discrete) values of variables into cohort means, thus resulting in information on individuals being lost (De Jong et al., 2004).

Table 2.1: Comparison of types of vehicle ownership models

Aspect	Aggregate time series model	Cohort models	Aggregate market models	Heuristic simulation models	Static disaggregate ownership models	Indirect utility models	Static disaggregate type choice models	Panel models	Pseudo panel models	Dynamic transaction models
Demand-supply	Usually only demand	Demand	Demand and supply of 2 nd hand cars;	Demand and supply of 2 nd hand cars;	Demand	Demand	Demand	Demand	Demand	Demand
			Equilibrium mechanism	Equilibrium mechanism						
Level of aggregation	Aggregate	Aggregate	Aggregate	Disaggregate	Disaggregate	Disaggregate	Disaggregate	Disaggregate	Aggregate	Disaggregate
Dynamic or static model	Dynamic	Dynamic	Dynamic	Static, but shift from new to old cars over time	Static	Static	Static	Dynamic	Dynamic	Dynamic
Long or short-run forecasts	Short, medium and long (saturation)	Medium and long	Short, medium and long	Medium and long	Long	Long	Long	Short and long	Short and long	Short & medium
Theory	No strong links	No strong links	Economic market equilibrium theory	Strong basic assumptions, can be at odds with theory	Can be based on random utility theory	Strong links	Can be based on random utility theory	Can be based on random utility or lifetime utility theory	Weak links with random utility theory	Parts can be based on random utility
Car use	Not included	Not included	Not included	Can be included, but insensitive (can be amended)	Included in some models (logsum)	Included	Included in some models (logsum)	Sometimes included, but in ad hoc way	Not included, but can be	Sometimes included, but in ad hoc way
Data requirements	Light	Light	Light	Moderate	Moderate	Heavy	Heavy	Very heavy	Moderate	Very heavy

Special treatment of business cars	Usually not, but possible	Usually not, but possible	Usually not, but possible	Usually	Done in recent models	Usually not, but possible	Usually not, but possible	Usually not, but possible	Usually not, but possible	Usually not, but possible
Car types	No car types	No car types	Limited number of car types	Limited number of car types	Very limited number	Very limited number of car types possible	Often very many car types (brand model-age)	Very limited number of car types possible, but could be combined with a type choice model	A very limited number of car types possible	Very limited number in duration model, but very many in-car type choice model

Impact of income	Yes	Yes	Yes (average and distribution)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Impact of car cost	Fixed and or variable cost sometimes included	None	Fixed and variable	Fixed and variable	Fixed cost often included; logsum includes variable cost	Fixed and variable (also on car use)	Purchase cost and fuel efficiency often included	No policy runs reported, but might be possible	Fixed and variable	Fixed and variable
Car quality impacts	No	No	No	Can be included, might have to work through cost	No	No	Yes	No, unless type choice added	No	Yes in type choice
Impact of licence holding	No	Yes	Yes	No	Possible	Possible	No	No, but possible	No, but possible	No, but possible
Sociodemographic impacts	Limited	Many possible	Limited	Many possible	Many possible	Many possible	Many possible	Many possible	Limited	Many possible
Attitudinal variables	Hard to include	Cohort-specific attitudes can be included	Hard to include	Hard to include	Can be included if specific questions in dataset	Hard to include	Can be included if specific questions in dataset	Can be included if specific questions in dataset	Can be included if specific questions in dataset	Can be included if specific questions in dataset
Scrappage included	No	No	Can be included	Can be included	No	No	No	Can be included	No	Can be included

Source: (De Jong et al., 2004)

2.4 Factors influencing vehicle ownership

It is necessary to explore the topic of factors influencing vehicle ownership, as these factors have an impact on the critical decision individuals make on whether or not to purchase a car. In this section, the literature on these factors is explored and these factors are considered by focusing on their effect on vehicle ownership. Many of these factors influence vehicle ownership levels at both local and national levels. The aim is not to review all possible influences, but rather the main variables focused on in the literature.

Every individual and household has a budget of the amount that is set aside for transport, and this imposes a constraint on the choice between vehicle ownership and vehicle use (Beca Carter Hollings & Ferner Ltd, 2003). It is without saying that a small family, with high expenses but reliant on a private vehicle, tends to economise on vehicle ownership by choosing to purchase a rather cheaper vehicle that will get them from point A to point B. The dynamics change with more disposable income and fewer dependents.

Some of the predominant influences on vehicle ownership include:

Button, Pearman, and Fowkes (1982) argued that income is the core determinant of strong vehicle ownership amongst consumers. Most econometric studies using both time-series and cross-sectional data have incorporated an income variable, and most of the studies place importance on establishing income elasticity of demand for vehicle ownership.

However, an attractive income is the main variable. Button et al. (1982) state in their research that allowance must be made for the varying techniques that may be employed in forecasting vehicle ownership. Several periods and areas are focused on by different research papers, and allowance must be made for the additional variables included in the models developed. Other studies such as that conducted by Fishwick and cited in Button et al (1982) have failed to find a noteworthy relationship between vehicle ownership and income.

Mokonyama and Venter (2007) agree with Button et al (1982) in finding that income is the most significant explanatory variable in household vehicle ownership models. The consumption of vehicle travel is a 'luxury good', where the increase in income makes the benefits of private vehicle travel attractive.

The relationship between income and vehicle ownership levels is validated by research conducted by Button, Ngoe, and Hine (1993), who analysed vehicle ownership in low-income countries. Their results show that among the more prosperous individuals in these countries an increase in vehicle ownership can be observed as a result of their prosperity.

Households have long been recognised as a key determinant of travel choices and having a strong influence on vehicle ownership levels and patterns. Households can be divided into

distinct categories according to characteristics such as the size of household, age, social status, gender distribution, and employed members of the household. Heggie (1979) gives a complete concept of the important nature of households in the modelling of vehicle ownership. This is done by focusing first on the household's needs in different life cycle stages and then linking this back to vehicle ownership.

Vehicle ownership tends to be highest in low-serviced suburbs and densely populated areas, and lowest in central business districts (CBDs) and surrounding areas. This pattern can be explained by the distance from the household to important services (Heggie, 1979). South Africa is a notable example of this pattern, where the aftermath of apartheid is displayed in the low levels of retail and services within black townships or suburbs. This increases the need amongst the township dwellers to own private transport, therefore increasing vehicle ownership.

Bjoner and Leth-Petersen (2004) reported in research titled "A disaggregate model of household vehicle ownership" (Cundill, 1986) that a household with married people had a tendency to increase vehicle ownership, since it appeared that separated couples who then become single have higher levels of vehicle ownership compared to singles who never got married.

Fuel is a variable that is a complement to trip choice, and differences in the price of fuel may, therefore, be assumed to influence vehicle ownership decisions. Fuel prices tend to display small environmental variations in a region or a country where it seems doubtful that fuel prices would prove significant in cross-sectional studies of vehicle ownership (Button et al., 1982).

Over time the impact of fuel price changes on the total vehicle stock is also arguable. A long-term increase in the price of fuel will reduce the demand for vehicles, but this decrease will not have a significant impact on the overall vehicle ownership. A higher fuel price will perhaps impact the type of vehicle that consumers will opt to purchase rather than the overall number.

Essentially, the introduction of fuel-efficient vehicles will become a popular option amongst most individuals, this at the cost of less fuel-economic types.

Dargay and Gately's (1999) paper on vehicle ownership reports on running costs as a variable influencing ownership, referring to fuel prices. In this paper, there is a support of the concept by Button et al. (1982) that increases in the fuel price influence vehicle ownership as far as fuel consumption is concerned.

Accessibility is included in vehicle ownership modelling in an attempt to display more directly the overall advantages that vehicle ownership adds to a household's wellbeing (Button et al., 1982). Accessibility does this by attempting to show the additional opportunities created and

becoming available with the ownership of a vehicle, while other models include the accessibility that vehicle ownership provides to alternative transport modes.

However, Gould (1969) argues that accessibility is a slippery concept, a common term used but not properly defined and bringing the problem of how to measure it. Care needs to be applied when using such a concept. Accessibility is relevant in vehicle ownership in terms of the need for private transport travelling and the quality of public transport services offered. The ability to make a trip allows for the full range of social and employment opportunities to be enjoyed.

The measures of accessibility may be in the form of replacement variables or indices that try to quantify the level of access offered to the household by owning a vehicle. The most common replacement/surrogate variables are population or suburban density of an area (Gould, 1969).

Mokonyama and Venter (2007) mention road density, population density and level of service as other variables that influence vehicle ownership. Investment in road infrastructure is often connected to an increase in household vehicle ownership; however, Mokonyama and Venter (2007) discovered that there is a range of road densities for a given level of vehicle ownership, particularly in high-income countries (Mokonyama & Venter, 2007).

Road investment has a strong relation with economic growth and not vehicle ownership per se (Ingram & Liu, 1997). Mokonyama and Venter (2007) state that further road infrastructure investments in South Africa are not expected to have much of an effect on the level of vehicle ownership.

Often in densely populated areas with a good working public transport system or service, the vehicle ownership ratio is lower than anticipated given the high per capita income in certain areas. The inverse relationship between vehicle ownership and population density is connected to shorter distances travelled between CBDs. However, vehicle ownership will impact the demand for public transport. It's not always that a community with a good working public transport, choose public transport over private vehicle ownership. Therefore, the relationship is not always inverse.

2.5 Review of South African literature on vehicle ownership forecasting

It is particularly important in South Africa to consider population groups when conducting any form of forecasting or predicting. This is because of the socio-economic similarity that characterises each population group. Another important consideration would be the availability and quality of applicable data on which forecasting procedures are based.

In their paper on vehicle ownership in South Africa, Marks and Brown (Marks & Brown, 1981) focused on reviewing vehicle ownership models in the local context and introduced the power

growth, forecasting model. They analysed vehicle ownership trends in South Africa, both nationally and regionally (focusing on the Pretoria-Witwatersrand-Vereeniging area).

The power growth forecasting model introduced in Marks and Brown's paper is an improved time series formulation in which incorporation of power functions is applied. This approach helps to overcome issues of externally determining the essential "saturation-level" constant by simultaneously determining this saturation level and other constants from past trends in the growth rates.

Marks and Brown (Marks & Brown, 1981) used the power growth model as opposed to well-known and well-utilised models such as the logistic time series model to forecast vehicle ownership in South Africa, though it is suggested to use both models. In the research literature, it is assumed the model is applied to time series data. Observations of past trends over the long term showed that the relationship between the percentage growth rate of vehicle ownership is not linear but rather curved, therefore, the application of the power growth model is clearer when compared to the logistic function.

The power growth model has the following advantages related to vehicle ownership forecasting (Marks & Brown, 1981):

- The growth curve convexity is inevitably accounted for with the use of a small integral.
- The saturation level is calculated by model calibration of existing time-series data as to the assumed value of saturation level as an exogenous input into the forecasting model.
- The model is flexible; it allows the forecaster to examine the interrelation effects of changes in the values of the constant integral and saturation level.

Disadvantages of the power growth model:

- Computer applications required for curve fittings.
- Model sensitivity high because errors in the saturation level and ownership level would be magnified by the power of the integral.

According to Marks and Brown (1981), the application of the power growth model to time series data is a major improvement on existing methods used for forecasting vehicle ownership.

The analysis of past vehicle ownership trends in South Africa is an important input in a time series-based forecast; however, the availability and quality of this historical data is a problem.

In their analysis of the trends in South African vehicle ownership, Marks and Brown (1981) made use of data such as official national motor vehicle licensing statistics over the period of analysis (1920-1977) and made available by the Department of Statistics; Pretoria-

Witwatersrand-Vereeniging home-based interview data; regional statistics for the same period as the national statistics; and statistics from censuses at both national and regional level.

In 2007, Mokonyama and Venter (Mokonyama & Venter, 2007) began their investigations into vehicle ownership modelling, with their research focusing more on the forecasting of household vehicle ownership in South Africa and looking specifically at alternative models to those existing ones, and making observations about future trends in vehicle ownership models.

The household vehicle ownership modelling approach explored is not dependent on the classification of the population into racial groups, but uses household income and spatial characteristics of the area of study, captured in terms of dwelling unit type. However, there may still be frequent social inequalities amongst population groups but systematic models of household vehicle ownership need not depend on racial groupings to produce an effective forecast.

Previous South African vehicle ownership modelling research distinguished between different population groups intending to reduce data collection errors that resulted from opposing economic profiles of the population groups (Mokonyama & Venter, 2007). By apartheid planning policy, racial differences made it easier to model average trip generation rates in areas housing different racial groups. However, owing to the growing spatial and economic integration brought about by the democratisation of the country, racial differentiation in transport modelling is difficult and fast becoming irrelevant.

In 2003 vehicle ownership in South Africa was 6 245 392 (RTMC, 2018); 74% of all households did not own vehicles (Mokonyama & Venter, 2007). Vehicle ownership in South Africa has been accelerating since then, but this is more in the metropolitan areas where congestion and environmental costs are evidence of increasing vehicle ownership (RTMC, 2018).

In their research, Mokonyama and Venter (Mokonyama & Venter, 2007) used a different model to the one used by Marks and Brown (1981) in forecasting vehicle ownership in South Africa. Mokonyama and Venter (Mokonyama & Venter, 2007) focused on forecasting for the metropolitan urban areas and used the household vehicle ownership model to forecast vehicle ownership. The primary reason for using this model is that this model does not rely on the classification of the population into race groups, but rather depends on market profit around the area of interest.

Household income and spatial characteristics of the area of interest are captured in categories of dwelling unit types. Mokonyama and Venter (Mokonyama & Venter, 2007) specifically used this model in the city of Johannesburg and further tested it around the Gauteng area.

In their research, Mokonyama and Venter (Mokonyama & Venter, 2007) proposed an aggregated cross-sectional model based on category analysis. The model developed tested variables such as dwelling unit type (classified as flats, houses, townhouses and 'other') including informal dwellings occupied by households between quintiles 1 and 4, shacks and leased rooms, household income and the location of the relative households within the area of research.

The use of dwelling type as a variable in research is as a result of the theory that a household's dwelling asset, whether owned or rented, conveys information concerning lifestyle choices, which can be associated with vehicle ownership. The dwelling type is further linked to residential density, which can affect the density of vehicle ownership.

The model used by Mokonyama and Venter (Mokonyama & Venter, 2007) utilised data from the 2002 Gauteng Household Travel Survey, Census 2001 by Statistics SA and land use data divided into City of Johannesburg transport zones. The transport zones were then subdivided into income groups, mainly low (< R1 999), medium (R2 000-R6 999) and high (R7 000+). Each zone was then classified into an income group according to the household income category fit.

An evaluation of the dwelling location variable was conducted by making an additional distinction according to the households residing within the N1-N3-N12 Johannesburg freeway ring road and households situated outside the ring road. The ring road was used as the functional boundary because of its historic role in bypassing what was viewed as the centre of Johannesburg, however, the calibration of the location concerning the ring road was irrelevant; statistically, there was no difference between household vehicle ownership within the ring and outside the ring.

This model was then, in the same paper by Mokonyama and Venter (Mokonyama & Venter, 2007) tested for transferability using the City of Tshwane as the study area, and making use of the same variables as those used for the City of Johannesburg. In aggregate, the model was found to perform nearly as well in Tshwane as in Johannesburg. The model predicted 82% (Mokonyama & Venter, 2007) of the actual vehicle fleet in Tshwane, the difference explained by the number of vehicles owned by non-household formations and non-residents.

Venter and Mohammad (Venter & Mohammed, 2013) published an article on estimating vehicle ownership. Their research explored a disaggregate vehicle ownership model and connected it to a model of household transport energy consumption to consider the unrevealed socio-economic and land-use variables steering energy consumption.

The development of a vehicle ownership model in this research focused on a metropolitan area, namely Nelson Mandela Bay. The data used in the research was pulled from the 2004

Nelson Mandela Metropolitan Area travel survey, provided by transport supply data. The research was then limited to private transport modes and this excluded freight and commercial transportation. The focus of this research was furthermore the end user's energy consumption, analysing the minimal amount of fuel in the case of the utilisation of road transport, or the electricity consumed per trip when a traveller was making use of rail transport.

Models of vehicle ownership are essential to the examination of transport energy consumption. Venter and Mohammad (Venter & Mohammed, 2013) estimated a household vehicle choice model for investigating the factors leading to decisions to purchase a vehicle and provided a variable of predicted vehicle ownership potentially used in energy models.

A multinomial logit model (MNL) was introduced in the research, as the model for predicting vehicle ownership. This model uses category-dependent variables and has three potential outcomes consisting of households with zero cars, one car, or two and more cars. This model involves explanatory variables such as monthly household income stated by a survey respondent, number of people working in the household and the size of the household, which was compared with income to analyse the possibilities of household size influencing vehicle ownership reliant on socio-economic standings.

In the research, it was found that amongst the explanatory variables used the correlations were below 0.5. This demonstrates that there is sufficient independence between the variables. Furthermore, Mohammad and Venter (2013) included land-use descriptors such as zonal population density and job accessibility.

Letshwiti, Stanway, and Mokonyama (Letshwiti et al., 2003) researched CSIR Transport assessing vehicle ownership trends and use within South Africa. This research explores the logistic curve theory as a method of forecasting vehicle ownership trends. The analysis is centred on information adopted from the National Transport Department database and the electronic National Traffic Information System (eNaTIS). Additionally, more data was obtained from the National Department of Minerals and Energy Affairs, Statistics South Africa and metropolitan vehicle registering authorities.

Letshwiti et al. (Letshwiti et al., 2003) adopt the logistic curve theory on assumptions that the ownership of a vehicle tends to follow an S-shaped curve as time passes. In this logistic curve theory, the curve asymptotes to some saturation value to represent the point of stability in the market of vehicles where the replacement rate is equal to the disposal rate. The quality of the forecast using this theory depends predominantly on the quality of the time series data utilised in the analysis.

In the research, Letshwiti and fellow researchers (2003) decrease the effects of factors affecting the quality of time series data by presenting the forecasts in terms of several vehicle

ownership saturation values, and from these values then interpolations of future vehicle ownership are made.

The logistic curve used in the research is represented by the following equation (Letshwiti et al., 2003):

$$C_t = \frac{s}{1 + b \exp(-aSt)}$$

Where: C_t = vehicle ownership at the given time t (cars/1000population)

b = constant from the derivation of the equation

a = constant from derivation of equation

S = saturation level (cars/1000 population)

t = time

The saturation value in the equation is attained by the assumption of a linear relationship between the percentage change in vehicle ownership and vehicle ownership. At the point where the percentage change in vehicle ownership is zero, the value of saturation links with the level of vehicle ownership. To obtain the constants 'a' and 'b' multiple regression can be used, or the constants can be determined by using limit conditions at time=t. The research makes use of regression to get the values of constant 'a' and 'b'.

The paper uses 1940 as the base year in the projection of vehicle ownership trends in South Africa.

2.6 Review of International literature on vehicle ownership forecasting

Unlike literature in South Africa, international literature on the topic of vehicle ownership is abundantly available. In this section of the literature review, international literature on vehicle ownership modelling is reviewed. This section intends to investigate what models and variables have been used internationally and how this can be adopted into research and better the model developed in the research.

Cundill (1986) conducted a study on vehicle ownership in Kenya. The focus of the study was to examine a few policies for conserving fuel within the transport sector, and part of the study involved the surveying of ownership and the use of private vehicles. The surveys aimed to provide background information that assisted in evaluating the fuel conservation policies involving vehicles.

Kenya is a developing country and thus many of the privately-owned vehicles are owned by individuals working within urban areas. Concerning this, surveys in the study were conducted

in four major urban parts of the country, consisting of Nairobi, Mombasa, Nakuru, and Kisumu. The surveys covered two parts, the first part being an interview survey which covered questions about the respondent, total household income and cars owned in the households, and the second part being a journey log survey that required car owners to keep track of vehicle usage over a given period. Both parts of the survey covered car-owning and non-car-owning households.

The analysis of vehicle ownership in the paper was concerned mainly with the relationship between vehicle ownership and income in Kenya. This followed the approach implemented by Bates, Gunn, and Roberts (Cundill, 1986) in the United Kingdom's Regional Highway Traffic model.

Cundill (Cundill, 1986) presents the quasi-logistics function as an analysis tool for vehicle ownership in Kenya, which was adopted from Bates, Gunn, and Roberts (Cundill, 1986) as stated above.

The quasi-logistic function is expressed as follows:

$$P = \frac{S}{1 + e^{-a|b}}$$

In the study it was found that the probability (P) of a household with very high income owning one or more cars can be linked to the household income (I) expressed by the quasi-logistics function above where S is the saturation level (the probability of a household owning one or more cars), while a and b are constants.

If assumed S is unity, the quasi-logistics function can be rewritten as:

$$\frac{P}{1 - P} = e^{a|b}$$

Or in logarithmic form:

$$\ln\left(\frac{P}{1 - P}\right) = a + b \ln(I)$$

The level of income known as the 'equi-probability income' | and equals $e^{-a/b}$, is the point where the probability of vehicle ownership equals 0.5. Graphically this point is where income level at the logarithmic form of the equation intercepts the Ln (I). Using |' this is written as:

$$P = \frac{1}{1 + \left(\frac{1}{r}\right)^{-b}}$$

To test the quasi-logistic function in Kenya, Cundill (1986) made use of data collected from all survey websites that were pooled in the research. Each income category plotted using the logarithmic form of the quasi-logistic function, is displayed in Figure 2.1 below. In the figure, error bars are displayed corresponding to \pm two standard errors in P, limited by the 95% probability.

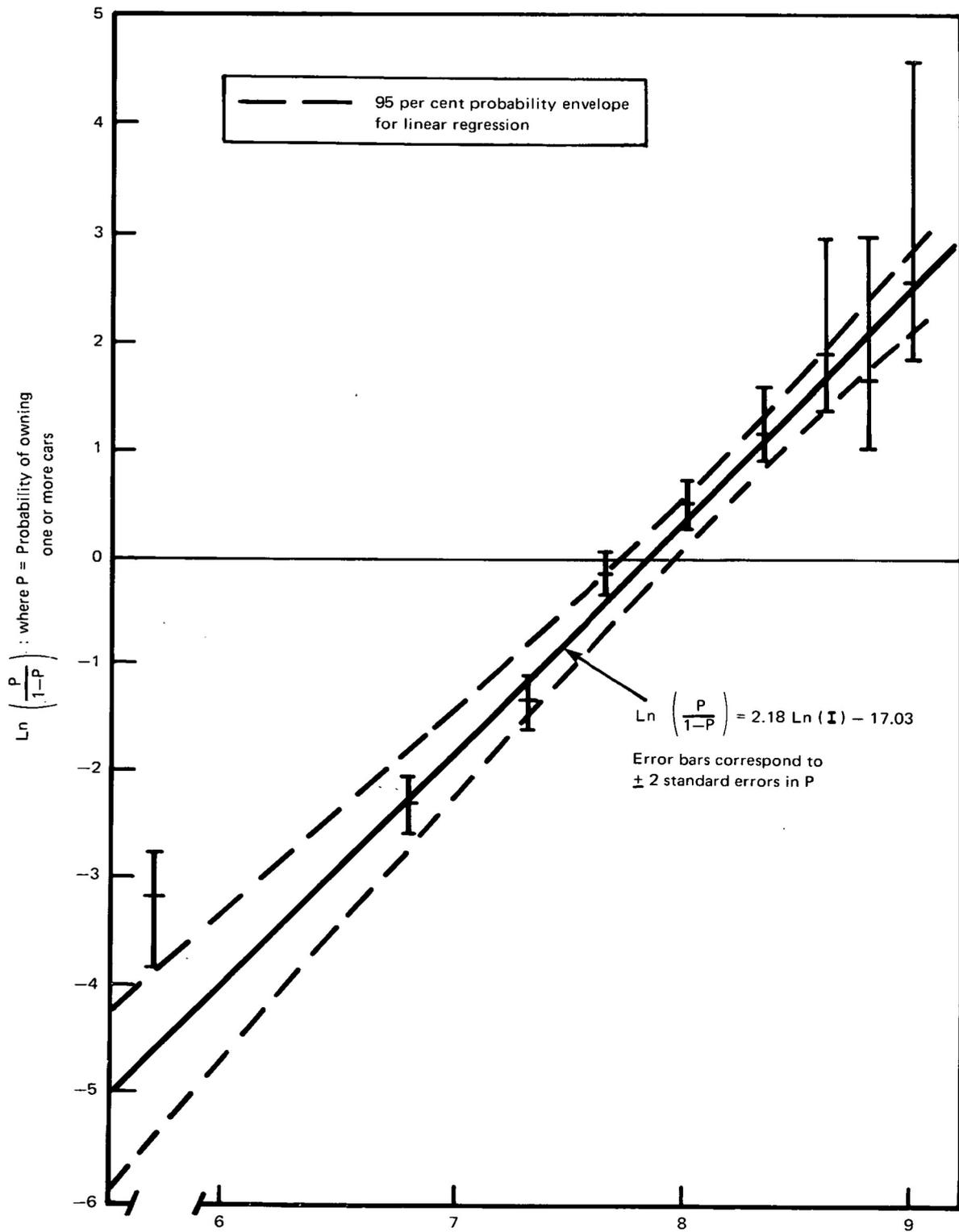


Figure 2.1: Quasi-logistic relationship for vehicle ownership.

Source: Cundill (1986)

Although P is binomially distributed in the research, a normal distribution was used to estimate it with a standard error 0, given by:

$$\sigma = \frac{P(1 + P)}{n}$$

Where n = number of interviews in the observed income category.

When the P-value of a given income category is close to zero or close to unity, it is prone to error resulting from misspecification of vehicle ownership or income. The research concluded by stating that its unfortunate comparison of results from the four urban areas was limited by statistical constraints arising from the decreased sample sizes.

Furthermore, it is not favourable that the vehicle ownership relationship with income can be used as a foundation of a model for dependable forecasting of future vehicle ownership or fuel use. This is because of the instability of overtime and it would need supporting data, specifically on income distribution, which is not always available.

Whelan (2007) focused on the modelling of vehicle ownership within Great Britain. In the research, it is noted that a range of vehicle ownership forecasting methodologies were used in the United Kingdom by the department of transport. These methods were used in assistance to the production of the national road traffic forecasts (NRTF) needed in supporting road scheme design and evaluation.

The NRTF was based on extrapolation techniques developed by the Transport and Road Research Laboratory. Models were aggregate, predicting ownership nationally and regionally making use of time series data, cross-sectional or pooled data. S-shaped growth curves market saturation and incorporate income, index of motor cost and time were used as descriptive variables in the presented models.

Whelan adopted the NRTF in his research paper, with improvements to its vehicle ownership forecasting methodology. The improvements were aimed at justifying increases in multi-vehicle households, assist in assessing the impact of company cars on ownership levels, identify market saturation and the impact of employment levels on the ownership of vehicles and introduce sensitivity to purchase and use costs within the model.

Data was extracted from the family expenditure survey (FES) and the national travel survey (NTS), with FES data covering the period 1971-1991 and NTS data covering 1991. Whelan identified the following explanatory variables in the model: household income, household structure, motoring costs, accessibility, the provision of company vehicles and license holding. De Jong et al. (2004) states that these variables are commonly used when developing a vehicle ownership model.

Household choices are further categorised and include households with zero, one, two or more vehicles. The model assesses household decisions to own vehicles in the stated categories by three binary digit models. The binary digit models are as follows:

First:

$$P_{1+} = \frac{S_{1ah}}{[1 + \exp(-U_{1+})]}$$

Second:

$$P_{2+|1+} = \frac{S_{2ah}}{[1 + \exp(-U_{2+|1+})]}$$

Third:

$$P_{3+|2+} = \frac{S_{3ah}}{[1 + \exp(-U_{3+|2+})]}$$

Where S is the saturation level by area (a) and household (h) and U is the utility of ownership.

$$U_{1+} = ASC_1 + b_1LPA + (c_1 + c_{h1}D_h + c_{a1}D_a)Y + d_1E + e_1O + f_1R$$

$$U_{2+|1+} = ASC_2 + b_2LPA + (c_2 + c_{h2}D_h + c_{a2}D_a)Y + d_2E + e_2O + f_2R + g_{21}CC_1$$

$$U_{3+|2+} = ASC_3 + b_3LPA + (c_3 + c_{h3}D_h + c_{a3}D_a)Y + d_3E + e_3O + f_3R + g_{31}CC_1 + g_{32}CC_2$$

Where: LPA = number of driving license per adult

Y = household income

R = index of vehicle use costs

D_h = vector of household type dummy variables

O = index of purchase costs

D_a = vector of area type dummy variables

ASC = vector of alternative specific constants

E = number of adults employed

b, c, d, e, f, g = parameter vectors to estimate

CC_1 = dummy variable if there is one company car in the household

CC_2 = dummy variable if there are two company cars in a household

Whelan defined the forecasting period of research at a 5-yearly interval from 2001 to 2031, allowing for estimations of trends in ownership. The research concluded with results from the forecasting, showing an increase in the number of households, including a greater increase in the number of single adult households. The number of vehicles was forecasted to display a 42% increase from 25.63 million to 36.35 million vehicles in Britain.

Romilly, Song, and Liu (1998) researched the modelling and forecasting of vehicle ownership in Britain; however, their research developed a model of per capita vehicle ownership, this being different from that applied to Britain by Whelan (2007). The model introduced by Romily et al. (1998) produces forecasts that are more trustworthy than those of the NRTF used by Whelan in his research on Britain.

In their paper Romily et al. (1998) focus on modelling and forecasting vehicle ownership and do not consider the influences of car use or any form of road traffic. The paper studies a single model that concurrently integrates casual and proxy variables, which will provide a variety of forecasts from appropriate assumptions on time paths of the independent variables.

The methodology used in developing the model of per capita vehicle ownership is econometric, using the error correction model. Romily et al. (1998) used data from the period 1953-1994 and the following variables in the model: real personal disposable income per capita, real motoring cost index, real bus fare index, unemployment rate, size of the road network, real interest rate, population and age structure.

The model is based on the assumption that several economic and social variables determine the aggregate level of per capita vehicle ownership over a given period. The cointegration regression of the model is re-estimated using data from the period 1953-1978, 1982 and 1988 respectively. Time trend and structural break dummies are incorporated in each of the cointegration regressions, while estimations from these cointegration regressions are used as parameters for the forecasting models.

The forecasting criteria used include: Mean absolute percentage forecasting error (MAPE) and root mean square forecasting error (RMSE). The model provides forecasted results for short-term and long-run income and provides a better in-sample forecasting performance than DoT.

In the Middle East minimal research had been done in the field of vehicle ownership, up until the very first research was published by Manski (1983). This research was based on

aggregate discrete choice models to analyse vehicle ownership in Israel. Also in Kuwait research was done on vehicle ownership, using household size, income and number of adults as causal variables for the model developed. For Asian households, it was found that income was the only significant variable in developing a vehicle ownership model.

Serag (2014) focused on modelling vehicle ownership in Egypt, with data for the period 1990-2009 extracted from World Bank Statistics, thus making 1990 the base year. Three models were developed and used in the forecasting of vehicle ownership in Egypt: the Log-Linear Model, Quasi-Logistic Model, and Gompertz Curve. In each model, the gross domestic product per capita (GDPPC) was used as the user income variable expressed with the 2000-constant price in US dollars.

Within each model Serag (2014) used time and GDPPC as independent variables, with the sum-square of error terms from each model used to compare the models' accuracy of the forecast. The models that were developed indicated that vehicle ownership depends on time and GDPPC. The log-linear model overestimated vehicle ownership and resulted in the highest vehicle ownership in the forecast. The forecasts from the quasi-logistic and Gompertz models were similar and gave considerable accuracy in the forecast. However, the Gompertz model had the highest forecasting accuracy.

The approach adopted in the research employed sensitivity analysis and offered a range of projections of future growth in per capita vehicle ownership based on changing scenarios of how income may change in the future.

Öğüt (2004) introduced S-Curve models to the Middle East, and in the research, Öğüt focused on Turkey. The period of analysis was 1970-2002, 1970 is the base year of the research in terms of data collection and observations. In these S-curve models, Gross National Product per capita (GNPPC) was used as the user income variable as opposed to the GDPPC used by Serag (2014). This income variable was expressed with the 1987-constant price in Turkish Lira.

The long-term vehicle ownership forecast models included the idea of a saturation level (the level at which income no longer is a restraint). Öğüt explored three S-Curve models, one of which was mentioned in the review of Serag's (2014) paper on Egypt, namely: Logistic curve, Power Growth Curve and Gompertz Curve. Each model applied GNPPC as variables, and the sum-square of error terms (SSET) was used to compare the forecasts of the models. The Power Growth and Gompertz curve models produced similar forecasts in the research, while the Logistic curve had an overvalued vehicle ownership forecast.

The Gompertz Curve model and Power Growth Curve model increased the reliability of the forecasts in this study.

In the case of Iran, data was collected from five cities, namely Tehran, Tabriz, Shiarz, Esfahan and Mashhad to observe the ownership of vehicles. The study divided the country into 56 zones. A nested Logit Model using a 25% random sample from mother and child tracking system (MCTS) data was proposed, then a two-level model of vehicle ownership with the use of simple multinomial logit models (Shaygan, Mamdoohi, & Masoumi, 2017). In each model, household and socio-economic conditions such as the number of employed members of the household and income levels of each household influenced the desire to own more than a single-vehicle. Additionally, job type was important, while the presence of males within each household, the age of the oldest child and household size were crucial factors influencing the purchase of a vehicle.

In the city of Esfahan, average vehicle ownership per capita was calculated, vehicle ownership was estimated according to the assumption of limited imports of vehicles, and using the number of vehicles to be produced private vehicle ownership was predicted. Later, disaggregate data was converted to aggregate, and aggregate variables were used as independent variables for modelling. In the study conducted in Iran, three groups of models were applied, namely a trend base model, regression model, and discrete choice model.

Direct models were presented to indicate the number of vehicles percentage share of each household in each zone. Multinomial logistic and ordinal regression were the discrete choice models used to predict vehicle ownership. The trend-based model determined the type of vehicle within the predicted vehicle ownership profile (Shaygan et al., 2017).

Prevedouros and An's (1998) study on vehicle ownership focused on Asian countries. The study was based on per capita vehicle ownership trends in 12 countries, obtaining data for research from the United Nations' Statistic Yearbook. The vehicle registrations were divided into population figures to produce passenger vehicles per thousand of the population, and this became the dependent variable for the model. The independent variables included deflator, gross domestic product (GDP) at both current and constant prices, unemployment rate, railway passenger mileage, and roadway mileage.

This research made use of time series data. Thus, time series regression models are most suitable for data analysis. For this application, the time series regression model used was AREG found in SPSS/PG+, a method that can handle missing data that may be present in a data set. An assumption was made that GDP growth rate would remain as constant in the period 2000-2005 as from 1990-2000, as reported by the Organisation for Economic Cooperation and Development (OECD) and the United Nations. The use of these databases improved the accuracy of the forecasts.

Yang, Jia, Liu, and Yin (2017) focused mainly on China, but with China being part of Asia the study did not adopt the model used by Prevedouros and An (1998) as the study was focused on China at the city level. The data for this study were collected from the China City Statistics Yearbooks for the years 1994-2012, this period being within the period in which China witnessed rapid economic development and urbanisation in its cities, with accompanying growth in vehicle ownership. Sampling consisted of 293 cities out of 336 due to missing data.

For analytical purposes, cities were ranked by total population based on a census conducted in 2010, which provided bases for grouping cities into size quartiles: small city, medium city, a large city and super-large city. The variables in the model used contained the vehicle ownership rate as the dependent variable, representing the number of privately owned vehicles per thousand persons. Vehicle ownership rate is a good indicator of willingness to own a vehicle, and its use eliminates the estimation bias connected with the total population size.

Independent variables included: Average income (a measure of wealth), population (population size and urbanisation rate), size of built-up area, road area per capita, urban population density, improvement in the road system, density in urban development on vehicle ownership and transportation alternatives (bus passenger volumes). Panel data analysis methods were applied to measure explanatory values.

The Fixed Effects Model (FEM) was adopted in the analysis to estimate and explore the relation pattern between vehicle ownership and factors that influence ownership. A theoretical model that included all factors in the estimation of vehicle ownership can be written in logarithmic form:

$$\ln y_{it} = \alpha_i + \beta_1 \ln(\text{Income}_{it}) + \beta_2 \ln(\text{TP}_{it}) + \beta_3 \ln(\text{Urb}_{it}) + \beta_4 \ln(\text{BUA}_{it}) + \beta_5 \ln(\text{Road}_{it}) \\ + \beta_6 \ln(\text{UrbDen}_{it}) + \beta_7 \ln(\text{Bus}_{it}) + \beta_8 \ln(\text{Taxi}_{it}) + \varepsilon_{it}$$

Where: $\ln y_{it}$ = vehicle ownership rate in a city i during year t (per 1000 persons)

Income = average annual income per capita

TP = total population (1000 persons)

Urb = urbanisation was the % of the population in an urban area

BUA = total built-up area

Road = area of roads per capita (m^2 per person)

UrbDen = urban population density (1000 persons per km^2)

Bus = yearly volume bus passengers m (1000 persons per trip)

Taxi = total number of taxis

α_i = city specific constant term in the regression model

ε_{it} = error term

All analyses were conducted in the statistical tool Stata.

2.7 Vehicle ownership forecasting in low-income countries

Efforts have been made over the years to develop cultured models for vehicle ownership forecasting within developing countries. These efforts look at patterns of vehicle ownership growth limited to low-income countries around the world. Traffic is growing, especially in poor nations, and vehicle ownership forecast is a great tool to assist the government in preparing ahead in terms of planning for higher volumes of traffic.

An increase in vehicle ownership implies potential increase in fuel prices within the nation. Growth in vehicle ownership brings about urbanisation and strains transport infrastructure and maintenance of these road facilities. It is valuable to the World Bank to know whether the trends in vehicle ownership patterns are displayed over a range of low-income countries for funding purposes (Button et al., 1993).

In most low-income countries traffic growth is measured mainly by vehicle ownership, though goods transported can also be important to traffic growth. In subsequent analyses per capita, vehicle ownership is used as the dependent variable, and this allows for easier comparison with other studies and limits statistical problems arising when using data from countries of a widely different population.

It is not uncommon to witness slow growth in vehicle ownership within developing countries, compared to developed countries. Button et al. (1993) suggest the use of a non-linear forecasting basis. A limited number of studies have applied aggregate data in observing vehicle ownership levels amongst low-income countries, while the difference between low-income countries and industrialised nations has been shown by using linear and log-linear specifications.

Button et al. (1993) applied the aggregate quasi-logistic approach in their study of vehicle ownership forecasting for low-income countries. This approach has been used in local and national forecasting within industrialised nations.

Equation:

$$p = \frac{s}{1 + e^{-a}x_1^{-b2} \dots x_a^{-ba}}$$

Where: p = probability of an individual owning a car

S = saturation level of vehicle ownership per capita

$x_1 x_2 \dots x_a$ = socio-economic influences on ownership

$b_1 b_2 \dots b_a$ = parameters

This equation is then converted into logarithmic form to produce the operational model. Values of P become the actual per capita vehicle ownership levels for each country included in the research. Parameters are determined by the linear regression process (applying least square methods).

Logarithmic equation:

$$\ln\left(\frac{p}{s-p}\right) = a + b_1 \ln x_2 + b_2 \ln x_n \dots b_n \ln x_n$$

Dummy variables for each country are included in the developed model to take account of national features, and a time trend is included as an independent variable. Income is included as the independent variable influencing per capita vehicle ownership at national level.

2.8 Summary and conclusion

The literature review indicated that there has been significant research conducted on the forecasting of vehicle ownership, using various models available. Literature in a South African context on this topic is however limited. The literature has shown that vehicle ownership forecasting models are vital for local and national governments and manufacturing industries. The ability to forecast provides industries with the capacity to provide for an approximated guaranteed demand.

Various models have been developed for the forecasting of vehicle ownership, providing alternatives for different purposes of the required forecast. All the models are used for forecasting; however, the choice of a model will depend on the type of forecasting that needs to be done.

Additionally, the literature review presented constructive awareness of what has been done internationally and within South Africa's borders, and the success of the various forecasts. Moreover, the important factors influencing vehicle ownership forecasting were also provided within the literature, together with the relationship between these factors and vehicle ownership, with the main influencing factors being income and geographical locations of the households. Different forecasting models apply to different data sets.

CHAPTER 3: DATA

3.1 Introduction

In the previous chapter the literature was reviewed concerning vehicle ownership forecasting, models used in forecasting, and factors influencing ownership of vehicles. This included both international and South African literature.

The purpose of this chapter is to outline the data collecting process and present some descriptive patterns and trends with the data. The Chapter also attempts to explain the trends in the data. As the research considers both mode ownership and vehicle forecasts, both data sources are considered and discussed.

This chapter commences by providing a detailed introduction of the National Household Travel Survey (NHTS) 2013. It then continues into discussing descriptive statistics and discuss the dependent and independent variables used in the mode ownership models. Finally, the chapter examines the Road Traffic Management Corporation (RTMC) and The National Traffic Information System (eNaTiS) data used for vehicle forecasting. Some preliminary analysis on the historical patterns of the data is presented.

3.2 Historical vehicle population

The process of conducting a forecast of the vehicle fleet requires the use of historical data. This study makes use of historical vehicle fleet data extracted from the Electronic National administration Traffic Information System (eNaTiS)¹ and the Road Traffic Management Corporation (RTMC). Historical data on South Africa's vehicle fleet is analysed to identify past trends and patterns. This chapter gives a historical background and overview of what the growth of the vehicle fleet in South Africa looked like over the years (1986-2016).

Figure 3.1 displays data per vehicles per 1 000 of the population for all countries of the world for the year 2014 (JA Van Rensburg & Krygsman, 2019; J Van Rensburg & Krygsman, 2015). South Africa ranked 84 out of the countries included, meaning in 2014 South Africa was within the top 50% of the 191 countries included in the dataset. South Africa had the third highest number of vehicles per 1 000 of the population of the BRICS countries with 165 vehicles per 1000 of the population, ahead of China and India. Russia recorded 293 vehicles per 1 000 of the population and Brazil 249 vehicles per 1 000 of the population.

¹ eNaTiS is a strategic state resource that enables the state to improve South Africa's existing road infrastructure and contributes in accomplishing continuous advancement of transportation infrastructure management.

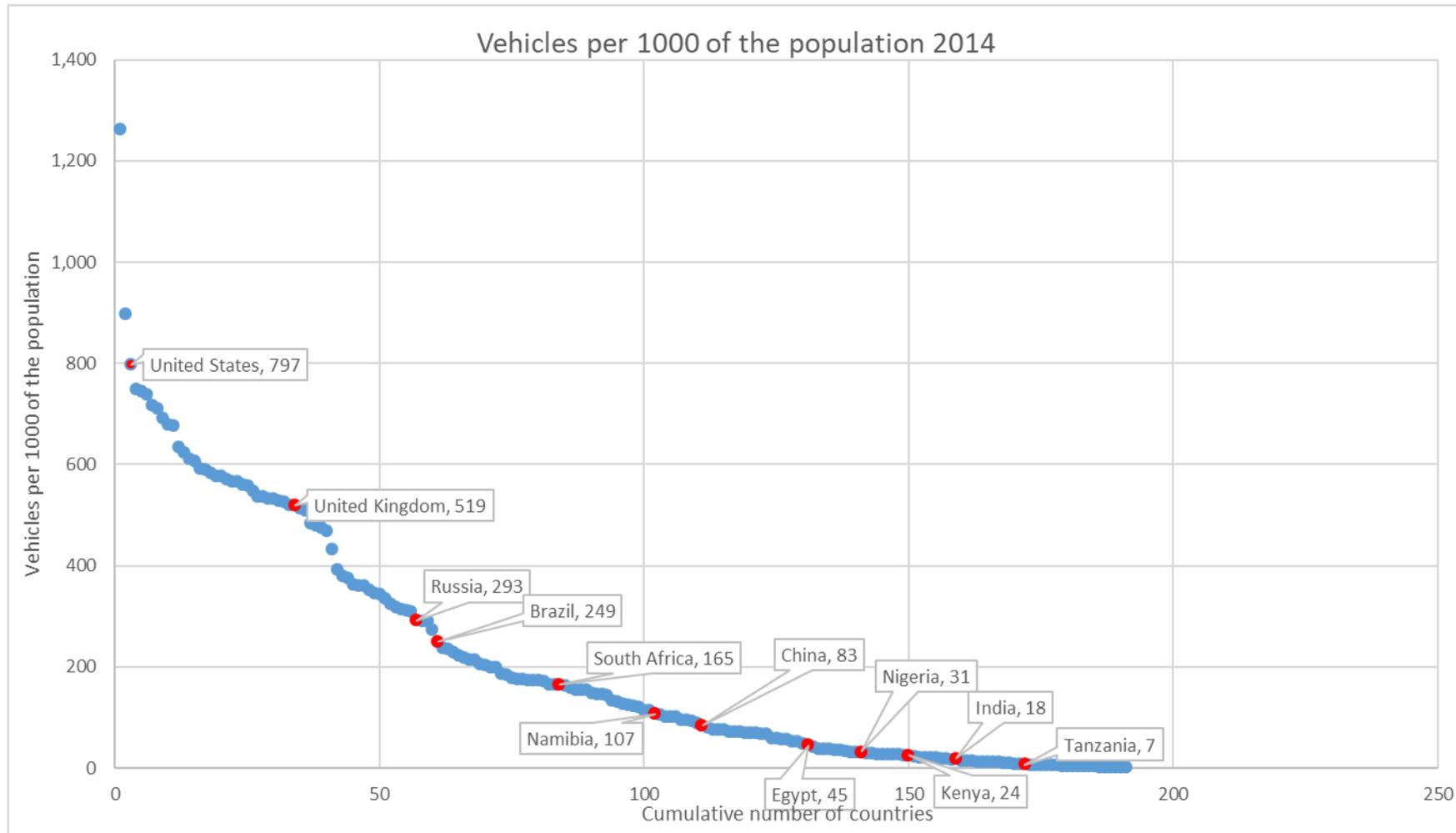


Figure 3.1: Scatterplot of countries comparing vehicles per 1 000 of the population in 2014.

Source: (JA Van Rensburg & Krygsman, 2019; J Van Rensburg & Krygsman, 2015)

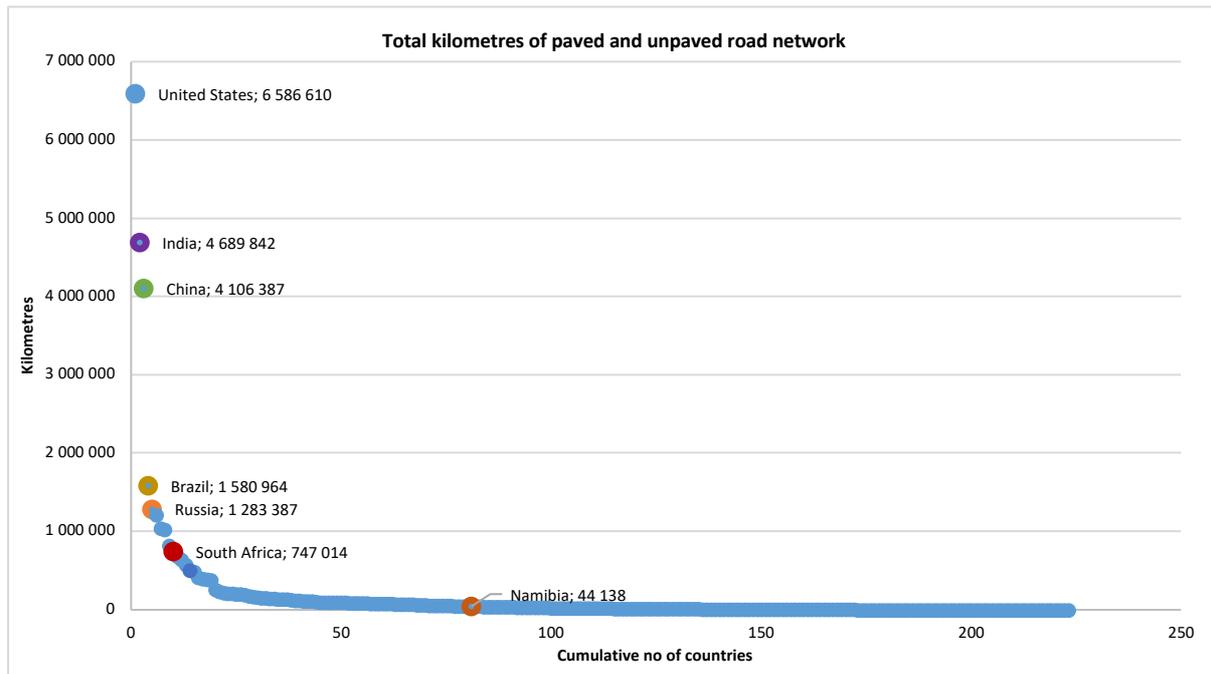


Figure 3.2: Scatterplot of countries comparing total kilometres of paved and unpaved road

Source: (JA Van Rensburg & Krygsman, 2019; J Van Rensburg & Krygsman, 2015)

South Africa's national roads represent an integrated transport network, which contributes and adds value to infrastructure, quality of life of commuters and contributes in a sustainable economy. In Figure 3.2, South Africa's road network is ranked 10th longest in the world, with a total of 747 014 kms of paved and unpaved road. South Africa being the only African country in the top 10 world longest road networks. Accessibility to road infrastructure positively influences ownership of vehicles. With new road infrastructure potential employment is created and access to new employment markets, with access to transportation and thus, providing income for the population.

South Africa's vehicle population is displayed in Figure 3.3, for the period 1986 to 2017 (RTMC, 2018). During the apartheid era the South African vehicle population was 4 228 523 in 1986 growing to 4 870 609 in 1993, reflecting an increase of 585 276. In 1994, barely one year into the post-apartheid era, the vehicle population had increased notably from the 1993 vehicle population of 4 870 609 to 5 314 411 vehicles, a result of the increased accessibility to social and household infrastructure.

Government expenditure focused on improving household welfare and democracy brought a new demand for vehicle ownership from those previously oppressed, as a desire for easier accessibility to economic opportunities or social spaces and services. The first democratic

election of 1994 was the major socio-economic event taking place in South Africa which led to the restoration of its Commonwealth membership and the lifting of sanctions (SAHO, 2019b). This restoration of membership meant international cooperation and an advancement in economic environment, social development and human rights for South Africa.

Three periods are sectioned in Figure 3.3, pre-election, transition and post-election. These periods reflect increasing annual vehicle growth rates. The compound growth rate in the pre-election period, 1986-1993, was 1.78% while a growth rate of 2.66% was observed in the transition period, 1993-2004. The post-election period, 2004-2017, showed the largest overall growth of 3.64%. Over the entire period, the vehicle growth average 3.05%.

In 1994 South Africa had a vehicle population of 5 314 411 with a population of 41 218 901 people (RTMC, 2018) (Statistics South Africa, 2018). This indicates that in 1994 South Africa had 130 vehicles per 1 000 of the population, with a recorded annual growth rate of 8.43%.

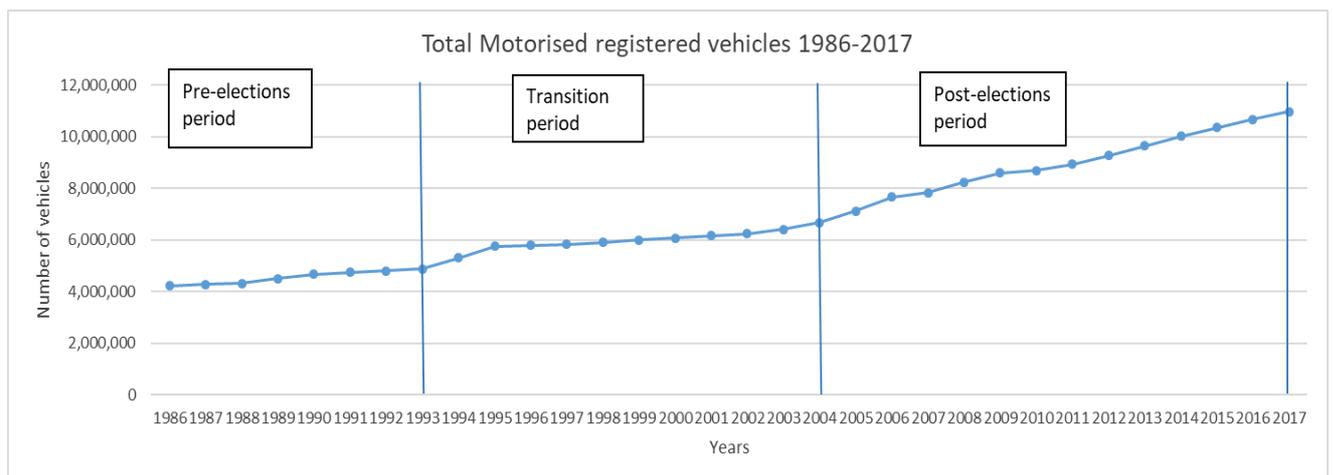


Figure 3.3: Annual vehicle population.

Source: RTMC (2018)

Figure 3.4 shows the relationship between the population and economic growth, observed using real GDP². The real GDP measures the growth of the South African economy, by comparing one quarter of the country's gross domestic product to the previous quarter. Real GDP is used as it is the most accurate measure of economic growth, it removes the effects of inflation on the data.

² Real GDP (also known as GDP constant local currency) is an inflation-adjusted measure that reflects the value of all goods and services produced by an economy in a given year, expressed in base-year prices and referred to as constant price.

The figure demonstrates the existence of a positive relationship between vehicle fleet, population and real GDP. The increase in the population, motivates an increase in the number of vehicles. The more people, the more vehicles. Real GDP reflects the size of the economy. This relationship is not uncommon, as the increase in economic growth brings social development and job creation to a country, resulting in an income increase for households (increased buying power) which could encourage a household to purchase a vehicle.

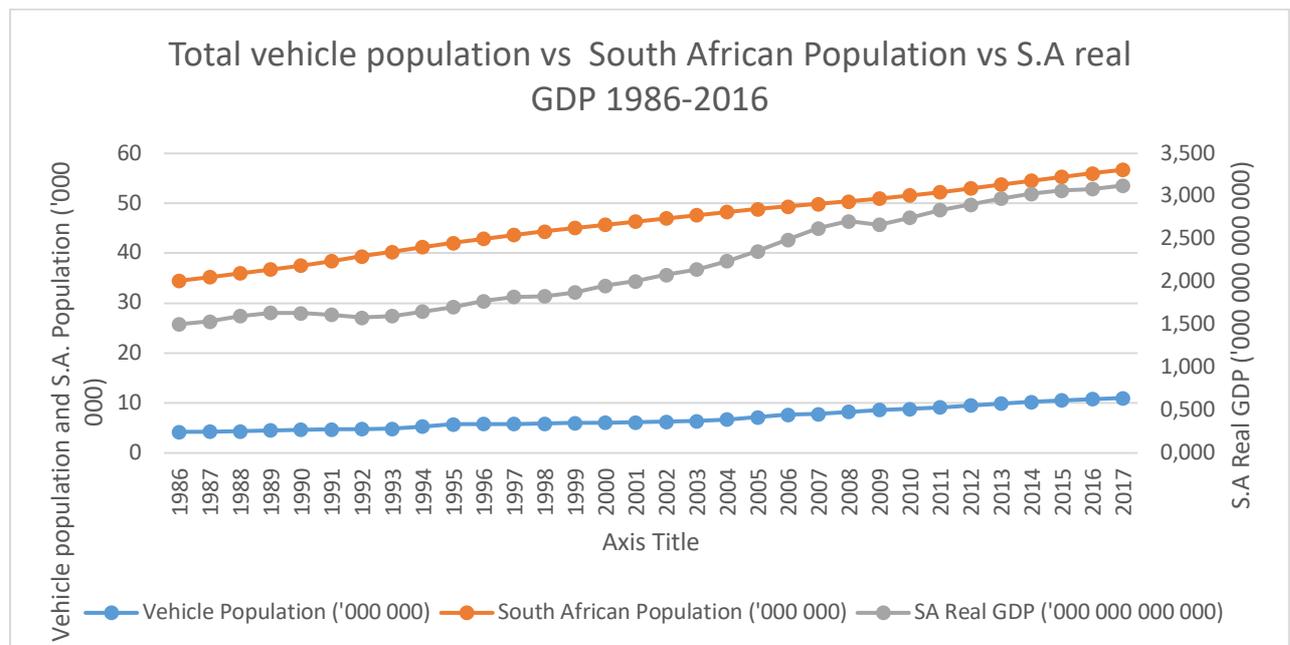


Figure 3.4: Comparing vehicle population, population and real GDP 1986-2016

Source: StatsSA (2018)

The complex relationship between vehicle population, real GDP and human population growth rates is presented in Figure 3.5. The average real GDP growth rate for 1987-2016 is 2.33%. A strong growth rate attracts investors and in turn economic growth.

The average growth rate of vehicle population for the period 1986-2017, is 3.05%, approximately twice that of the general population, of 1.62%. The growth rate of the South African general population appears more stable compared to that of real GDP and vehicle population. Vehicle population grows at a higher annual average growth rate, compared to real GDP. Graphically the growth rate strength of both the vehicle population and real GDP is shown to be constantly fluctuating, with that the growth of real GDP at times laying in the negative.

In 1992 South Africa experienced the mass killing of protestors outside of Bisho, now known as the Bisho massacre. The impact of this political event shock the economy and resulted in a plunge in real GDP growth during 1992, dropping into negatives. In 1994 the impact of a

major economic event in South Africa, namely the advent of democracy is indicated by the high peak in vehicle population evident in the graph, with GDP and population still showing constant growth after 1994.

Between 1995 and 1996 there is a significant decrease in vehicle population growth, the growth remains at a low growth rate until 2002. This reflects that in real terms South African households are becoming poorer. Between 1995 and 2003, the number of households earning less than R1 000 a month had increased from 24 to 49 per cent (Lombard, Cameron, Mokonyama, & Shaw, 2007).

In 2008, the effect of the global financial crisis, which led to a period of recession in 2009 is seen in the large drop in GDP and vehicle population shows a steady decrease from 2008 till 2009, and starts to increase again in 2010, with the rest of the economy.

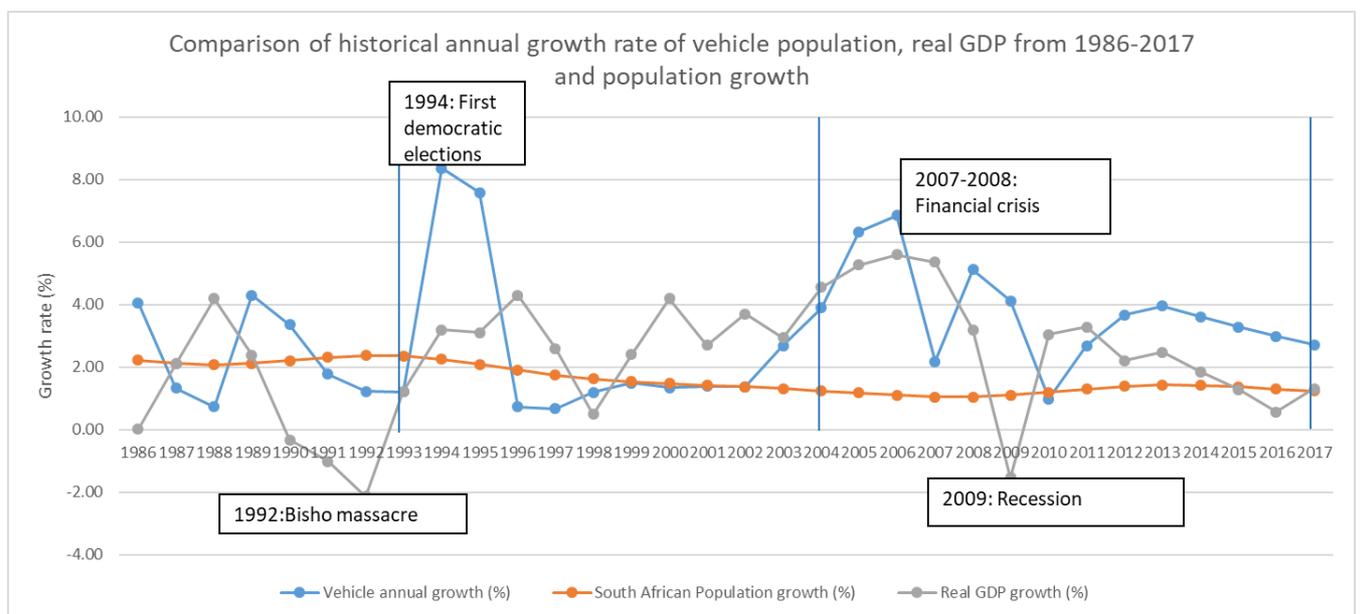


Figure 3.5: Comparison of vehicle population, human population and real GDP growth rate

Sources: RTMC (2018); The World Bank (2017), (BBCNews, 2018), (SAHO, 2019a)

Figure 3.6 presents a scatterplot of the relationship of real GDP to vehicles per 1 000 of the population. A logarithmic trend line is fitted.

The independent variable in Figure 3.6 is real GDP and the dependent variable is vehicles per 1 000 people. The logarithmic equation of the trend line is:

$$\text{Vehicles per 1000 of the population} = 94.282 \ln(\text{Real GDP}) - 2529.2$$

The equation states that, if real GDP increase by 1%, then vehicles per 1 000 of the population are expected to increase by $94.282/100 = 0.94282$ units. Therefore, the elasticity is 0.94, this

means for every 1% increase or decrease in real GDP, the number of vehicles per 1 000 of the population will change by 0.94282 vehicles.

The value of the R square in the context of the analysed data represents the proportion of the change in vehicles per 1 000 people that is explained by real GDP. Therefore, 91.64% of the change in vehicles per 1 000 of the population in South Africa, is explained by the change in real GDP.

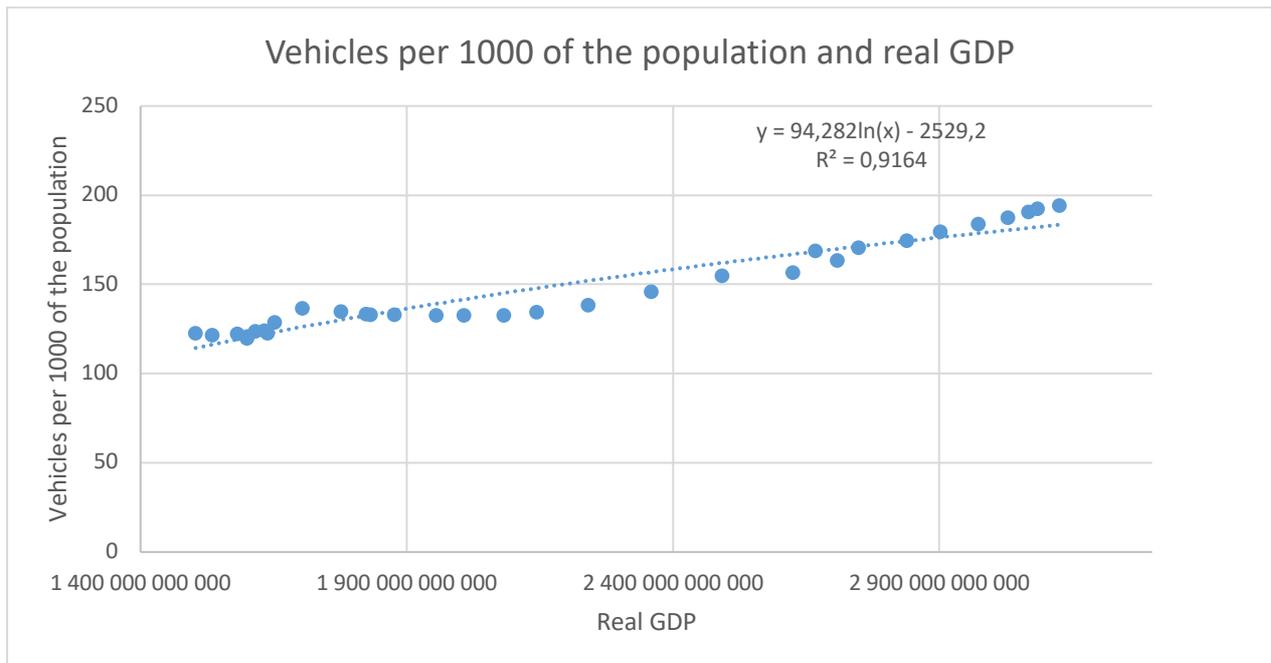


Figure 3.6: Scatterplot representing real GDP with trend line

Source: The World Bank (2017) (vehicles per 1 000 of population calculated using World Bank and eNaTIS data)

3.3 2013 South African Household Travel Survey (NHTS)

3.3.1 Introduction to 2013 National Household Travel Survey

The National Household Travel Survey (NHTS) 2013 is recorded as the second NHTS conducted in South Africa, with the first being the National Household Travel Survey launched in 2003 (Statistics South Africa, 2014). The questionnaire used in conducting the 2013 survey was based on the 2003 NHTS questionnaire and adjusted according to information needs within industries. The questions used in the survey are attached as Appendix A. Statistics South Africa (StatsSA) implemented the 2013 NHTS from February to March 2013, and preceding the main survey a pilot survey was rolled out using a smaller scale, for training and testing the contents of the survey and the questionnaire.

Sample design for the NHTS 2013 was based on the census of 2011 enumeration areas³ (EAs) and on a two-staged random stratified sampling (sampling by subpopulations). The stratified sampling first selected a sample of 5 034 primary sampling units (PSUs) from the dwelling structure of the Census 2011, at a transport analysis zone and provincial classification level.

In the selected sample of 5 034 PSUs, 22 were found to be unoccupied and thus the survey continued with 5 012 PSUs, sampling 51 341 dwelling units (DUs) from the remaining PSUs. However, of the 51 341 DUs sampled, 849 DUs presented incomplete or no questionnaires. In the sampled PSUs, 4 957 had at least one responding household, 5 had all DUs with households out of the given sampling scope; however the remaining 50 PSUs had sampled DUs that had no responding households. Table 3.1 shows a summary of the PSUs per province.

Table 3.1: Summary of the number of primary sampling units

Province	Number of PSUs	Average number of dwelling units per PSU	Total number of dwelling units
Western Cape	559	10	5,528
Eastern Cape	710	11	7,497
Northern Cape	206	10	2,103
Free State	350	10	3,601
KwaZulu-Natal	965	10	9,806
North West	388	9	3,628
Gauteng	1,025	10	10,683
Mpumalanga	366	10	3,794
Limpopo	443	11	4,701
RSA	5,012	10	51,341

Source: Statistics South Africa (2014)

The NHTS collected data using three phases (Statistics South Africa, 2019):

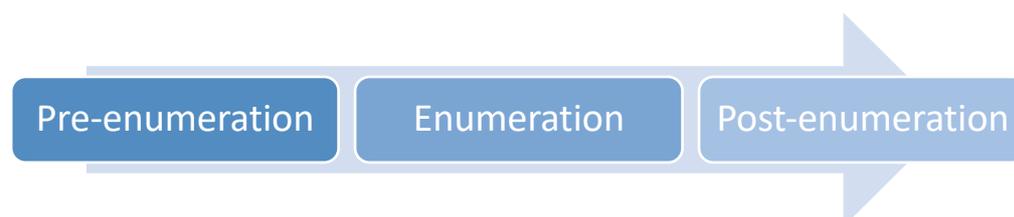


Figure 3.7: Three phases of data collection.

^{3 3} An enumeration area refers to the geographic areas covered by the Census 2011 representatives. It is an area composed of one or more adjoining blocks. The EAs cover all the territory of South Africa. EAs are only used in the collection of census data.

- Pre enumeration: Involved planning, funding, publicity, questionnaire design, printing, recruitment and training. Publicity included informing potential respondents of the upcoming survey and its purpose. Training of provincial survey coordinators, fieldwork coordinators, fieldwork supervisors and fieldworkers.
- Enumeration: This phase made use of 51 341 DUs across the nine provinces. Surveyors were allocated within the DUs, equipped with questionnaires. Questionnaires were completed, quality assurance took place in this phase and data was finally captured.
- Post-enumeration: Data was processed and converted from the questionnaire to electronic format, coding of open-ended questions, analysis of the data, compilation of metadata, data and report dissemination (continuous process).

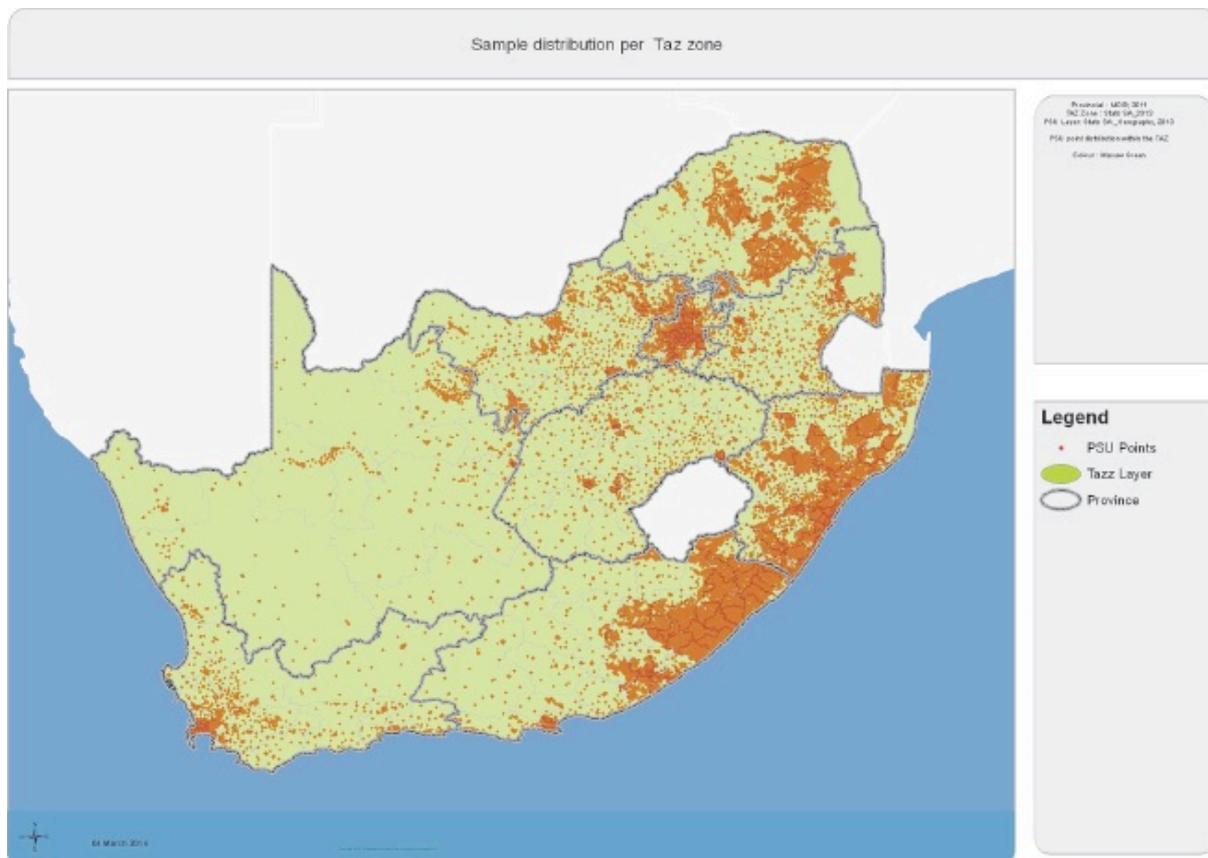


Figure 3.8: Sampled travel analysis zones.

Source: Statistics South Africa (2014).

3.3.2 Limitations of the National Household Travel Survey

The study conducted in the collection of the NHTS data had limitations as the sample design was exclusive of households and individuals who lived in institutions such as boarding houses,

residential hostels, military barracks and hospital accommodation. The time period in which the survey had to be conducted and surveyors trained was limited, such that it made it difficult to implement surveys in the same way in all sampled districts. This potentially could have caused errors in the manner the survey was conducted Statistics South Africa (2014).

3.3.3 Key findings in the National Household Travel Survey 2013

Table 3.2 summarises the response rates of the sampled households and the percentages of the households in the National Household Travel Survey.

Table 3.2: Survey response rates

Response code	Label	Frequency	Percentage	Cumulative frequency	Cumulative Percentage
1	Response	43 462	82.37%	43 462	82.37%
2	Non-response	5 314	10.07%	48 776	92.45%
3	Out of scope	3 986	7.55%	52 762	100%

Source: Statistics South Africa (2014)

An overview of the findings embodied in the NHTS results shows that beyond half of the travellers dwell in four provinces of the country, namely Gauteng, KwaZulu-Natal, Eastern Cape and Western Cape. Percentage proportions of this population in terms of dwelling type are as follows: around 85% of individuals residing in urban, metropolitan and rural areas reported to have travelled within the time period of the survey, with only 75.4% citizens living in rural areas likely to travel.

In the category of education institution and level of education, the survey found that learners and individuals that attended pre-school ABET and literacy classes walked to the varying institutions. Of those who attended higher education institutions approximately 30.5% used taxis and 24.7% used private vehicles for the trip. This shows that there is a connection between level of education and travel mode.

Workers in the survey were from age 15 years and older, and the travel patterns and choices of workers indicated that 38.4% of them made use of private vehicles, while the majority made use of public transportation and some mainly residents in more rural areas walked all the way to work places.

Concerning households, the survey concluded that 32.6% of households' choice of travel mode was influenced by travel time⁴, the cost to travel was an important influencing factor to 26.1% of the households, and 9.25% of the households stated that flexibility was a major influence on their choice of travel mode. Safety from road accidents was deemed a crucial

⁴ The time spent travelling from point of origin to destination.

influencing factor by 8.7% of the households surveyed. A comparison between the results obtained from the first NHTS 2003 and the recent NHTS 2013 shows that there was a significant increase in the percentage of households owning or having access to vehicles, increasing from 22.9% in 2003 to 28.5% in 2013 (Statistics South Africa, 2014).

In conclusion, the province of Gauteng had the highest percentage (39.3%) of people aged 18 and older in possession of a driver's licence compared to the other eight provinces. The Western Cape had the second highest figure of 36.4% of licence holders, and KwaZulu-Natal was third with 20.8%. The absolute number of licence holders in 2013 were higher compared to those surveyed in 2003. With the increase in household vehicle ownership, this increase is an anticipated outcome. A majority of those households owning or having access to vehicles dwelled within urban areas.

3.4 Descriptive statistics

Household vehicle ownership decisions are influenced by countless factors; in this study a number of explanatory variables are identified to assist in the analysis of the National Household Travel Survey data. The following variables are identified: Main dwelling, income quantiles, household expenses and geographical location. In this data chapter population groups, income category and level of education are included only as a base of comparison and explanation of identified trends.

The National Household Travel Survey data analysed is divided into three main categories, namely: households with 0 vehicles, households with 1 vehicle and lastly, households with 2 or more vehicles. Figure 3.9 shows an overview of the percentage split of vehicles owned per household surveyed within each province. This includes data of households that did not specify how many vehicles they own and have access to. In this figure it is observed that the Western Cape and Gauteng have the largest percentage of surveyed households owning two or more vehicles at 17.2% and 15.4% respectively, with the Eastern Cape producing the largest percentage of households owning no (zero) vehicles per household at 83.8% of all surveyed households within the province.

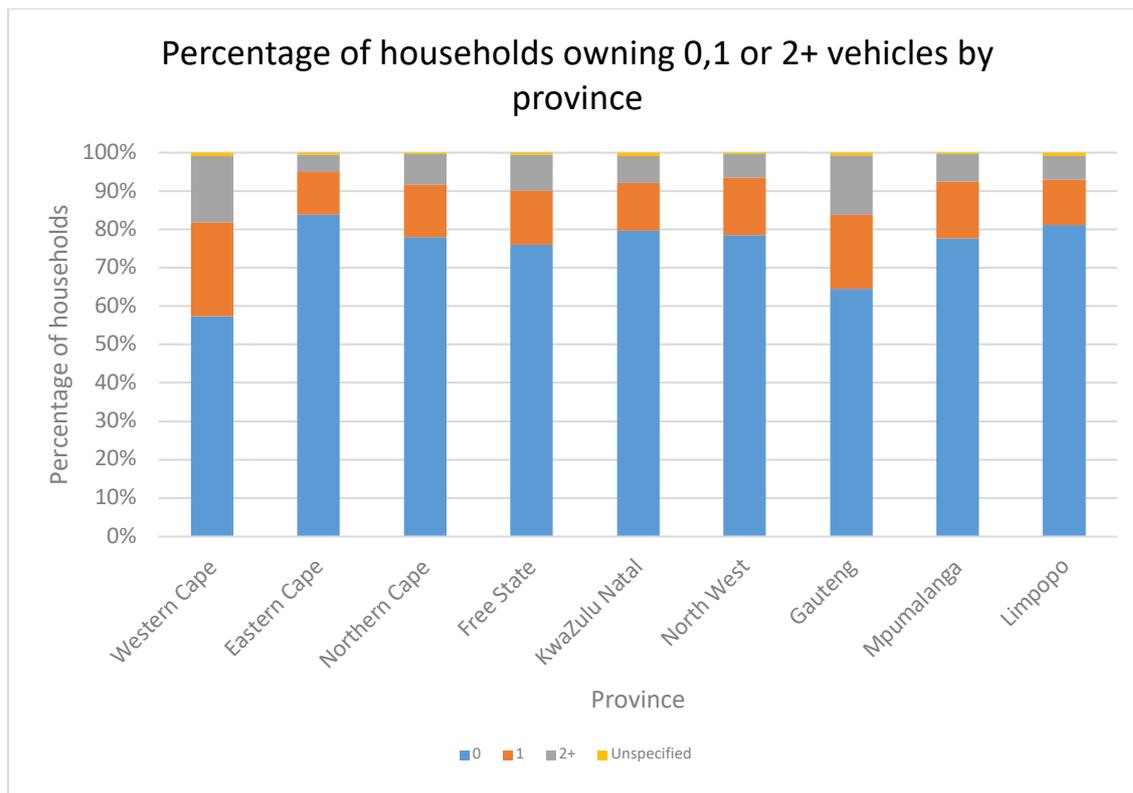


Figure 3.9: Percentage of households by province

Source: (Statistics South Africa, 2014)

The relationship between vehicle ownership and household geographical location displayed by Figure 3.10 shows that households within the metropolitan areas own more vehicles, with 15.7% of households owning two or more vehicles and 87.10% of households residing in rural areas owning no vehicles. The geographical location has an effect on travel experience from one spatial location to another, and the decision to purchase is influenced by factors such as access to road infrastructure within that specific location. This supporting figures to Figure 3.9 above.

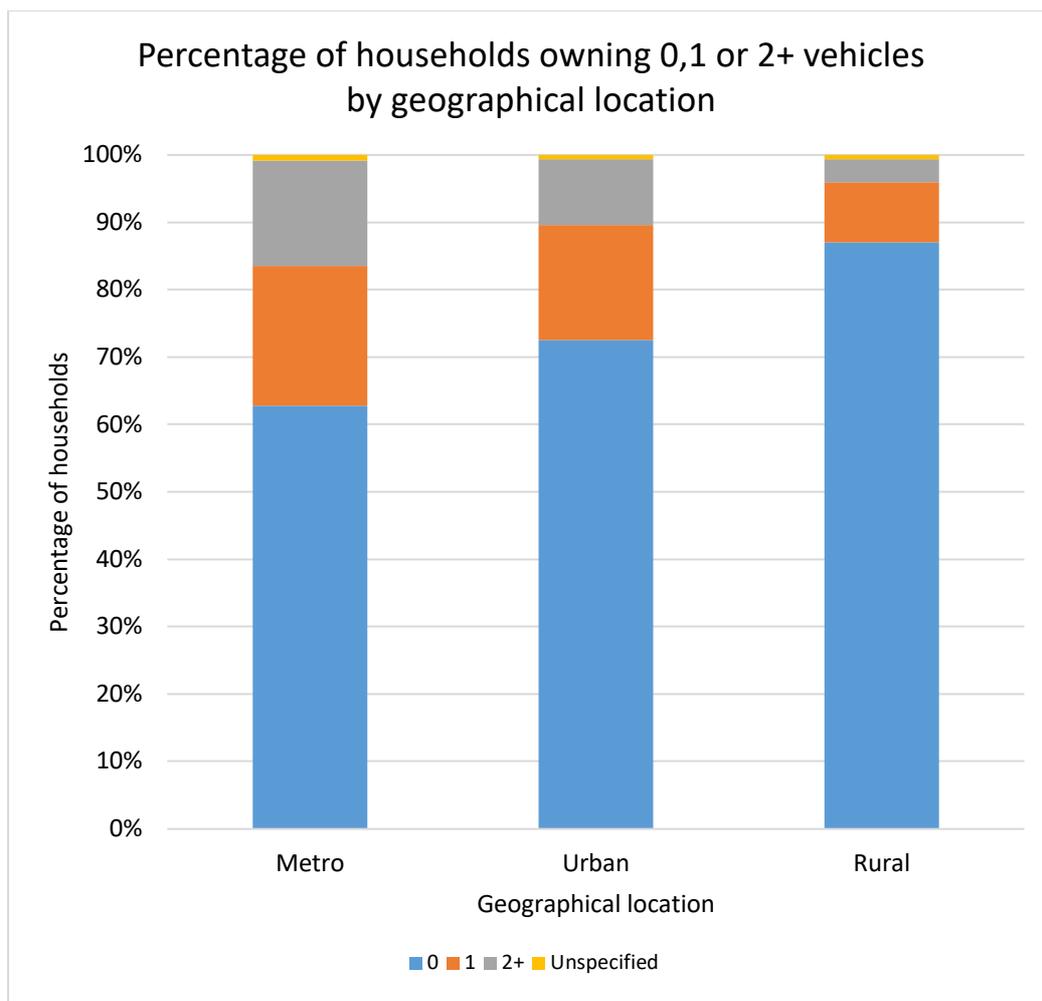


Figure 3.10: Household percentage by geographical location

Source: (Statistics South Africa, 2014)

The percentages of households owning vehicles by main dwelling are demonstrated by Figure 3.11. The outcomes of the relationship between the main dwelling and household vehicle ownership support the conclusions made from the relationship between geographical location and households' vehicle ownership. Among households residing in informal dwellings 90.23% have no vehicles. Most informal dwellings include houses made from more traditional materials and are dwellings such as huts, with these mostly found in rural areas, while 14.76% of households residing in formal dwellings, such as concrete block structures own two or more vehicles, and are found in more metropolitan areas.

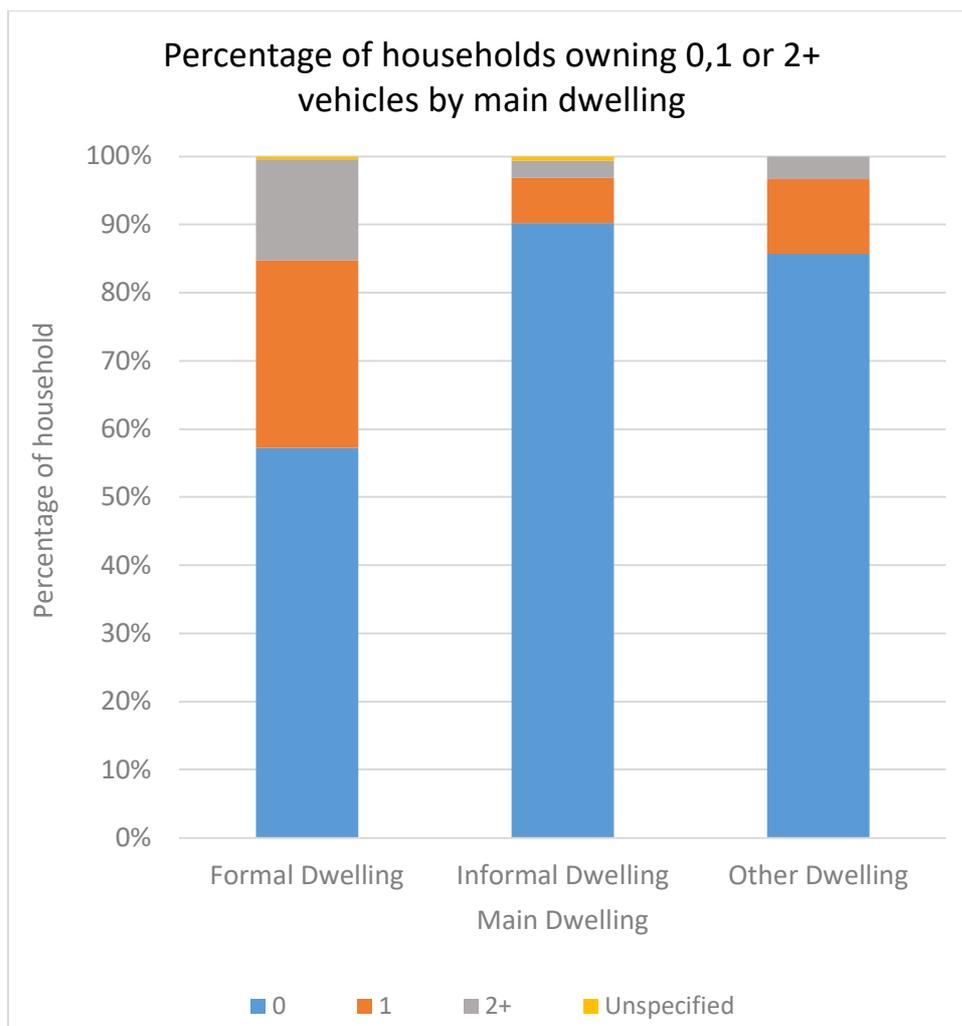


Figure 3.11: Household percentage by main dwelling

Source: (Statistics South Africa, 2014)

The observations in Figure 3.12 showing income quantiles correlate with the observations in Figure 3.13 showing the annual income categories for households.

Total monthly household income per capita is used to calculate income quantiles. In other words, the quintiles are divided as follows: 20% of individuals with the lowest incomes make up quintile 1 (R0-R395,) between 20% and 40% fall within quintile 2 (R395.11-R828), individuals between 40% and 60% are quintile 3 (R828.33-R1 600), those between 60% and 80% are quintile 4 (R1 6001-R4 017) and 20% of individuals who earn the highest incomes make up quintile 5 (R4 023-R222 000) (Statistics South Africa, 2014)

Households in the highest income quantile show the lowest percentage of households without vehicles and the highest percentage of households with two or more vehicles. This finding is not unexpected as it is common that the more income one has the easier it is to purchase private vehicles.

The analysis in the income quintiles is supported by Figure 3.12. The number of vehicles owned by households increases steadily with income, from 11.63% of households owning two or more vehicles and earning between R0-R6 000 annually, to 65.28% of households owning two or more vehicles and earning R192 001 or more annually.

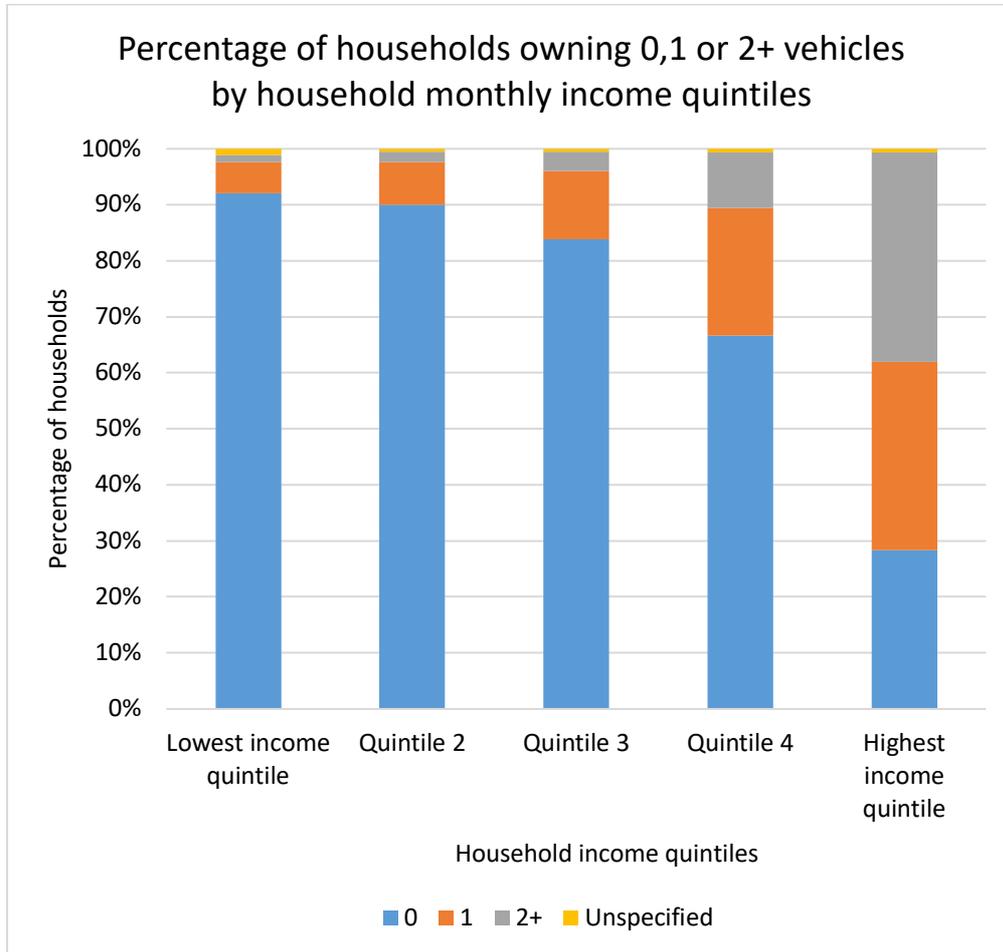


Figure 3.12: Household percentage by income quintiles

Source: (Statistics South Africa, 2014)

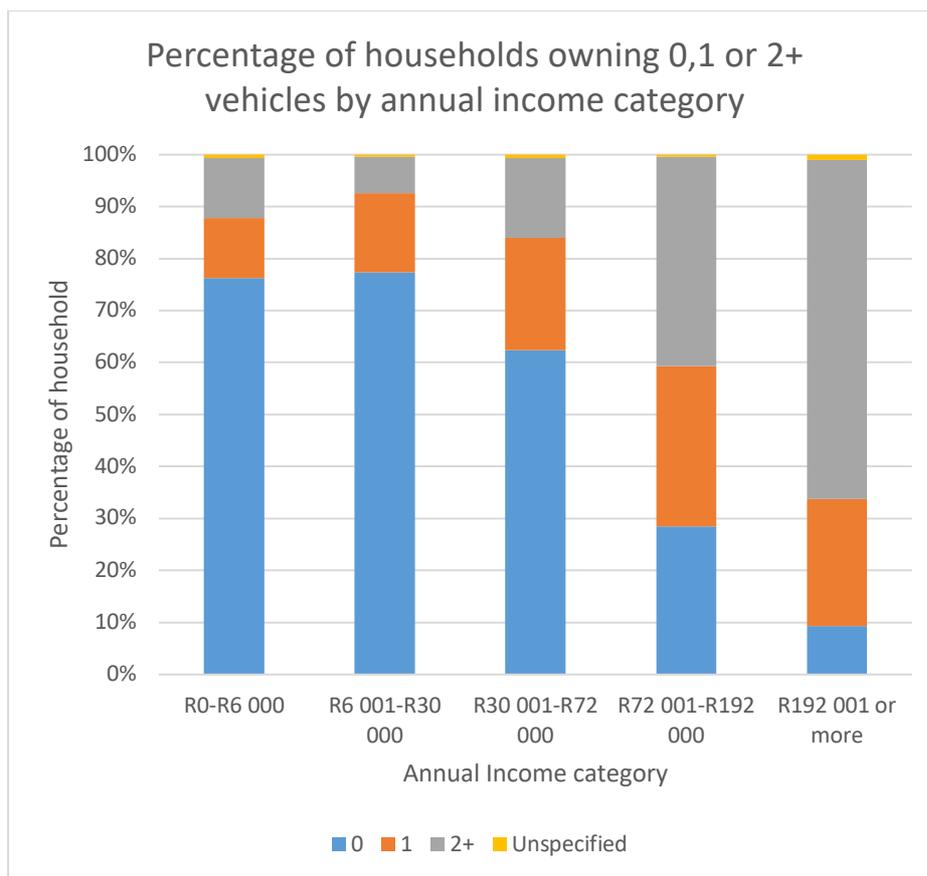


Figure 3.13: Household percentage by income category

Source: (Statistics South Africa, 2014)

Household vehicle ownership is shown to be directly related to household income and expenses; as income increases vehicle ownership within households tends to increase. Within our scope of data used in this study, Figure 3.14 shows that the sampled households display a directly related relationship of vehicle ownership and expenses; the more vehicles a household has the higher the expenses to be handled within households.

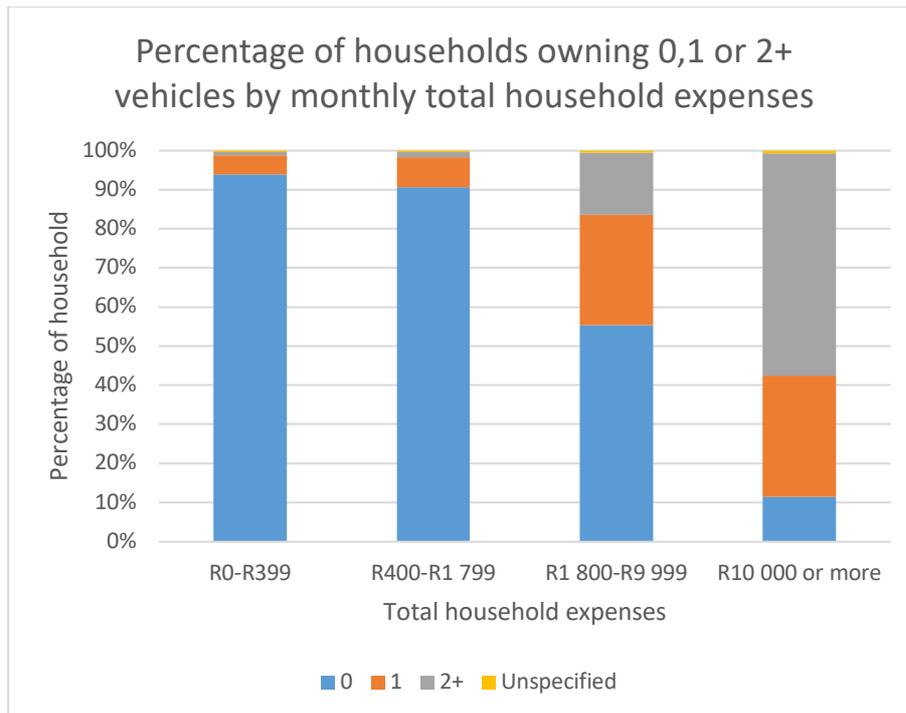


Figure 3.14: Household percentage by total household expenditure

Source: (Statistics South Africa, 2014)

Figure 3.15 shows that 80.59% of black households do not own private vehicles, compared to 8.41% of white households not owning private vehicles.

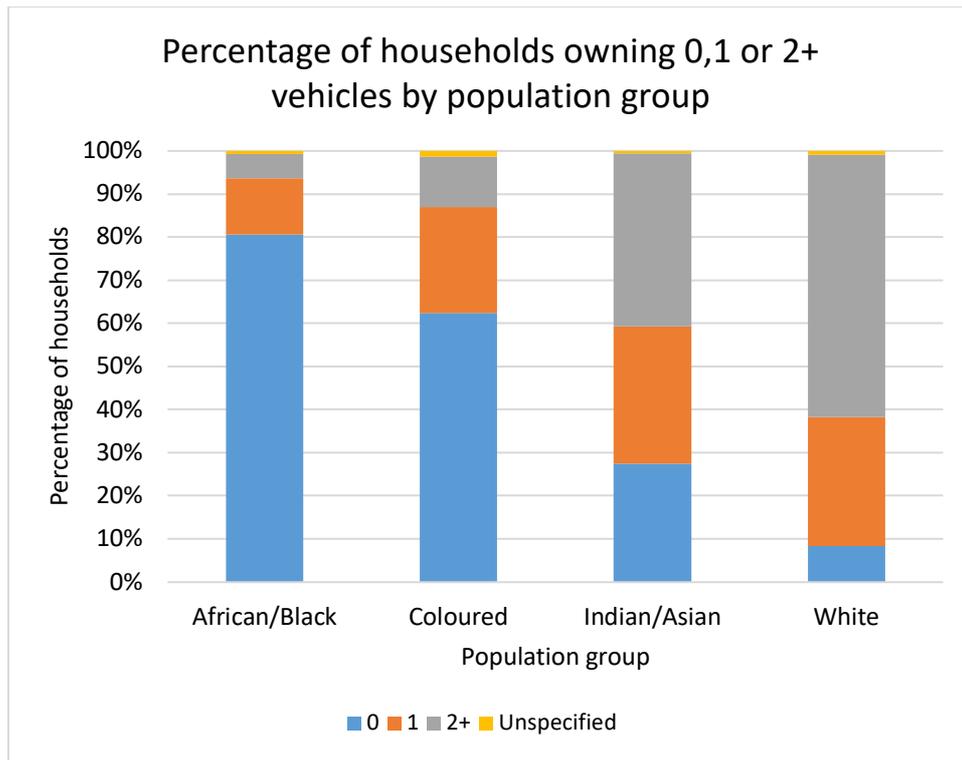


Figure 3.15: Household percentage by population group

Source: (Statistics South Africa, 2014)

Vehicle ownership increases with the level of education attained by individuals within the households, mainly because the level of education relates to the level of income. As shown in Figure 3.16, 84.51% of the households that have attended Adult Basic Education and Training learning centres are without vehicles. This percentage drops to 37.76% for those who attended higher education institutions and 37.5% for those with home based education/home schooling. Households that attended higher education institutions and home schooling have 34.65% and 36.88% of households owning two or more vehicles respectively.

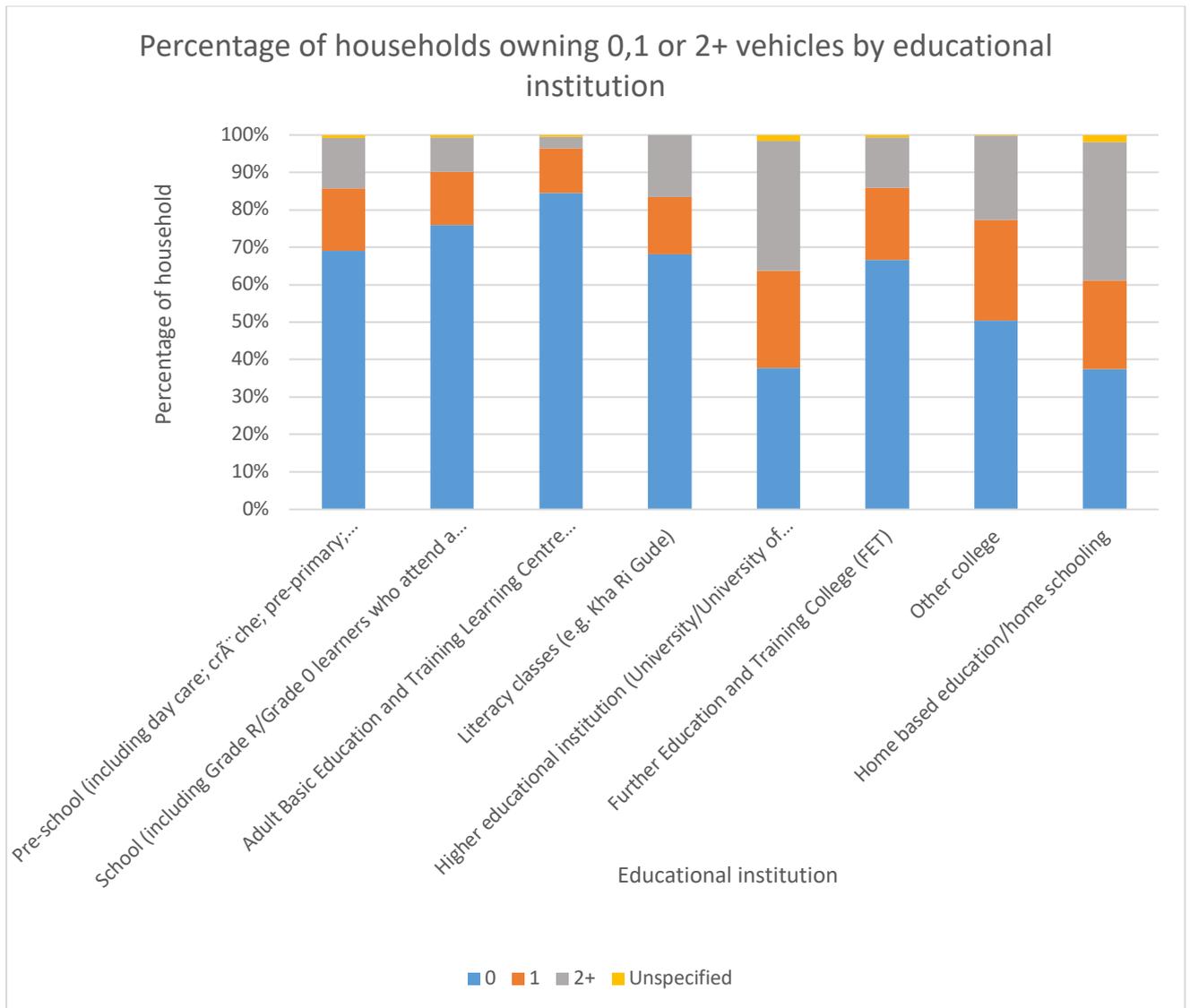


Figure 3.16: Household percentage by education institution

Source: (Statistics South Africa, 2014)

Figure 3.17 gives highest grades attained by percentage of households. The trend seen in Figure 3.17 agrees with that in Figure 3.16, as it is observed that the higher the grade attained the less households without vehicles get and the more households own vehicles, while 37.89% of households with tertiary education (including doctorates) own two or more vehicles and 77.96% of households with only primary education neither own nor have access to vehicles.

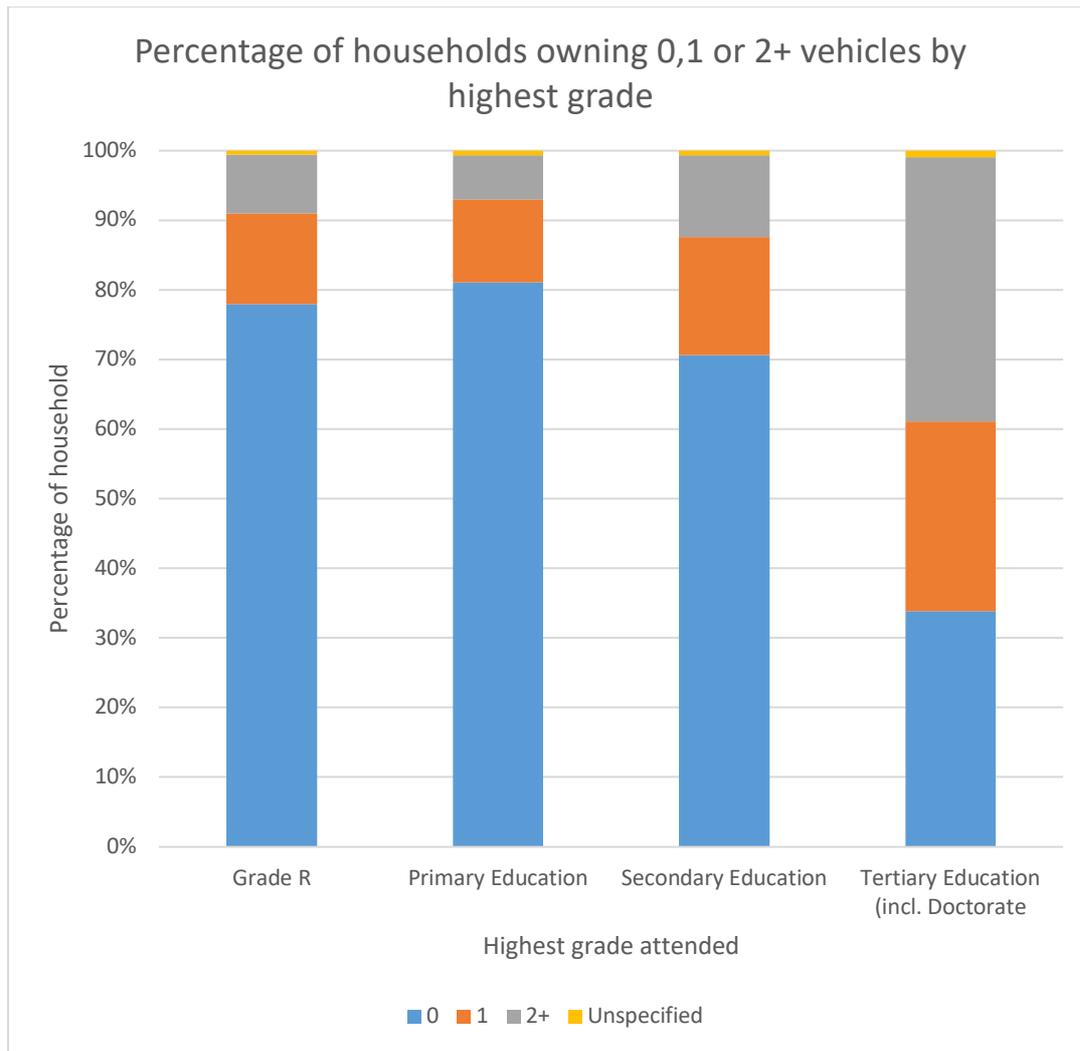


Figure 3.17: Household percentage by highest grade achieved

Source: (Statistics South Africa, 2014)

3.5 Summary and conclusion

The focus of this chapter was the data used in the study, to investigate vehicle prediction and vehicle ownership for households. The section included the collection process of the data presented in the National Household Travel Survey, how the data was sampled and when the surveys were conducted. The description of the data variables was discussed and graphically presented.

Vehicle forecasting data was compiled from two different data sources, both sources containing historical vehicle population data. The annual average growth rates were given and factors impacting the historical patterns were discussed. The South African human population, together with the real GDP were presented as factors affecting the vehicle population.

This chapter gave a brief overview of the historical data required in the forecasting of vehicle ownership and household vehicle ownership, summarising past trends that may have an impact on the expected forecast for future vehicle ownership within South Africa. Households owning no vehicles dominated the results.

CHAPTER 4: METHODOLOGY

4.1 Introduction

As discussed in Chapter 3, the purpose of the study is to forecast the vehicle fleet for South Africa and build a model for vehicle ownership model for South Africa using the National Household Travel Survey of 2013. This chapter presents and describes the research design, the methodology adopted in the research, and the model specifications used to predict and study household vehicle ownership using an appropriate model. The chapter is organised in two sections: the research questions and the research design.

4.2 Research questions

The purpose of this research is to firstly analyse vehicle ownership decisions for South African households and secondly to develop a forecasting model to predict the number of vehicles. The National Household Travel Survey (2013), the most recent survey conducted and Statistics South Africa on travel behaviour is used for the vehicle ownership model. Data provided by the Road Traffic Management Corporation (RTMC), together with Electronic national administration Traffic Information System (eNaTIS) on the vehicle fleet is used for the vehicle forecasting model.

These two research questions will provide the following answers

- What socio-economic factors influence household vehicle ownership?
- What will the growth in vehicles be for South Africa for the period 2018 - 2038?

These research questions can be structured as sub questions, including:

- What are the main factors influencing household vehicle ownership?
- How far do these factors influence vehicle growth?

4.3 Research design

The study takes the form of a quantitative study, making use of a secondary data analysis approach to gather the required data. The secondary data analysis approach is a practice of utilising quantitative data already collected. This approach saves time and avoided duplication of research data and efforts (Cole, 2018). The quantitative approach concentrates on numeric data collected using rigid techniques such as questionnaires or interviews and detailed, convergent reasoning rather than divergent reasoning.

Convergent reasoning is connected with analysis and decision making, taking ideas and sorting, evaluating and analysing it, whereas, divergent reasoning follows the process of

developing ideas with no discussion or analysis of the ideas (Boogaard, 2018). The aim of such a method is counting and classifying variables and building statistical models and figures, explaining what is observed from the data (Explorable.com, 2009).

Quantitative research makes use of tools such as surveys, questionnaires or software to collect numerical or measurable data required for the research and analysis. Quantitative studies look at mainly interpreting data and finding trends and seasonality within the analysed data. This approach is the most appropriate method applicable to this study, given that the secondary data used for the purpose of the study is numerical and objective in the questions asked in the questionnaires used to gather the data. Furthermore, the quantitative approach to research requires the researcher to remain objective from the subject matter.

The study implements a Multinomial Logistics Regression model (MRL) as the vehicle ownership model. The Multinomial logit model was first employed by Train and Lave (1978) in a study conducted to estimate a car type choice model, with the purpose of evaluating transportation energy consumption policies. Subsequently, there have been more studies on the logit models in vehicle modelling. Krygsman (2001) use logistics regression to estimate the choice between car and public transport use for The Netherlands (Krygsman & Dijst, 2001). Cirillo (2010) applies the Multinomial logit model in the study titled "Automobile Ownership Model" conducted in Maryland to forecast vehicle ownership. This study follows a very similar approach to estimate mode ownership for South Africa.

This regression model is an important method for categorical data analysis, a good fit for this research as the data analysed is categorical and the dependent variable has more than two categories. It is a model used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables (LaerdStatistics, 2018).

This model is limited by 6 data assumptions; (1) Dependent variable should be nominal, (2) one or more independent variable that are ordinal, nominal or continuous, (3) independence of observations and dependent variable categories need to be mutually exclusive and exhausted, (4) no multicollinearity, (5) a linear relationship should exist between independent variables and dependent variable and (6) no outliers should exist within the data.

Multivariate analysis techniques commonly require the basic assumption of normality and continuous data. Where risk measure scales such as psychometric scales used in questionnaires, ordinal and nominal scales are usually considered unsuitable for a multivariate analysis technique. For this reason, multinomial regression is used where the mentioned basic assumptions for multivariate analysis techniques tend to be violated. Therefore, the

multinomial logistics regression is widely used as a problem tool, in the fields of medicine, psychology, mathematical finance and engineering (Bayaga, 2010).

CHAPTER 5: MODEL SPECIFICATION

5.1 Introduction

This chapter explores the forecasting models and household vehicle ownership model and recognises the suitable model and model framework. The chapter describes the fundamental mathematical framework of these models and the model structures of forecasting methods and multinomial logit modelling frameworks used for analysis in the research.

5.2 Vehicle forecast

In the vehicle forecasting section, the study uses extrapolation and trend analysis as a methodology type in analysing the data. To conduct vehicle forecasting, historical data is necessary to acquire an understanding of future developments. Adopting the extrapolation and trend analysis, assumes that the future serves as a logical extension of what has already happened in the past, and thus predictions can be made by establishing and deducing the correct trend from the available data.

This approach to forecasting may work in particular situations; however, the steering forces shaping the trends within the historical data must be studied attentively (National Research Council, 2010). Considering forces shaping trends may change, relying solely on historical data to forecast may be difficult. However, the methods used in the data analysis apply trend extrapolation which tracks changes and forces impacting trends.

To conduct vehicle forecasting, the study uses Microsoft Excel and SPSS tools. Forecasting methods for quantitative time series data include methods such as naïve method, regression trends, exponential smoothing, time series decomposition, ARIMA models, etc.

The study applies three methods for vehicle forecasting, and these include Holt's method, conducted using Microsoft Excel. This forecasting method takes into account that the data used is not stationary and may exhibit an upward or downward gradual trend over time. It imitates the forces or factors that may have a long-term impact on the time series, whether it be a constant impact or gradually as time goes by (Forecasting853, 2018). Holt's method is known as a double exponential smoothing technique, which allows two smoothing parameters namely α and β . where α is the smoothing parameter for the base level and β being the smoothing parameter for the trend.

The second method applied to conduct vehicle forecasting is the use of the forecasting sheet function found in Excel (Microsoft, 2019). Concerning this method, the forecast function predicts future values using the existing time-based data and the additive error, additive

trend and additive seasonality version of the exponential smoothing (ETS AAA) algorithm, which is a seasonal algorithm, meaning it computes a changing trend equation with seasonal adjustments. The seasonal algorithm (ETS AAA) makes use of an equation to model time series data, the algorithm in the equation used by the forecasting sheet function accounts for the additive error, additive trend and additive seasonality (Power BI Team, 2014).

The third method is the statistics tool software Statistics Package for the Social Science (SPSS). The software SPSS automatically determines the best fitting model between ARIMA and exponential smoothing to analyse the historical data inputted (SPSS, 2015).

5.2.1 Holt's method

The Holt's method is a computational method, conducted in Microsoft Excel. This model is also known as the double exponential smoothing technique, an attractive smoothing model for forecasting data with trends. Holt's method consists of three distinct equations that operate together to produce a final forecast of the data being observed (Hyndman & Athanasopoulos, 2018). The method allows for two smoothing parameters, one for the overall smoothing and the other parameter for the trend smoothing equation.

Method characteristics (ForecastIT, 2010):

- The method automatically fits a smoothing equation to the data
- Estimating a smoothing equation, which helps in minimising the probability of errors occurring between the actual data and model estimates.

Holt's method equation (Forecasting853, 2018):

$$\hat{Y}_{t+n} = E_t + nT_t$$

Where: $E_t = \alpha Y_t + (1 - \alpha)(E_{t-1} + T_{t-1})$

$$T_t = \beta(E_t - E_{t-1}) + (1 - \beta)T_{t-1}$$

With $0 \leq \alpha, \beta \leq 1$

In the above equation: T_t =denotes the trend (slope) adjustment factor at the time period t

E_t = denotes the base level at time period t

\hat{Y}_{t+n} = calculates the final forecast for the period $t + n$.

α and β = smoothing parameters, α being the smoothing parameter for the base level (E_t) and β being the smoothing parameter for the trend.

Similar to simple exponential smoothing (SES) a time series forecasting method used for data without a trend or seasonality, the base level equation expresses that E_t is the weighted

average of the observation Y_t and $E_{t-1} + T_{t-1}$) giving a one-step-ahead trending forecast for time t . The trend equation displays T_t is the weighted average of the estimated trend at time t based on $(E_t - E_{t-1})$ and T_{t-1} , the previous estimate of the trend (Hyndman & Athanasopoulos, 2018).

5.2.2 Forecasting sheet

The forecast sheet is another option available for forecasting trend series data in Microsoft Excel. Excel offers various forecasting tools, and importing data into Excel and making use of the built-in tools and formulas to forecast is a simple process.

The forecasting function used in Excel makes use of the Exponential Smoothing (ETS) method (Adam, 2018). The exponential smoothing method, similar to the moving average method is based on smoothing historical data trends. Conversely, this algorithm achieves smoothing by identifying seasonality patterns within the historical data and confidence intervals. The forecasting sheet is only available on Microsoft Excel 2016 and later.

The forecast sheet algorithm makes use of the functions “FORECAST.ETS” in the forecast output column and “FORECAST.CONFINT” function used for the interval values. These formulas are populated based on the inputs in the options section of the sheet.

The function “FORECAST.ETS” is used to conduct the exponential smoothing within the forecasting sheet. This function is similar to Holt's method mentioned above. It calculates a future value based on the additive error, additive trend and additive seasonality version of the exponential triple smoothing (ETS) algorithm, smoothing out irrelevant inconsistencies in the data trends (Cheusheva, 2019). The algorithm does this by detecting seasonality patterns if there are any, and confidence intervals.

The syntax of the Excel forecasting sheet “FORECAST.ETS” is as follows (Cheusheva, 2019):

FORECAST.ETS (*target date*, *values*, *timeline*, [*seasonality*], [*data completion*], [*aggregation*])

- **Target date** (required): this represents the data point for which to forecast a value. A date, time or a number can denote it.
- **Values** (required): a range or selection of the historical data for which a forecast for future values is required.
- **Timeline** (required): a selection or range of data with a constant step between them.
- **Seasonality** (optional): a number representing the interval of the seasonal pattern within the data.
 - 1 or omitted (default) – seasonality is automatically detected in excel, making use of whole numbers.

- 0 – Represents no seasonality within the data, therefore a linear forecast will be applicable.
- Seasonality is limited to 8 760 (number of hours in a year), anything higher will produce an output in the #NUM! error.
- **Data completion** (optional): accounts for missing points.
 - 1 or omitted (default) – fills in the missing points using the average of the neighbouring points (linear interpolation)
 - 0 – Treats the missing points as zeros.
- **Aggregation** (optional): helps stipulate how to aggregate multiple data values with the same timestamp.
 - 1 or omitted (default) – the AVERAGE function used to aggregate
 - Other options are: 2-COUNT, 3-COUNTA, 4-MAX, 5-MEDIAN, 6-MIN, and 7-SUM.

5.2.3 SPSS forecasting

SPSS is a statistics software package used for complex statistical data analysis. In the case of the data used for vehicle forecasting, the forecasting feature within SPSS is used in forecasting the time series vehicle population data (Foley, 2018). Once more, like the previous methods, using the forecasting feature on SPSS, the time series modeller proceeding approximates exponential smoothing; additionally, it estimates univariate autoregressive Integrated Moving Average (ARIMA) and multivariate ARIMA (also known as transfer function model) models for time series and produces forecasts for the vehicle population data.

The SPSS forecasting modeller attempts to automatically identify and estimate the best-fitting ARIMA or exponential smoothing model for the vehicle population data (dependent variable series).

Assumptions for the forecasting in SPSS:

- The vehicle population data (dependent variable) and years (independent variable) are time series, meaning each case represents a point in time separated by a constant time interval.
- The data used is assumed stationary.
- Independent variables do not have any missing values in the period used for estimation.

The abovementioned models and methods are applied throughout the study to achieve the objectives of the research.

5.3 Vehicle ownership

5.3.1 *Multinomial logit model*

The multinomial logit model is based on the random utility theory (Wong & Lin, 2011) and has been established to be suitable for the prediction of vehicle ownership levels (Bhat & Pulugurta, 1998). Though the multinomial logit model is widely used in vehicle ownership modelling, there are other methodologies accepted in literature. However, the multinomial logit model provides the anticipated properties for modelling the probability of a household owning a certain number of vehicles.

The dependent variable, households' access to cars/bakkies/station wagons/4x4s available for private use consists of twelve mutually exclusive categories; for this study, these categories are reduced to four categories. The choice of categories for the independent variable includes zero, one and two or more vehicles for a household. The decision-maker in this study is the household, as it is assumed that the household has the overall responsibility for the number of vehicles owned.

The multinomial logit/logistics regression model is extensively used in estimating discrete choice models (Kim, 2011). Discrete choice models are utilized in explaining a choice from a set of two or more mutually exclusive alternatives (Discrete choice analysis, 2010). In this paper, this model is developed for the probability of households owning a certain number of vehicles. Discrete choice models are based on random utility maximisation theory. This theory assumes that the preference of the household (decision maker) for an alternative is influenced by a utility, a measure of value to the household. In summary, the utility maximisation rule takes note that a household will select the alternative that maximises utility amongst the available choices. Consequently, these households within the studied data will choose to own a certain number of vehicles that bring the most value and benefits to the household (Ben-Akiva & Bierlaire, 1999).

For the vehicle ownership model, assume that utility is:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

$$\text{And } V_{nj} = \alpha_j + \beta_{nj}x_n$$

Where:

U = Utility of individual

n = Choosing alternative

$j = 0, 1$ or $2+$ (where $j=0$ for households with access to no vehicles, $j=1$ for households that own 1 vehicle, $j=2+$ for households that own 2 or more vehicles)

ε_{nj} = The random variable in the utility, which is not noticeable.

Where x_n is the vector of explanatory variables, of individual n , α_j and β_{nj} are the parameter vectors to be estimated. The probabilities that household n chooses alternative i is given by:

$$P_{ni} = \Pr (U_{ni} > U_{nj}, \forall j \neq i)$$

The multinomial logit model assumes that the random variable/error components ε_{nj} are independently and identically distributed and follows a Gumbel distribution. Given the distribution of the unnoticed components of utility, the probability that the households will choose alternative i is:

$$P_{ni} = \frac{e^{v_{ni}}}{\sum_{j \in J_n} e^{v_{nj}}}$$

Where: n = Choosing alternative

P_{ni} = The probability of the decision-maker choosing alternative i

v_{ni} = Systematic component of the utility alternative choosing alternative j

$j = 0, 1$ or $2+$ (where $j=0$ for households with access to no vehicles, $j=1$ for households that own 1 vehicle, $j=2+$ for households that own 2 or more vehicles)

A property of importance for the multinomial logit model is Independence from Irrelevant Alternatives (IIA) (Ben-Akiva & Bierlaire, 1999). This property suggests that the comparative odds between two alternatives are the same no matter what other alternatives are available.

The discrete choice model is measured by the maximum likelihood approach (IBM, 2019); using Statistics Package for Social Sciences (SPSS) for the estimation purpose.

5.4 Summary and conclusion

This chapter presented the models and methods applied in the study. The model specification process explained the independent and dependent variables of the models and methods mentioned, together with the equations of the multinomial logistics regression model. The chapter mathematically defined the relationship between the dependent and independent variables in each method and model. Non-related independent variables are excluded from the model

From the model specifications given above, it can be concluded that all the models applied in the study apply exponential smoothing of the data in the process of forecasting it. To analyse

a vehicle forecast, three methods are implemented as a means of comparison to determine the most accurate forecasted values for the number of vehicles within the forecast period of 2018 to 2038. The data used in the analysis has no seasonality present, and thus the models focus only on trends when conducting forecasts.

The household vehicle ownership section of the study applies the multinomial logistics regression as a model of prediction as the data used is categorical data, this model being most fitted for the type of data used in the study of household vehicle ownership.

CHAPTER 6: RESULTS ANALYSIS AND DISCUSSION: VEHICLE FORECAST

6.1 Introduction

In the previous chapters a base was formed for the performance of the data analysis and discussion. In the literature review chapter the background of what models and methods have been used in the forecasting of vehicle ownership in both developed and developing countries was presented. In the same chapter vehicle forecasting in the context of South Africa was also presented. This was done in order to give an overview of what has been done in South Africa concerning the forecasting of vehicle ownership.

The model specification was presented in Chapter 5, where the methodologies and procedures of forecasting were discussed in detail for both vehicle ownership and vehicle forecasting. Chapter 5 goes into the detail of the procedures involved in each model and method used in the forecasting process of the analysis. The statistics tool SPSS used in forecasting vehicle ownership using multinomial logit regression and used in vehicle forecasting was briefly introduced in the section and is further evaluated in this chapter.

The main objectives of this chapter are to present and analyse the result outputs from SPSS for vehicle ownership and vehicle forecasting. The focus is on the vehicle population in the year 2038; however, the chapter gives attention to the independent variables mentioned in previous chapters and their impact on the forecasted vehicle population numbers. Results are presented in the form of graphs and tables, discussed accordingly.

This chapter presents and discusses the results produced by applying the models. The chapter begins by presenting results for the vehicle forecasting using the three selected methods, and lastly the results and discussion for household vehicle ownership are presented. The chapter concludes by providing an overview of the results.

6.2 Vehicle forecast

The historical data used in conducting forecasts for this section was discussed in the methodology section of the study. In this vehicle forecast section the results from the Holt's method, forecasting sheet and SPSS are analysed and discussed.

6.2.1 *Holt's method results*

The Holt's method was performed using Microsoft Excel and the results of the forecast are presented in Table 6.1 below for the period 2018-2038.

The vehicle population in South Africa, as displayed in Table 6.1 below is expected to reach 16 472 593 vehicles in 2038. This represents an increase of 5 505 025 vehicles from 10 967 568 vehicles in 2017. The predicted vehicle population for the period 2018 to 2038 grows at an annual average growth rate of 1.87%.

The parameters α^5 and β^6 calculated in the Holt's method to produce the predicted vehicle population are 1 and 0.46 respectively. These parameters are optimised through the minimisation of the mean squared error (MSE), in the process of producing the forecasts, by changing the initial α and β .

Table 6.1: Forecasted vehicle population for period 2018-2038 using Holt's method.

Year	Predicted Vehicle population
2018	11,272,354
2019	11,532,363
2020	11,792,373
2021	12,052,382
2022	12,312,391
2023	12,572,400
2024	12,832,410
2025	13,092,419
2026	13,352,428
2027	13,612,437
2028	13,872,447
2029	14,132,456
2030	14,392,465
2031	14,652,474
2032	14,912,484
2033	15,172,493
2034	15,432,502
2035	15,692,511
2036	15,952,521
2037	16,212,530
2038	16,472,539

The Holt's method parameters are summarised in Table 6.2 below. The smoothing constants of α and β determine the sensitivity of the forecasts to the changes in demand for vehicles

⁵ Alpha, also known as the base value. The value of alpha determines the weighting of the past data values in setting the baseline for the forecast. A high alpha leads to increased weight given to recent observations and a low alpha means more uniform weighting.

⁶ Beta, also known as the trend value. It determines the degree of change in recent data trends and should be valued compared to older trends when forecasting.

over the forecasted period. It is important to note that parameters have a substantial impact on the quality of any forecast.

The optimal alpha for the vehicle population using Holt’s method is 1, which indicates that the forecasted vehicle population is more responsive to more recent historical vehicle population data, and vehicle demand is changing rapidly. The optimal value of β 0.46 assumes that the trend changes at an average rate over time, and this β value deems that the future is uncertain because the errors in a trend estimation become more notable with the forecasting of more than one period ahead. This model is attempting to estimate a short-term trend.

Table 6.2: Vehicle population Holt's method parameters

Alpha	1
Beta	0.46

In Figure 6.1 below the actual vehicle population and forecasted vehicle population are plotted together with trend lines. It is clear that the predicted vehicle populations display the same trend as the actual vehicle populations, as seen in the graph below where the lines of actual vehicles and predicted vehicles in the initial period 1986 to 2017 move collectively. The trend lines for both graphs are moving in an upwards (positive) direction. This trend analysis plot illustrates that the model (the Holt’s method) implemented in this data fits closely with the actual data.

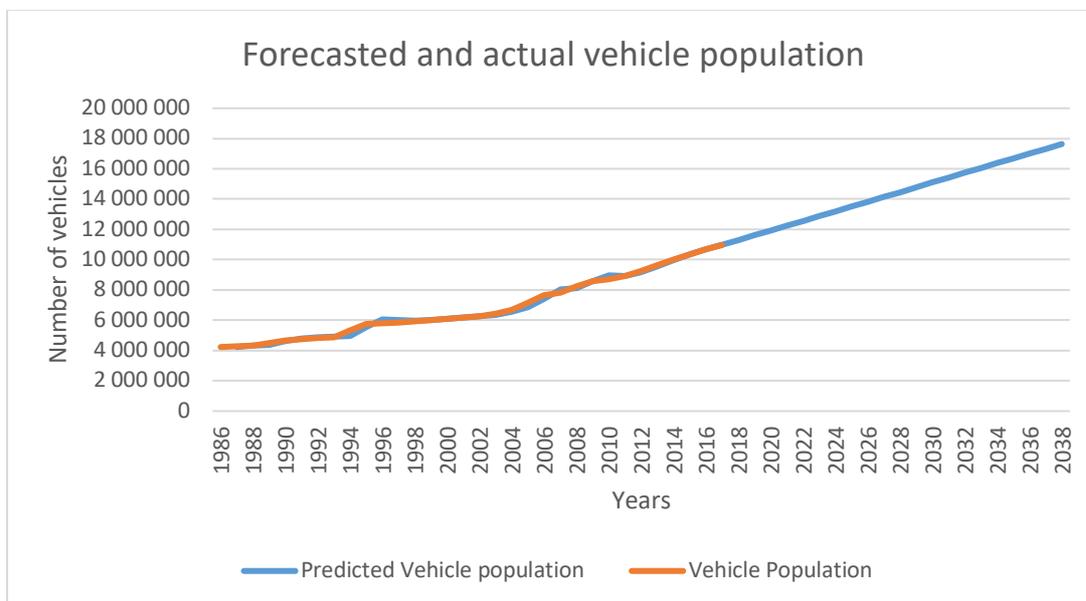


Figure 6.1: Holt's forecasted and actual vehicle population

In Table 6.3 below the prediction results for South Africa’s real GDP are presented in local currency. The forecasted results extracted from the Holt’s method show that the real GDP for

South Africa in local currency is approximated to reach 3 931 500 019 539.18 in 2038. This real GDP is only an approximated value, as factors that affect real GDP are not taken into consideration in the analysis, and the method only uses the trend detected in historical data to generate the predicted values. These predicted real GDPs for the period 2018 to 2038 grow at an annual average growth rate of 1.09%.

Table 6.3: Real GDP forecasted values for 2018-2038 using Holt's method

Year	Predicted Real GDP
2018	3 163 199 524 739.91
2019	3 201 614 549 479.83
2020	3 240 029 574 219.74
2021	3 278 444 598 959.65
2022	3 316 859 623 699.57
2023	3 355 274 648 439.48
2024	3 393 689 673 179.40
2025	3 432 104 697 919.31
2026	3 470 519 722 659.22
2027	3 508 934 747 399.14
2028	3 547 349 772 139.05
2029	3 585 764 796 878.96
2030	3 624 179 821 618.88
2031	3 662 594 846 358.79
2032	3 701 009 871 098.70
2033	3 739 424 895 838.62
2034	3 777 839 920 578.53
2035	3 816 254 945 318.44
2036	3 854 669 970 058.36
2037	3 893 084 994 798.27
2038	3 931 500 019 538.18

These parameters α and β calculated in Table 6.4 below are still produced through the Holt's method of forecasting. Parameters are optimised through the minimisation of the mean squared error (MSE) in the process of producing the forecasts, by changing the initial α and β .

The optimal alpha for the South African real GDP in local currency using Holt's method is 1, which shows that the forecasted real GDPs are more responsive to more recent historical data for real GDP when estimating final values. This alpha also indicates that the real GDP in South Africa changes quickly in response to the effects of past economical events. The optimal value of β 0.45 assumes that the trend changes at an average rate over time, as is the trend in the vehicle population above. This β value estimates that the future is uncertain because the errors in a trend estimation become more notable with the forecasting of more than one period ahead. This model is attempting to estimate a short-term trend.

Table 6.4: Real GDP Holt's parameters

Alpha	1
Beta	0.45

In Figure 6.2 below the actual real GDP and forecasted real GDP are plotted together with trend lines. Observing the trend lines, it is evident that these two graphs follow the same trend. As seen in the graph below, the line graphs of actual real GDP and predicted real GDP in the historical period 1986 to 2017 move collectively. The trend lines for both graphs display an upward positive trend within the historical and predicted values. This trend analysis plot illustrates that the Holt's model fits closely with the actual data.

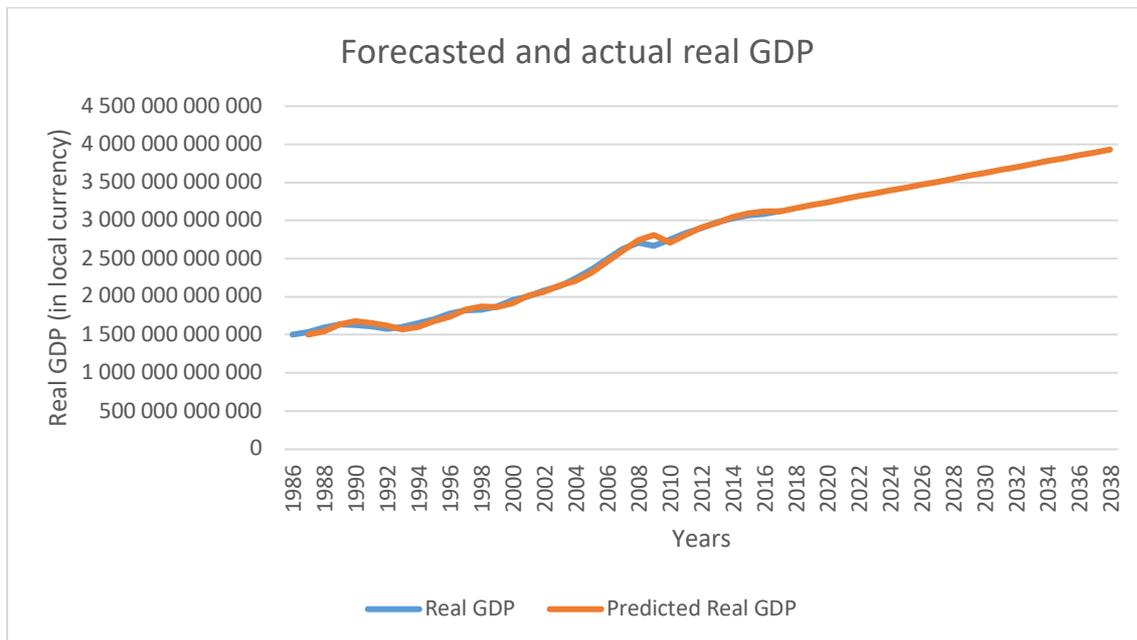
**Figure 6.2: Holt's forecasted and actual real GDP**

Table 6.5 below presents the forecasted values using Holt's method for the South African population for the forecast period 2018 to 2038. The forecast approximates the South African population to reach 71 452 499 people by 2038. This prediction does not include any factors that may affect the number of people in the future, such as war, epidemics, natural calamities or migrations. These predicted population values for the period 2018 to 2038 show an annual average growth rate of 1.09%. This average annual growth rate is the same as that of real GDP, thus, it can be concluded from this observation that South African real GDP and population increase at the same average annual proportion for the forecast period, expected to have a similar trend.

Table 6.5: South African population forecasted values for 2018-2038 using Holt's method

Year	Predicted South African Population
2018	57,418,839
2019	58,120,522
2020	58,822,205
2021	59,523,888
2022	60,225,571
2023	60,927,254
2024	61,628,937
2025	62,330,620
2026	63,032,303
2027	63,733,986
2028	64,435,669
2029	65,137,352
2030	65,839,035
2031	66,540,718
2032	67,242,401
2033	67,944,084
2034	68,645,767
2035	69,347,450
2036	70,049,133
2037	70,750,816
2038	71,452,499

The optimal parameter alpha for the South African population using Holt's method presented in Table 6.6 is 1, which shows that the forecasted population is more influenced by more recent historical data for the South African population when estimating final values. The optimal value of β is 1 in the case of South Africa's population; this assumes that the trend changes rapidly from one period to the next, and this β value estimates that the future is certain because the errors in a trend estimation become less notable with the forecasting of more than one period ahead. This model is attempting to estimate a long-term trend.

Table 6.6: South African population Holt's parameters

Alpha	1
Beta	1

In Figure 6.3 below, the actual South African population and forecasted population plotted together with associated trend lines are presented. Observing the trend lines, it is evident that these two graphs follow the same trend. As seen in the graph below the line graphs of actual population and predicted population in the historical period 1986 to 2017 move together. The trend lines for both graphs show an upward positive trend within the historical and predicted

values. The trend analysis plot illustrates that the model (Holt’s method) applied in this data fits closely with the actual data used to forecast.

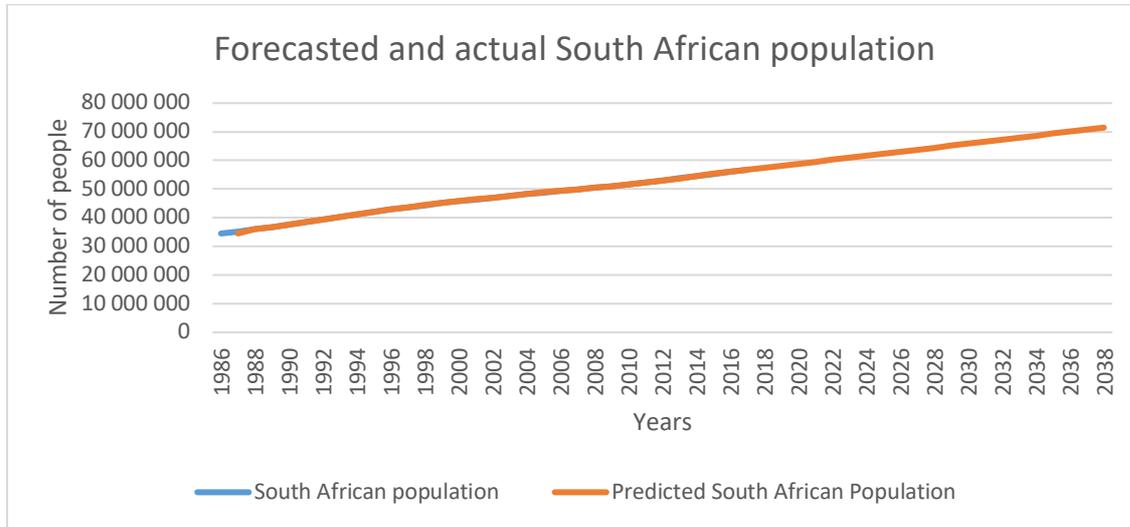


Figure 6.3: Holt's forecasted and actual South African population

6.2.2 Forecast sheet results

In this sub-section of vehicle ownership forecasts, the results from the Excel tool forecast sheet are presented. The purpose of the utilisation of this tool is to compare the outcomes to the previous forecasting method applied and the SPSS statistics tool results to follow. The comparison is done for a more accurate forecast of South Africa’s vehicle population, human population and real GDP.

Table 6.7 below presents the results of the vehicle population forecast for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. These confidence bounds define the range between which the forecasted vehicle population value is expected to lie; they represent the limits of the range. The vehicle population forecasts for each year within the forecast period are shown in Table 6.7. It can be observed from the table that by the year 2038 South Africa’s vehicle population according to the forecast sheet tool is predicted to be 15 516 278.

This number of vehicles are predicted to increase at an average annual growth rate of 1.62% in the period 2018 to 2038. These values are forecasted with no other influencing variable included.

Table 6.7: Vehicle population forecasted values and confidence bounds for 2018-2038 using Forecast sheet

Years	Forecast	Lower Confidence Bound	Upper Confidence Bound
2018	11,237,200	10,932,720	11,541,680
2019	11,451,154	10,902,499	11,999,809
2020	11,665,108	10,787,884	12,542,332
2021	11,879,062	10,613,272	13,144,852
2022	12,093,016	10,389,153	13,796,879
2023	12,306,970	10,121,492	14,492,448
2024	12,520,924	9,814,314	15,227,534
2025	12,734,877	9,470,611	15,999,144
2026	12,948,831	9,092,741	16,804,922
2027	13,162,785	8,682,635	17,642,935
2028	13,376,739	8,241,918	18,511,560
2029	13,590,693	7,771,984	19,409,402
2030	13,804,647	7,274,047	20,335,247
2031	14,018,601	6,749,179	21,288,022
2032	14,232,555	6,198,338	22,266,772
2033	14,446,509	5,622,381	23,270,636
2034	14,660,463	5,022,089	24,298,836
2035	14,874,416	4,398,171	25,350,662
2036	15,088,370	3,751,277	26,425,464
2037	15,302,324	3,082,007	27,522,641
2038	15,516,278	2,390,916	28,641,640

The forecasting sheet also produces the parameters alpha and beta, additionally it gives a parameter gamma, and these parameters are displayed in Table 6.8. The alpha for the South African vehicle population for the forecasting sheet tool outputs is 0.75, which shows that the forecasted vehicle population is impacted more by the recently observed historical data for South Africa's vehicle population when estimating final values. The optimal value of β in this method is 0.75; this assumes that the trend changes at an average rate from one period to the next, and this β value estimates that the future of the vehicle population is not certain as the errors in a trend estimation become notable with the forecast of more than one period ahead. This model is attempting to estimate a short-term trend.

In the forecast sheet is the gamma parameter, which is the seasonal component in the forecast. The gamma for the vehicle population is 0, thus meaning there is no seasonality within the vehicle population data given.

Table 6.8: Vehicle population forecast sheet parameters

Alpha	0.75
Beta	0.75
Gamma	0.00

Figure 6.4 displays the actual South African vehicle population and forecasted population plotted together with associated upper and lower confidence bound lines present. The forecast sheet tool applies exponential smoothing in the generating of a forecast, thus the graph below displays exponential trend lines. Observing the exponential trend lines, it is evident that these

two graphs follow the same trend. As seen in the graph below, the trend lines of actual vehicle population and predicted vehicle population in the historical period 1986 to 2017 move collectively. The trend lines for both graphs show a steady increasing positive trend within the historical and predicted values, the growth for 2037 and 2038 is predicted to be slower than that of prior years, this is shown by the flatter looking end of the forecast line. The trend analysis plot illustrates that the forecast sheet method applied in this data fits closely with the actual data used to forecast.

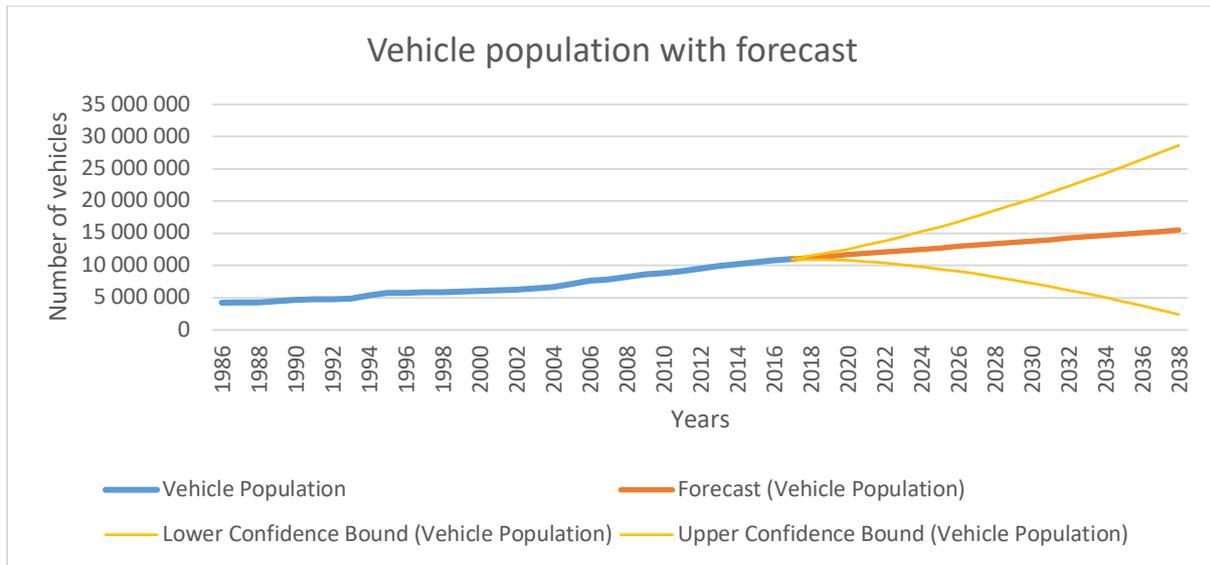


Figure 6.4: Forecast sheet forecasted and actual vehicle population, with confidence bounds

Table 6.9 presents the results of the South African real GDP for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. The real GDP forecasts for each year within the forecast period are shown in Table 6.9 below, and it can be observed from this table that, by the year 2038 real GDP according to the forecast sheet tool is predicted to be 4 356 309 783 580.11 in the local currency.

These real GDP values are predicted to increase at an average annual growth of 1.57% in the period 2018 to 2038.

These forecasted values are forecasted with no other influencing variable included.

Table 6.9: Real GDP forecasted values and confidence bounds for 2018-2038 using forecast sheet

Year	Forecast (Real GDP)	Lower Confidence Bound	Upper Confidence Bound
2018	3,185,550,194,946.36	3,098,792,589,217.25	3,272,307,800,675.46
2019	3,244,088,174,378.04	3,127,309,697,724.18	3,360,866,651,031.91
2020	3,302,626,153,809.73	3,162,051,924,180.44	3,443,200,383,439.02
2021	3,361,164,133,241.42	3,200,233,708,882.25	3,522,094,557,600.59
2022	3,419,702,112,673.11	3,240,677,840,550.51	3,598,726,384,795.71
2023	3,478,240,092,104.80	3,282,755,067,963.71	3,673,725,116,245.89
2024	3,536,778,071,536.48	3,326,082,246,942.27	3,747,473,896,130.70
2025	3,595,316,050,968.17	3,370,405,596,393.98	3,820,226,505,542.36
2026	3,653,854,030,399.86	3,415,546,758,284.19	3,892,161,302,515.52
2027	3,712,392,009,831.55	3,461,374,728,246.16	3,963,409,291,416.94
2028	3,770,929,989,263.23	3,507,789,937,647.25	4,034,070,040,879.22
2029	3,829,467,968,694.92	3,554,714,616,892.24	4,104,221,320,497.60
2030	3,888,005,948,126.61	3,602,086,652,878.26	4,173,925,243,374.96
2031	3,946,543,927,558.30	3,649,855,506,780.30	4,243,232,348,336.30
2032	4,005,081,906,989.99	3,697,979,405,753.24	4,312,184,408,226.74
2033	4,063,619,886,421.67	3,746,423,354,007.70	4,380,816,418,835.65
2034	4,122,157,865,853.36	3,795,157,688,779.64	4,449,158,042,927.09
2035	4,180,695,845,285.05	3,844,157,009,172.44	4,517,234,681,397.66
2036	4,239,233,824,716.74	3,893,399,366,562.89	4,585,068,282,870.58
2037	4,297,771,804,148.43	3,942,865,642,517.48	4,652,677,965,779.37
2038	4,356,309,783,580.11	3,992,539,063,734.59	4,720,080,503,425.64

The parameters alpha, beta and gamma for the South African real GDP are displayed in Table 6.10. The alpha for the South African real GDP for the forecasting sheet tool outputs is 0.9, which shows that the forecasted real GDP is influenced more by the recent historical data observed for the South African vehicle population when estimating final values. The optimal value of β in this method is 0, and this assumes that the trend does not change from one period to the next. This β value estimates that the future of the real GDP is uncertain as the errors in the trend estimation become extremely notable with the forecasting of more than one period ahead. This model is attempting to estimate a short-term trend.

In the forecast sheet, the gamma parameter is the seasonal component in the forecast. The gamma γ for the real GDP is 0, thus meaning there is no seasonality within the real GDP data given.

Table 6.10: Real GDP forecast sheet parameters

Alpha	0.90
Beta	0.00
Gamma	0.00

In Figure 6.5 the actual South African real GDP and forecasted real GDP are plotted together, with associated upper and lower confidence bound lines. Observing the exponential trend lines, it is evident that these two graphs follow the same trend, as seen in the graph below where the trend lines of actual vehicle population and predicted vehicle population in the

historical period 1986 to 2017 move collectively. However, the rate at which each trend line moves is different. The trend lines for both graphs show an upward positive trend within the historical and predicted values, but the historical real GDP trend line increases at a higher rate and lies above the confidence bound. The trend analysis plot illustrates that the forecast sheet method applied in this data closely fits the data used to predict, as it still lies within forecast confidence bounds and moves in the same direction as the historical data trend line.

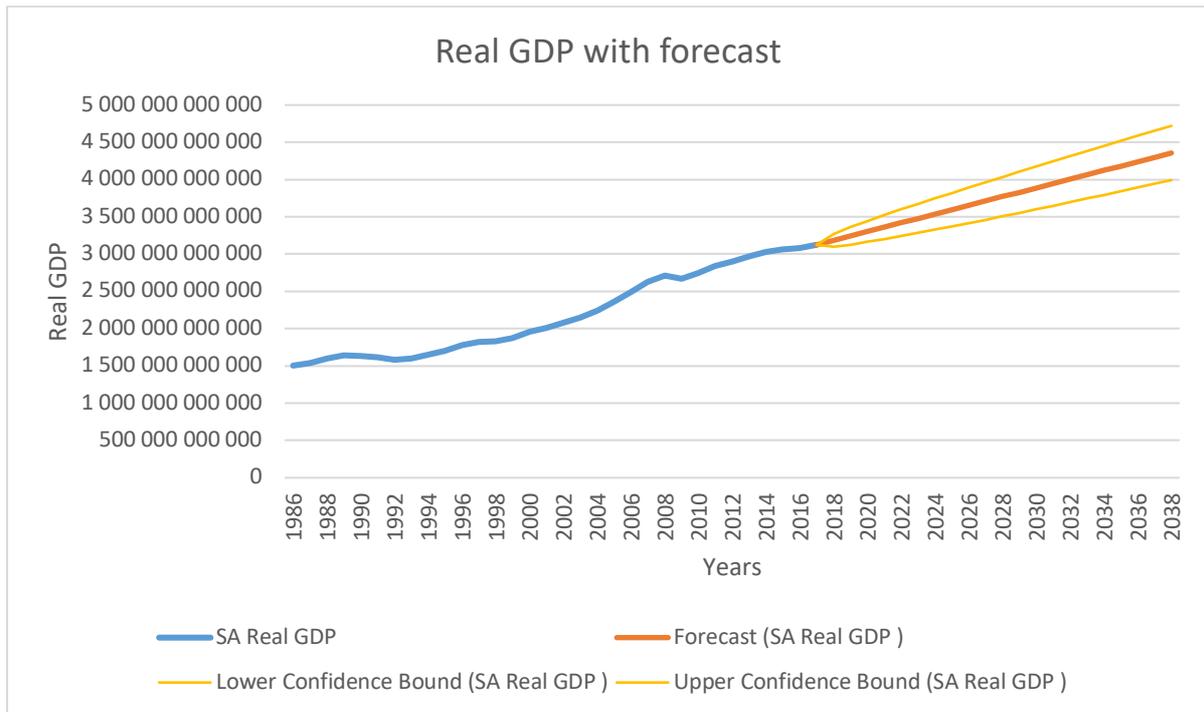


Figure 6.5: Forecast sheet forecasted and actual real GDP, with confidence bounds

Table 6.11 below presents the results of the South African population forecast for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. The population forecasts for each year within the forecast period are shown in Table 6.11 below, and as can be observed from the given table,, by the year 2038 South Africa’s human population according to the forecast sheet tool is predicted to be 71 213 365 people.

The population is predicted to increase at an average annual growth rate of 1.08% in the period 2018 to 2038. The forecasted population numbers are forecasted with no other influencing variable included.

Table 6.11: South African population forecasted values and confidence bounds for 2018-2038 using forecast sheet

Years	Forecast (South African Population)	Lower Confidence Bound	Upper Confidence Bound
2018	57,407,427	57,196,150.16	57,618,704.71
2019	58,097,724	57,784,106.78	58,411,341.85
2020	58,788,021	58,385,271.90	59,190,770.49
2021	59,478,318	58,991,423.65	59,965,212.50
2022	60,168,615	59,599,573.21	60,737,656.71
2023	60,858,912	60,208,289.83	61,509,533.84
2024	61,549,209	60,816,786.29	62,281,631.15
2025	62,239,506	61,424,591.88	63,054,419.32
2026	62,929,802	62,031,409.88	63,828,195.08
2027	63,620,099	62,637,046.83	64,603,151.90
2028	64,310,396	63,241,374.06	65,379,418.43
2029	65,000,693	63,844,305.35	66,157,080.90
2030	65,690,990	64,445,783.17	66,936,196.85
2031	66,381,287	65,045,769.92	67,716,803.85
2032	67,071,584	65,644,242.13	68,498,925.41
2033	67,761,881	66,241,186.45	69,282,574.85
2034	68,452,178	66,836,596.94	70,067,758.13
2035	69,142,474	67,430,473.08	70,854,475.74
2036	69,832,771	68,022,818.38	71,642,724.20
2037	70,523,068	68,613,639.31	72,432,497.04
2038	71,213,365	69,202,944.51	73,223,785.60

The parameters alpha and beta and gamma for the South African population are displayed in Table 6.12 below. The alpha for the South African population extracted from the forecasting sheet tool outputs is 1, and this is the highest value alpha can take. This population alpha shows that the forecasted population is altered more by the recently observed historical data for the South African population when estimating final values in the forecast period. The optimal value of β in this method is 0.1, and this assumes that the trend does not show a significant change from one period to the next. This β value estimates that the future of the population is uncertain, as the errors in the trend of the forecasted values become extremely notable with the forecast of more than one period ahead. This beta value concludes that the model is attempting to estimate a short-term trend.

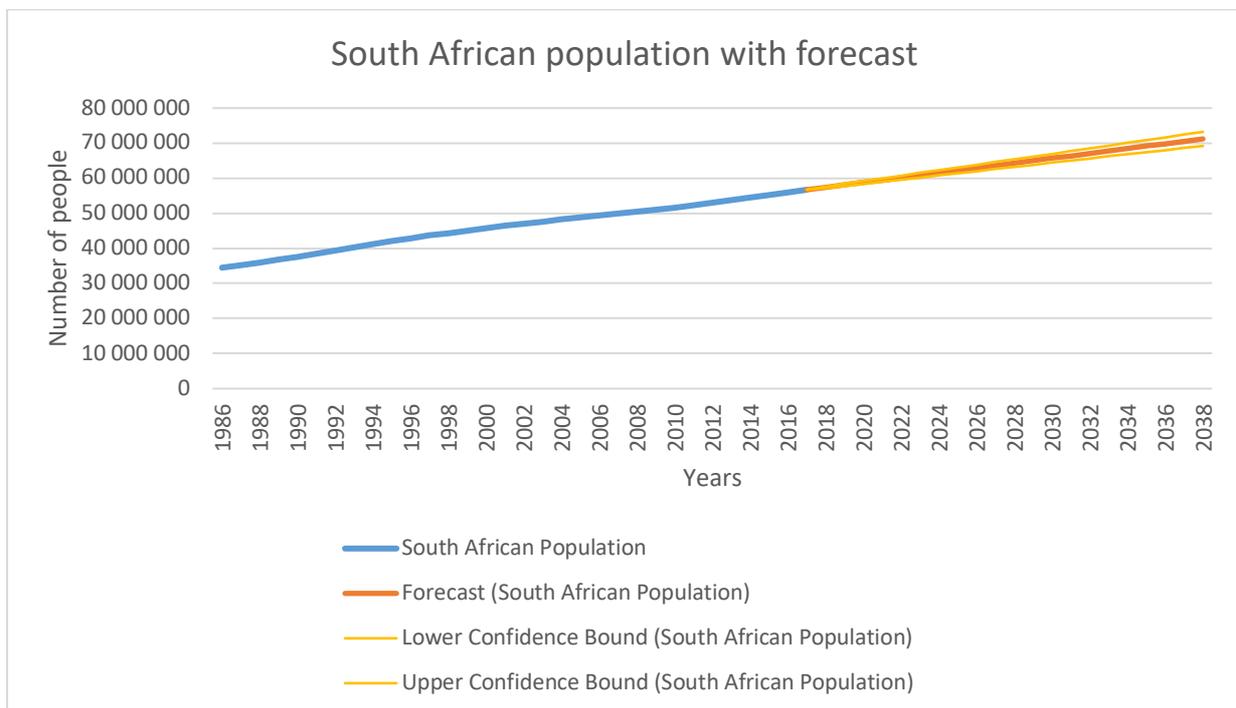
In the forecast sheet, the gamma parameter is the seasonal component in the forecast. The gamma in the result for the population is 0, thus meaning there is no seasonality within the population data given.

Table 6.12: South African population forecast sheet parameters

Alpha	1.00
Beta	0.10
Gamma	0.00

In Figure 6.6 below, the historical South African population and forecasted population are plotted together with associated upper and lower confidence bound lines, which from the graph are shown to not have much of a difference between the forecasted and the bound values. Observing the exponential trend lines, it is evident that these two graphs follow the same trend, as seen in the graph below the trend lines of historical population and predicted population in the historical period 1986 to 2017 move together (both are increasing).

However, the rate at which the trend lines move differs. The trend lines for both graphs show an upward positive trend within the historical and predicted values, but the historical population trend line increases at a higher rate and lies above the confidence bound. The trend analysis plot illustrates that the forecast sheet method applied in this data closely fits the data used to predict forecasted values, as the forecasted population trend line still lies within forecast confidence bounds and moves in the same direction as the historical data trend line.

**Figure 6.6: Forecast sheet forecasted and actual South African population, with confidence bounds**

6.2.3 SPSS Results

This sub-section is similar to those above, and provides the forecasted results of the components involved in SPSS forecasting tool. SPSS automatically matches the data to the most appropriate model when analysing the data for forecasting.

Table 6.13 below presents the results of the vehicle population forecast for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. The vehicle population forecasts for each year within the forecast period are shown in the table, where it can be observed that by the year 2038 the vehicle population according to the SPSS forecasting function is predicted to be 16 013 065.

The number of vehicles is predicted to increase at an average annual growth rate of 1.75% in the period 2018 to 2038. These values are forecasted with no other influencing variable included.

Table 6.13: Vehicle population forecasted values and confidence levels for 2018-2038 using SPSS

Year	Vehicle Population-Model_1		
	Forecast	UCL	LCL
2018	11,251,881	11,537,103	10,966,658
2019	11,489,940	12,034,463	10,945,417
2020	11,727,999	12,577,806	10,878,192
2021	11,966,058	13,161,666	10,770,451
2022	12,204,118	13,781,949	10,626,286
2023	12,442,177	14,435,547	10,448,807
2024	12,680,236	15,120,015	10,240,458
2025	12,918,295	15,833,371	10,003,220
2026	13,156,354	16,573,970	9,738,739
2027	13,394,414	17,340,415	9,448,412
2028	13,632,473	18,131,506	9,133,439
2029	13,870,532	18,946,193	8,794,871
2030	14,108,591	19,783,551	8,433,632
2031	14,346,651	20,642,754	8,050,547
2032	14,584,710	21,523,061	7,646,359
2033	14,822,769	22,423,801	7,221,737
2034	15,060,828	23,344,364	6,777,292
2035	15,298,887	24,284,191	6,313,584
2036	15,536,947	25,242,767	5,831,126
2037	15,775,006	26,219,617	5,330,395
2038	16,013,065	27,214,300	4,811,829

The exponential smoothing model parameters for the vehicle population are presented below in Table 6.14. SPSS fitted the vehicle population data to a Brown forecasting model, this model

uses two different smoothed series centred at different points in time, similar to the Holt's method.

In table 6.14 the estimate of the alpha (level and trend) is given as 0.813. This population alpha level shows that the forecasted vehicles are altered more by the recently observed historical data for South African population when estimating final values in the forecast period. The alpha trend is also 0.813 and assumes that the trend shows a rapid change from one period to the next. This trend estimates that the future of vehicles is certain, as the errors in the trend of the forecasted values become less notable with the forecasting of more than one period ahead. This trend value concludes that the model is attempting to estimate a long-term trend for the South African vehicle forecasts.

All of the model's parameter shows to be statistically significant at the level 5%, p-value (sig) < 0.05.

Table 6.14: Vehicle population SPSS parameters

Exponential Smoothing Model Parameters						
Model			Estimate	SE	t	Sig.
Vehicle Population-Model_1	No Transformation	Alpha (Level and Trend)	0.813	0.088	9.259	0.000

In Figure 6.7 below, the historical South African vehicle population and forecasted vehicle population are plotted together with associated upper and lower confidence bound lines. Observing the forecast fit line, it is evident that these two graphs follow the same trend. As seen in the graph below the fit line and the forecast line move jointly. The trend analysis plot illustrates that the SPSS forecasting method applied in this data closely fits the data used to predict forecasted values, as the forecasted vehicle population lies within the confidence bounds.

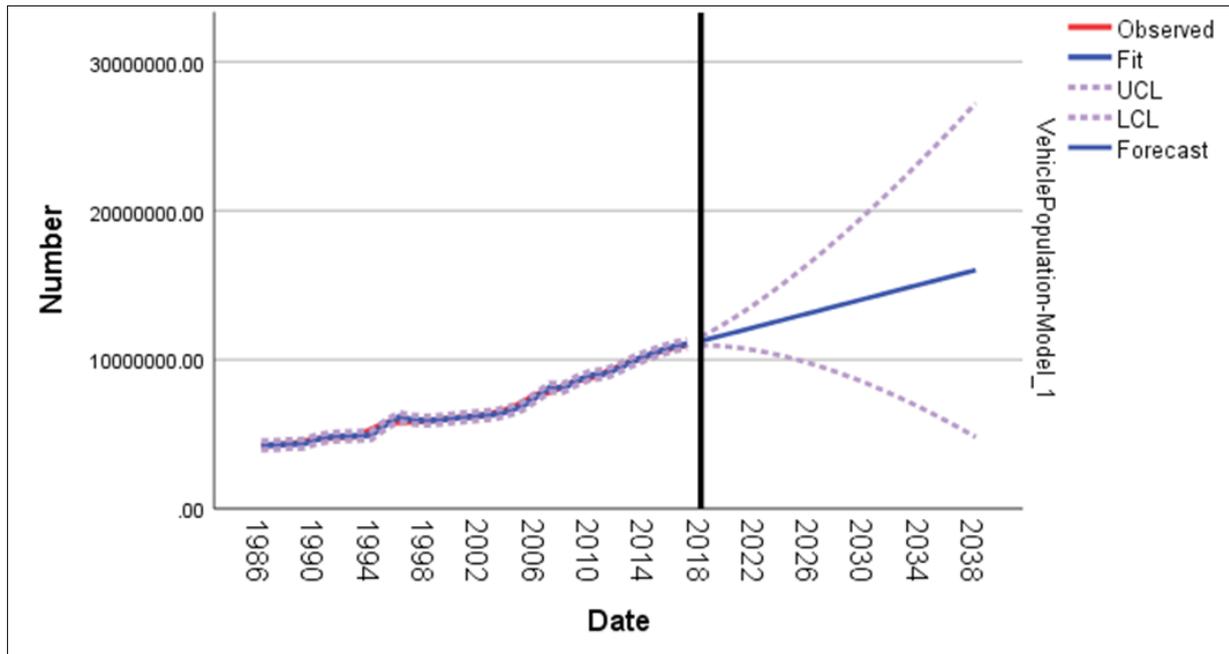


Figure 6.7: SPSS forecasted and actual South African population, with confidence level and fit

Table 6.15 below presents the results of the South African real GDP in local currency forecast for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. It can be observed from the given table that, by the year 2038 real GDP in local currency according to the SPSS forecasting function is predicted to be 3 883 705 362 363,34.

This value of real GDP in local currency is calculated at an average annual growth of 1.03% in the period 2018 to 2038. This forecasted values are forecasted with no other influencing variable included.

Table 6.15: Real GDP forecasted values and confidence levels for 2018-2038 using SPSS

Forecast			
SA Real GDP -Model_1			
Model	Forecast	UCL	LCL
2018	3,160,662,338,397.51	3,242,658,306,639.96	3,078,666,370,155.06
2019	3,196,814,489,595.80	3,361,737,628,493.47	3,031,891,350,698.14
2020	3,232,966,640,794.09	3,496,482,642,972.96	2,969,450,638,615.22
2021	3,269,118,791,992.38	3,644,940,131,154.98	2,893,297,452,829.79
2022	3,305,270,943,190.68	3,805,668,628,658.22	2,804,873,257,723.13
2023	3,341,423,094,388.97	3,977,586,903,905.66	2,705,259,284,872.28
2024	3,377,575,245,587.26	4,159,851,757,380.26	2,595,298,733,794.26
2025	3,413,727,396,785.55	4,351,783,633,849.46	2,475,671,159,721.65
2026	3,449,879,547,983.84	4,552,820,388,053.58	2,346,938,707,914.11
2027	3,486,031,699,182.13	4,762,487,234,828.11	2,209,576,163,536.15
2028	3,522,183,850,380.43	4,980,376,372,799.06	2,063,991,327,961.79
2029	3,558,336,001,578.72	5,206,132,653,396.92	1,910,539,349,760.51
2030	3,594,488,152,777.01	5,439,443,184,635.19	1,749,533,120,918.83
2031	3,630,640,303,975.30	5,680,029,588,402.42	1,581,251,019,548.18
2032	3,666,792,455,173.59	5,927,642,102,886.04	1,405,942,807,461.14
2033	3,702,944,606,371.88	6,182,055,002,475.64	1,223,834,210,268.12
2034	3,739,096,757,570.17	6,443,062,980,330.71	1,035,130,534,809.63
2035	3,775,248,908,768.47	6,710,478,248,704.71	840,019,568,832.23
2036	3,811,401,059,966.76	6,984,128,184,039.70	638,673,935,893.82
2037	3,847,553,211,165.05	7,263,853,392,125.04	431,253,030,205.06
2038	3,883,705,362,363.34	7,549,506,101,765.36	217,904,622,961.32

The exponential smoothing model parameters for South African real GDP is presented below in Table 6.16. SPSS fitted the real GDP data to a Brown's linear exponential smoothing model, which is a simple trend model and makes use of two smoothing series positioned at varying points in the forecast period. The forecasting formula used in the generation of the forecast values is based on an extrapolation of a line through the two positions. The Brown's model is similar to the Holt's model.

In the table the estimate of the alpha (level and trend) is given as 1. This population alpha level shows that the forecasted population is altered more by the recently observed historical data for South African population when estimating final values in the forecast period. The alpha trend is also 1 in the given table, and this assumes that the trend shows a rapid change from one period to the next. This trend estimates that the future of the population is certain, as the errors in the trend of the forecasted values become less notable with the forecasting of more than one period ahead. This trend value concludes that the model is attempting to estimate a long-term trend for the South African population forecasts.

All of the model's parameter shows to be statistically significant at the level 5%, p-value (sig) < 0.05.

Table 6.16: Real GDP SPSS parameters

Exponential Smoothing Model Parameters						
Model			Estimate	SE	t	Sig.
SA Real GDP Model_1	No	Alpha (Level Transformat and Trend) ion	0.873	0.087	10.035	0.000

In Figure 6.8 below, the historical South African real GDP and forecasted real GDP are plotted together with associated upper and lower confidence bound lines. Observing the forecast fit line, it is evident that these two graphs follow the same trend, as seen in the graph below where the fit line and the forecast line move jointly. The trend analysis plot illustrates that the SPSS forecasting method applied in this data closely fits the data used to predict forecasted values, as the forecasted vehicle population lies within the confidence bounds.

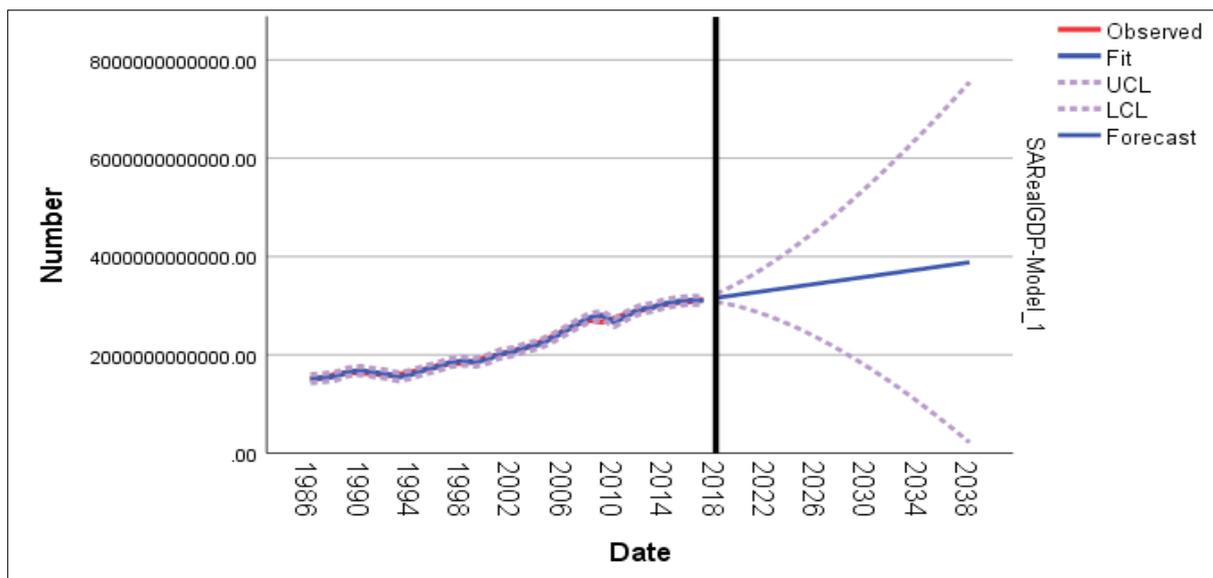


Figure 6.8: SPSS forecasted and actual real GDP, with confidence levels and fit

Table 6.17 below presents the results of the South African population forecast for the stated forecast period of the study, together with the lower and upper confidence bounds for each forecasted year. It can be observed from the given table that by the year 2038 South Africa's population according to the SPSS forecasting function is predicted to be 71 452 500 people.

This value of the South African population is calculated at an average annual growth of 1.09% in the period 2018 to 2038. This forecasted values are forecasted with no other influencing variable included.

Table 6.17: South African population forecasted values and confidence levels for 2018-2038 using SPSS

Forecast			
Model	South African Population-Model_1		
	Forecast	UCL	LCL
2018	57418839	57490931.64	57346746.46
2019	58120522	58281725.89	57959318.31
2020	58822205	59091950.59	58552459.71
2021	59523888	59918755.00	59129021.39
2022	60225571	60760223.36	59690919.13
2023	60927254	61614972.63	60239535.96
2024	61628937	62481946.89	60775927.79
2025	62330620	63360306.70	61300934.09
2026	63032303	64249364.22	61815242.66
2027	63733986	65148542.63	62319430.35
2028	64435670	66057349.20	62813989.88
2029	65137353	66975356.70	63299348.48
2030	65839036	67902190.10	63775881.18
2031	66540719	68837516.71	64243920.66
2032	67242402	69781038.83	64703764.64
2033	67944085	70732487.97	65155681.60
2034	68645768	71691620.41	65599915.26
2035	69347451	72658213.58	66036688.18
2036	70049134	73632063.21	66466204.65
2037	70750817	74612980.92	66888653.04
2038	71452500	75600792.26	67304207.80

The exponential smoothing model parameters for the South African population are presented below in Table 6.18. SPSS fitted the population data to a Brown's linear exponential smoothing model as with the real GDP above.

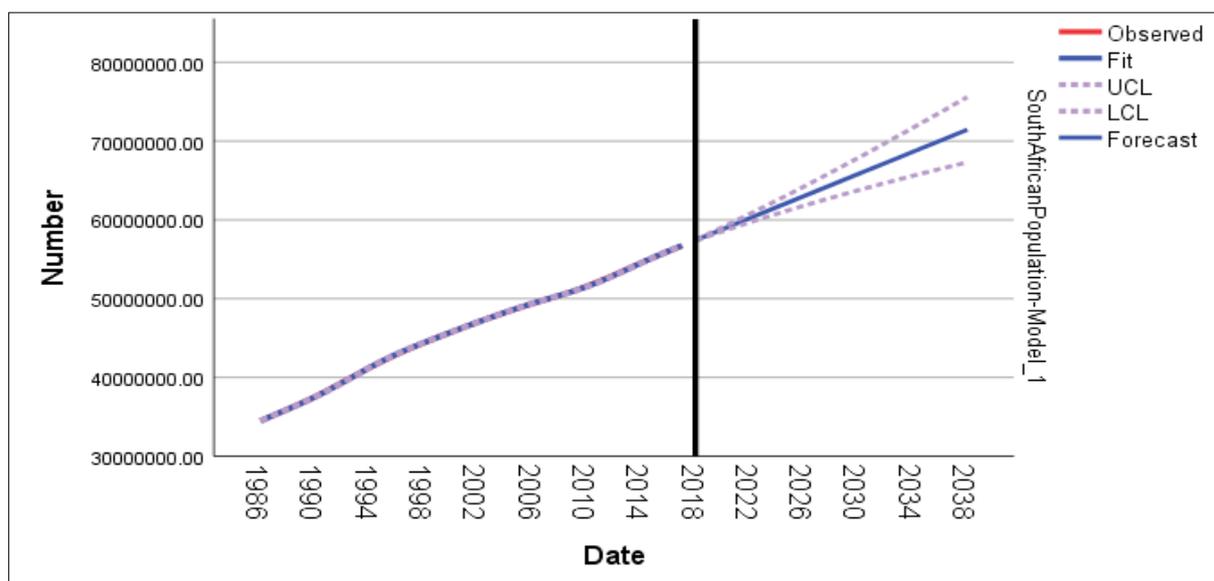
In the table the estimate of the alpha (level and trend) is given as 1, and this population alpha level shows that the forecasted population is altered more by the recently observed historical data for South African population when estimating final values in the forecast period. The alpha trend is also 1 in the given table, this assumes that the trend shows a rapid change from one period to the next. This trend estimates that the future of the population is certain, as the errors in the trend of the forecasted values become less notable with the forecasting of more than one period ahead. This trend value concludes that the model is attempting to estimate a long-term trend for the South African population forecasts.

All of the model's parameters are shown to be statistically significant at the level 5%, p-value (sig) < 0.05.

Table 6.18: South African population SPSS parameters

Exponential Smoothing Model Parameters						
Model			Estimate	SE	t	Sig.
South African Population-Model_1	No Transformat ion	Alpha (Level and Trend)	1.000	0.017	60.205	0.000

In Figure 6.9 below, the historical South African population and forecasted population are plotted together with associated upper and lower confidence bound lines. Observing the forecast fit line, it is evident that these two graphs follow the same trend. As seen in the graph below the fit line and the forecast line move jointly. The trend analysis plot illustrates that the SPSS forecasting method applied in this data closely fits the data used to predict forecasted values, as the forecasted vehicle population lies within the confidence bounds.

**Figure 6.9: SPSS forecasted and actual South African population, with confidence levels and fit**

6.2.4 Comparison and conclusion of forecast models and methods for vehicle forecasting

The historical data used in each forecasting method or model is the same; however the forecast results from each model or method applied differ, though the difference is not notable in some instances. The different forecast results are to be expected, as each model applies different formulas to smooth out the historical data in producing the forecasts. Below, Table 6.19 compares the outcomes of each model for each variable forecasted.

Evaluating the vehicle population forecast results from all three of the applied methods, namely Holt's method, the forecasting sheet tool and the SPSS statistics software, it can be observed that the Holt's method and SPSS results are approximately similar, with negligible

differences. Analysing the graphical representation of these forecasted values, it is noted that the graphs move in similar and close pattern. The average growth rates of the methods have a small difference, with the Holt's method growing vehicle population at an average annual growth rate of 1.68%, forecast sheet at 1.62% and SPSS at 1.75%. Thus, the forecast sheet line is forecasted below the two graphs forecasting a lower vehicle population. The forecast sheet has under forecasted in comparison to the other two methods.

It can be concluded that the vehicle population in South Africa approximated at an average of 16 000 627 vehicles in 2038. Averaged from the predicted vehicles in 2038 for the three methods.

Table 6.19: Holt's method vs forecast sheet vs SPSS forecasted vehicle population

Vehicle population forecast			
Years	Holt's method	Forecast	SPSS
2018	11 272 354	11 237 200	11 251 881
2019	11 532 363	11 451 154	11 489 940
2020	11 792 373	11 665 108	11 727 999
2021	12 052 382	11 879 062	11 966 058
2022	12 312 391	12 093 016	12 204 118
2023	12 572 400	12 306 970	12 442 177
2024	12 832 410	12 520 924	12 680 236
2025	13 092 419	12 734 877	12 918 295
2026	13 352 428	12 948 831	13 156 354
2027	13 612 437	13 162 785	13 394 414
2028	13 872 447	13 376 739	13 632 473
2029	14 132 456	13 590 693	13 870 532
2030	14 392 465	13 804 647	14 108 591
2031	14 652 474	14 018 601	14 346 651
2032	14 912 484	14 232 555	14 584 710
2033	15 172 493	14 446 509	14 822 769
2034	15 432 502	14 660 463	15 060 828
2035	15 692 511	14 874 416	15 298 887
2036	15 952 521	15 088 370	15 536 947
2037	16 212 530	15 302 324	15 775 006
2038	16 472 539	15 516 278	16 013 065

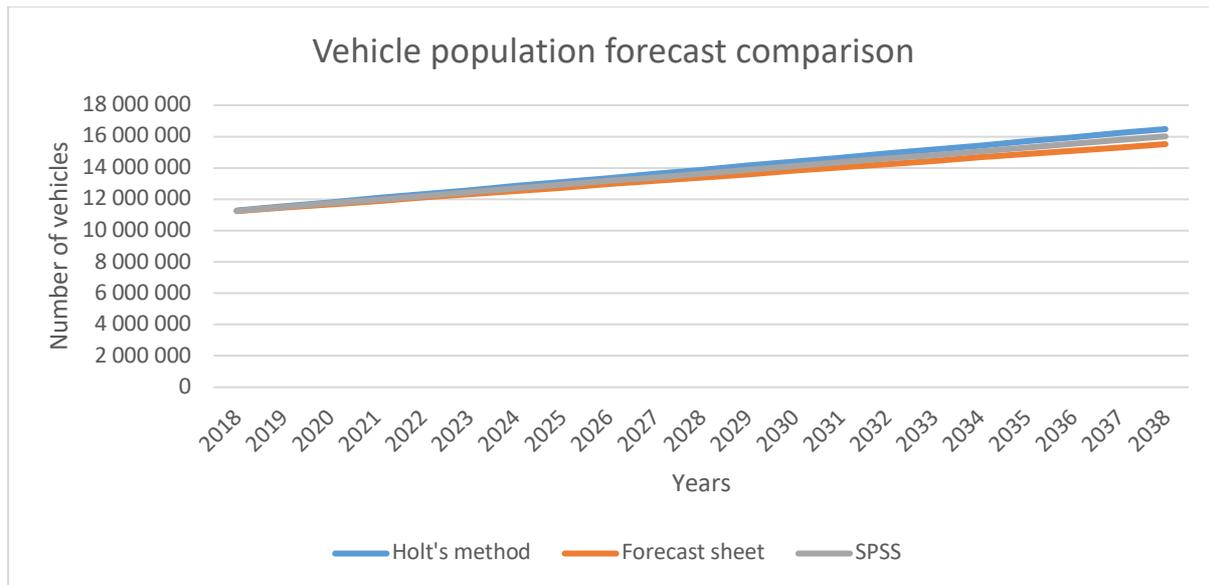


Figure 6.10: Holt's method vs forecast sheet vs SPSS forecasted vehicle population

Table 6.20 and Figure 6.11 for the real GDP are presented below, comparing the three forecasting methods. Studying these real GDP forecast results from all three of the applied methods, the results show that the Holt's method and SPSS results from the forecast are approximately similar with minor differences. Analysing the graphical representation of these forecasted values, the Holt's and SPSS line graphs move jointly and display the same pattern; however, the forecast sheet line is above the two graphs forecasting a much higher real GDP. The forecast sheet appears to overestimate the forecasts. This is the result of the 1.57% annual average growth rate, compared to the low rates of 1.09% and 1.03% average growth rates for the Holt's and SPSS results respectively.

It can be concluded that the real GDP for South Africa is expected to lie between 3 883 705 362 363.34 and 3 931 500 019 538.18 in local currency by the year 2038.

Table 6.20: Holt’s method vs forecast sheet vs SPSS forecasted real GDP

South African real GDP forecast			
Years	Holt's method	Forecast sheet	SPSS
2018	3 163 199 524 739,91	3 185 550 194 946,36	3 160 662 338 397,51
2019	3 201 614 549 479,83	3 244 088 174 378,04	3 196 814 489 595,80
2020	3 240 029 574 219,74	3 302 626 153 809,73	3 232 966 640 794,09
2021	3 278 444 598 959,65	3 361 164 133 241,42	3 269 118 791 992,38
2022	3 316 859 623 699,57	3 419 702 112 673,11	3 305 270 943 190,68
2023	3 355 274 648 439,48	3 478 240 092 104,80	3 341 423 094 388,97
2024	3 393 689 673 179,40	3 536 778 071 536,48	3 377 575 245 587,26
2025	3 432 104 697 919,31	3 595 316 050 968,17	3 413 727 396 785,55
2026	3 470 519 722 659,22	3 653 854 030 399,86	3 449 879 547 983,84
2027	3 508 934 747 399,14	3 712 392 009 831,55	3 486 031 699 182,13
2028	3 547 349 772 139,05	3 770 929 989 263,23	3 522 183 850 380,43
2029	3 585 764 796 878,96	3 829 467 968 694,92	3 558 336 001 578,72
2030	3 624 179 821 618,88	3 888 005 948 126,61	3 594 488 152 777,01
2031	3 662 594 846 358,79	3 946 543 927 558,30	3 630 640 303 975,30
2032	3 701 009 871 098,70	4 005 081 906 989,99	3 666 792 455 173,59
2033	3 739 424 895 838,62	4 063 619 886 421,67	3 702 944 606 371,88
2034	3 777 839 920 578,53	4 122 157 865 853,36	3 739 096 757 570,17
2035	3 816 254 945 318,44	4 180 695 845 285,05	3 775 248 908 768,47
2036	3 854 669 970 058,36	4 239 233 824 716,74	3 811 401 059 966,76
2037	3 893 084 994 798,27	4 297 771 804 148,43	3 847 553 211 165,05
2038	3 931 500 019 538,18	4 356 309 783 580,11	3 883 705 362 363,34

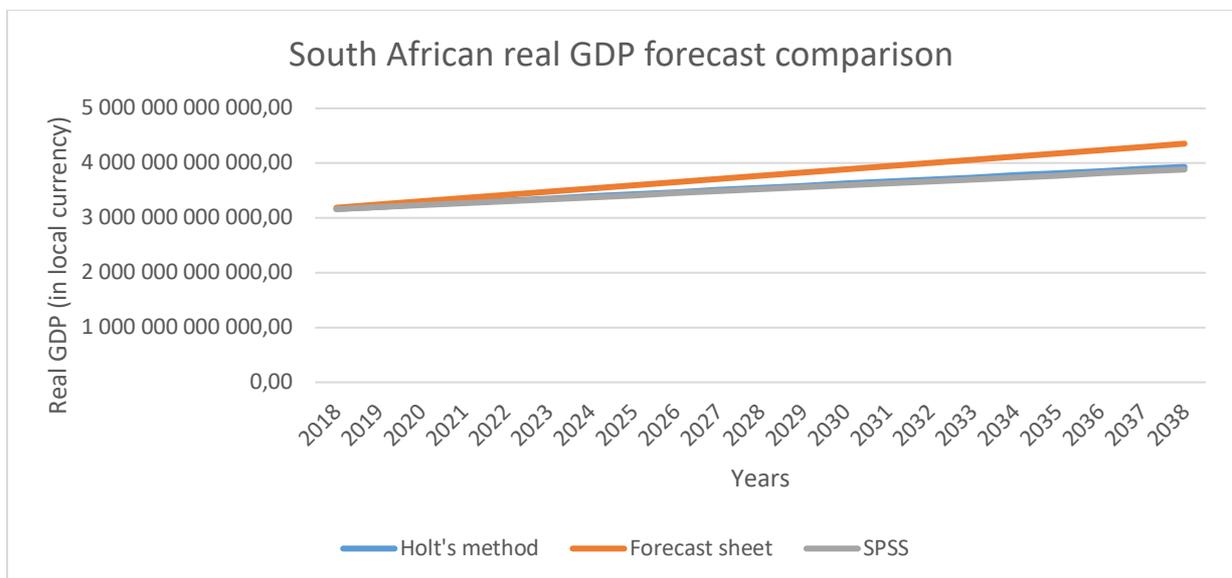


Figure 6.11: Holt’s method vs forecast sheet vs SPSS forecasted real GDP

Table 6.21 and Figure 6.12, comparing the results from the three methods for the South African population forecasts, are presented below. It is observed from these results that all

three methods or models show approximately similar forecasts with small differences. Furthermore, the graphical representation of these forecasted values shows that all three line graphs move conjointly and display the same pattern. The annual average growth rates of the variables are similar and therefore the line graphs display the same pattern. The line for the Holt's method moves at an average growth rate of 1.09%, forecast sheet at 1.08% and lastly the SPSS line moves at 1.09%.

It can be concluded that South Africa's population is approximated at an average of 71 372 788 people in the year 2038. Averaged from the predicted values of the three forecasts. The population will lie between 71 213 365 and 71 452 500 people by the year 2038.

Table 6.21: Holt's method vs forecast sheet vs SPSS forecasted South African population

South African population forecast			
Years	Holt's method	Forecast sheet	SPSS
2018	57 418 839	57 407 427	57 418 839
2019	58 120 522	58 097 724	58 120 522
2020	58 822 205	58 788 021	58 822 205
2021	59 523 888	59 478 318	59 523 888
2022	60 225 571	60 168 615	60 225 571
2023	60 927 254	60 858 912	60 927 254
2024	61 628 937	61 549 209	61 628 937
2025	62 330 620	62 239 506	62 330 620
2026	63 032 303	62 929 802	63 032 303
2027	63 733 986	63 620 099	63 733 986
2028	64 435 669	64 310 396	64 435 670
2029	65 137 352	65 000 693	65 137 353
2030	65 839 035	65 690 990	65 839 036
2031	66 540 718	66 381 287	66 540 719
2032	67 242 401	67 071 584	67 242 402
2033	67 944 084	67 761 881	67 944 085
2034	68 645 767	68 452 178	68 645 768
2035	69 347 450	69 142 474	69 347 451
2036	70 049 133	69 832 771	70 049 134
2037	70 750 816	70 523 068	70 750 817
2038	71 452 499	71 213 365	71 452 500

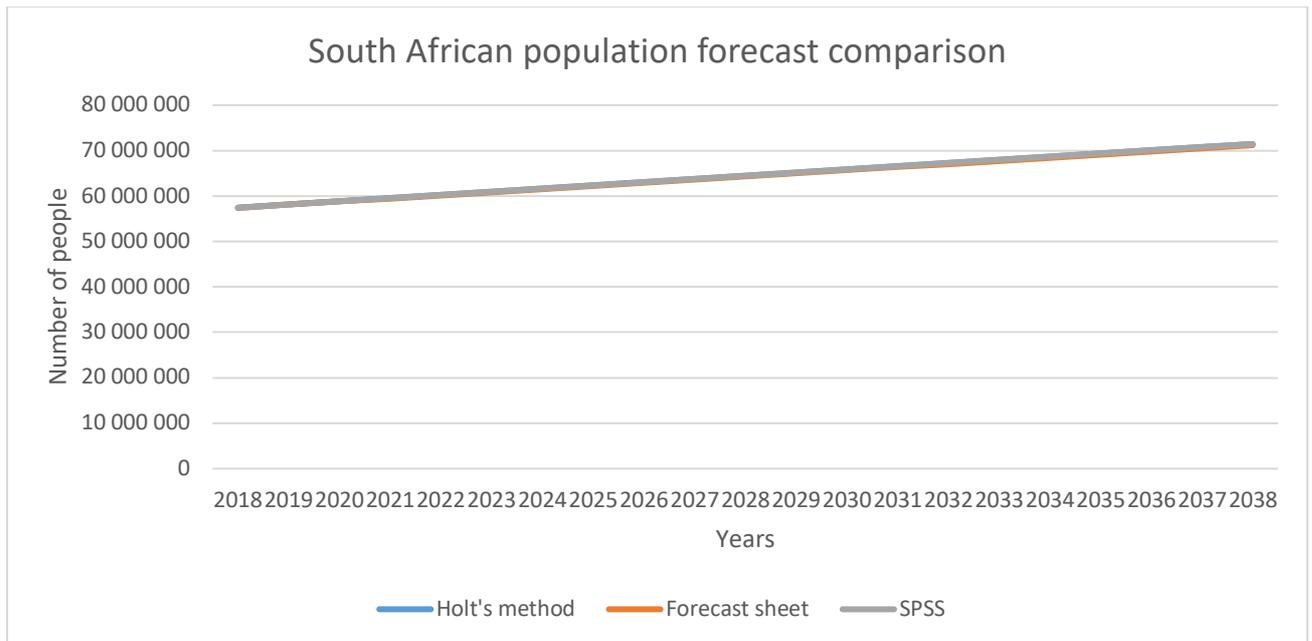


Figure 6.12: Holt’s method vs forecast sheet vs SPSS forecasted South African population

To measure the goodness of fit of the models, the root mean squared errors (RMSE) of the models are represented in Table 6.22 below. The RMSE indicates the absolute fit of the model to the data – how close the observed data points are to the predicted values of the applied model. The lower the values of the RMSE the better the fit, giving a good measure of the accuracy of the model prediction.

The table below shows that both the vehicle population forecasts and South African real GDP are better measured by the forecast sheet tool, as this gives the lowest RMSE values for the variables. The South African population is better fitted by the SPSS statistics software forecast function (Brown’s linear exponential smoothing model) according to this analysis; as stated in the table this model gives the lowest RMSE value.

Table 6.22: Vehicle population, real GDP and South African population RMSE for all three forecast methods

Root mean squared errors (RMSE)			
	Holt's method	Forecast sheet	SPSS
Vehicle population	142 880,54	106 446,10	136 650,31
South African real GDP	40 239 082 947,63	38 483 754 257,11	40 203 690 928,02
South African population	137 498,63	85 103,00	35 347,94

Therefore, it can be concluded that though the different models yield different forecast values with minimal difference in value, all forecasts display an upward increasing pattern. Consequently, an increase in vehicle population, real GDP and South African population is expected to increase between the stated ranges stated in the discussion above.

CHAPTER 7: VEHICLE OWNERSHIP MODEL

7.1. Vehicle ownership

The purpose of this section is to present the results from the vehicle household ownership estimation.

A household vehicles ownership model is estimated to determine the probability of owning zero, one or two or more household vehicles. The NHTS 2013 data was used to estimate the model. Table 6.23 shows the frequencies of responses obtained from the NHTS, there is a low vehicle accessibility in South Africa as most of the population is poor and do not have the financial capabilities of owning a vehicle. A multinomial logistics regression model (MNL) is used and the reference category is households owning zero vehicles. This means the results should be interpreted against this reference category.

Table 6.23: National Household Travel Survey 2013 household vehicle access summary

Data Source	Number	Percentage
NHTS 2013		
Households with access to 0 vehicles	32372	74,48%
Households with access to 1 vehicles	6682	15,37%
Households with access to 2+ vehicles	4128	9,50%
Households not reporting	280	0,64%
Total Households	43462	

Own table source (Statistics South Africa, 2014a)

Four explanatory variables that are considered to have an influence on the choice of household vehicle ownership, including: (a) include main dwelling, (b) income quintiles, (c) total household expenses and (d) geographical location. Statistics Package for Social Sciences (SPSS) is used in estimating the multinomial logistics regression model.

Table 6.24 identifies whether there is a relationship present between the number of vehicles owned by households and the identified independent variables. The chi-square serves as a basis for the testing of the strength relationship (Bayaga, 2010). The hypothesis for the null model is that the variables do not influence number of vehicles owned, while the alternative hypothesis indicates that the number of vehicles owned are statistically influenced by the independent variables.

In this analysis the probability of the model chi-square (16105,985) was 0.000, less than the significance level 0.05 ($p=0.000$, < 0.05), therefore indicating that the final model including the independent variables, predicts significantly better or more accurately than the intercept model

without independent variables. The null hypothesis is rejected and it is concluded that in this model there is evidence supporting the notion that there is a relationship between the number of vehicles owned by households and the identified independent variables.

Table 6.24: Model fitting information

Model Fitting Information				
Model	-2 Log Likelihood	Likelihood Ratio Tests		
		Chi-Square	df	Sig.
Intercept Only	17247,542			
Final	1141,557	16105,985	22	0,000

Table 6.25 presents a portion of the case processing summary table that includes all variables, response (dependent) variables and explanatory variables. It is observed that the number of valid observations used within the model is 40 107, distributed amongst the three categories of household vehicle ownership. The marginal percentage column shows the portion of the valid observations obtained in each response variable group; 75.4% of the valid cases own no vehicles, 15.3% of the valid cases own one vehicle and 9.3% own two or more vehicles.

Table 6.25: Case processing summary

Case Processing Summary			
		N	Marginal Percentage
Car_Household	0	30250	75,4%
	1	6131	15,3%
	2+	3726	9,3%
Valid		40107	100,0%
Missing		117166	
Total		157273	
Subpopulation		147 ^a	
a. The dependent variable has only one value observed in 38 (25.9%) subpopulations.			

Subpopulation

Subpopulation addresses the number of subpopulations that are contained within the data. The subpopulation a part of the total population data, includes one combination of the explanatory variables specified for this model. The footnote in Table 6.25 indicates how many of the combinations of the explanatory variables consist of records that have the same value in the response variable. In this model there are 147 combinations that are found in the data and 38 of the combinations are observed with the same response variable categories.

Missing Cases

Missing specifies the number of cases in the data where data is missing for the response variable and any of the explanatory variables. In the model, it is found that there are 117 166 cases in the dataset with missing data. Scales can still be calculated if at least two thirds of the cases are complete (Bayaga, 2010). A category of “unspecified”, “do not know” or “refuse” are not valid decisions with regards to this data analysis, it is considered as missing and set as such in SPSS.

By chance accuracy

The proportional by chance accuracy can be calculated by estimating the share of the cases for each group, based on the number of cases in each of the response variables. This is calculated by summing the squared proportion of cases in each group. In the case of the abovementioned cases the chance accuracy is $(0.754^2 + 0.153^2 + 0.093^2) = 0.600574 = 60.01\%$. The benchmark that is used to characterise the multinomial logistics regression model as useful is a 25% improvement over the calculated rate of accuracy reached by chance (Bayaga, 2010). The proportional by chance accuracy criterion is: $(1.25 * 0.600574) * 100 = 75.07\%$. This proportion rate is compared with the overall percentage of the final model.

Pseudo R-square

SPSS calculates three pseudo R-square values for logistics regression (Table 6.26). The pseudo R-square is a statistically produced ordinary least squares (OLS) regression and is a correlation measure computed by multinomial logistics regression, which attempts to estimate the strength of the relationship of the model or provide the amount of variance explained in the household vehicle ownership by the independent variables included in the model.

The Cox and Snell is based on log-likelihoods and considers sample size; Nagelkerke adjusts Cox and Snell so the value of 1 can be realised; and McFadden is the transformation of the likelihood ratio statistic intended to mimic an R-square. The pseudo R-squared presented in Table 6.26 suggests that the model including number of vehicles owned by households and independent variables is a better fit, than a model with just the number of vehicles owned by household as a variable. These statistics are most useful in the comparison of competing models for the same data.

Table 6.26: Pseudo R-Square

Pseudo R-Square	
Cox and Snell	0,331
Nagelkerke	0,433
McFadden	0,279

7.1.1. Evaluating the usefulness of the model

The pseudo R-square provides a measure of the magnitude of association between household vehicle ownership and the independent variables. However, it fails to provide the accuracy or errors in the model. Table 6.27 compares the successful predictions of households owning 0, 1, 2+ vehicles to number of households owning 0, 1 or 2+ actually observed. This assesses the fit of the model to the data.

The by chance accuracy rate was computed above by calculating the proportion of the cases for each group based on the number of cases in each group and then summing and squaring the proportion of cases in each group. This by chance accuracy calculation gave 75.07%; the benchmark used in the multinomial model is a 25% improvement over the rate achieved by chance accuracy. Therefore, the classification rate in the study should be at least 25% more than the proportion by chance accuracy of 60.01% for the multinomial logistics regression to be satisfactory.

Table 6.27 shows that the overall classification accuracy rate is 79.5%, which is greater than the proportional by chance accuracy rate criterion calculated as 75.07%, this suggests that the criterion is satisfied and the model is acceptable.

Table 6.27: Classification table

Classification				
Observed	Predicted			Percent Correct
	0	1	2+	
0	29078	932	240	96,1%
1	4006	1382	743	22,5%
2+	1185	1126	1415	38,0%
Overall Percentage	85,4%	8,6%	6,0%	79,5%

7.1.2. Relationship between the independent variables and dependent variable

A further analysis of the relationship of the number of vehicles owned by households and the identified independent variables is conducted using the likelihood ratio tests, which evaluate the overall relationship between the number of vehicles owned by households and each individual independent variable. Table 6.28 presents the likelihood ratio tests for the model and displays the significance of each individual independent variable.

Observations of the table show that each of the independent variables has a p-value of 0.00, which is less than the level of significance at 0.05. Thus, there is a relationship between the

number of vehicles owned by households and each of the independent variables, and as a result of these relationships all four of these variables should be included in the model.

Though these relationships exist between the dependent and independent variables, it does not mean that each independent variable is statistically significant.

Table 6.28: Likelihood ratio tests

Likelihood Ratio Tests				
Effect	-2 Log Likelihood of Reduced Model	Likelihood Ratio Tests		
		Chi-Square	df	Sig.
Intercept	1141.557 ^a	0,000	0	
Main_Dwelling	1839,022	697,465	4	0,000
Household_Inc_Quintiles	3681,275	2539,718	8	0,000
Geo_Location	1158,910	17,352	4	0,002
Household_Expenses	4723,402	3581,844	6	0,000
The chi-square statistic is the difference in -2 log-likelihoods between the final model and a				
a. This reduced model is equivalent to the final model because omitting the effect does not				

7.1.3. Model estimation

Table 6.29 shows the parameter estimate results from the multinomial model, together with the statistically significant (according to the chi-squared tests) predictor variables obtained from the analysis. This table also shows the coefficient estimates (B), standard errors, the Wald Statistics, the odds ratio Exp (B) together with the corresponding 95% confidence intervals.

Table 6.29: Parameter Estimates

Parameter Estimates									
Car Household ^a	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for		
							Lower Bound	Upper Bound	
1.00	Intercept	0,640	0,385	2,758	1	0,097			
	[Main_Dwelling=Formal Dwelling]	0,602	0,377	2,554	1	0,110	1,826	0,873	3,823
	[Main_Dwelling=Informal Dwelling]	-0,473	0,380	1,551	1	0,213	0,623	0,296	1,312
	[Main_Dwelling=12.00]	0 ^b			0				
	[Household_Inc_Quintiles=R0-R395]	-1,829	0,062	860,753	1	0,000	0,161	0,142	0,181
	[Household_Inc_Quintiles=R395.11-R828]	-1,665	0,055	919,159	1	0,000	0,189	0,170	0,211
	[Household_Inc_Quintiles=R828.33-R1 600]	-1,333	0,050	724,749	1	0,000	0,264	0,239	0,290
	[Household_Inc_Quintiles=R1 601-R4 017]	-0,798	0,045	308,766	1	0,000	0,450	0,412	0,492
	[Household_Inc_Quintiles=R4 023-R222 000]	0 ^b			0				
	[Geo_Location=Metro]	0,167	0,043	15,034	1	0,000	1,182	1,086	1,285
	[Geo_Location=Urban]	0,147	0,042	12,232	1	0,000	1,158	1,067	1,258
	[Geo_Location=Rural]	0 ^b			0				
	[Household_Expenses=R 0-R399]	-2,877	0,123	546,606	1	0,000	0,056	0,044	0,072
	[Household_Expenses=R 400-R1 799]	-2,263	0,079	825,407	1	0,000	0,104	0,089	0,121
	[Household_Expenses=R 1 800-R9 999]	-1,063	0,075	203,243	1	0,000	0,345	0,298	0,400
	[Household_Expenses=R 10 000 or more]	0 ^b			0				
2+	Intercept	0,676	0,774	0,762	1	0,383			
	[Main_Dwelling=Formal Dwelling]	1,364	0,769	3,147	1	0,076	3,912	0,867	17,662
	[Main_Dwelling=Informal Dwelling]	-0,311	0,777	0,160	1	0,689	0,733	0,160	3,363
	[Main_Dwelling=12.00]	0 ^b			0				
	[Household_Inc_Quintiles=R0-R395]	-2,724	0,111	605,423	1	0,000	0,066	0,053	0,082
	[Household_Inc_Quintiles=R395.11-R828]	-2,554	0,090	803,618	1	0,000	0,078	0,065	0,093
	[Household_Inc_Quintiles=R828.33-R1 600]	-2,196	0,073	894,792	1	0,000	0,111	0,096	0,128
	[Household_Inc_Quintiles=R1 601-R4 017]	-1,385	0,056	608,932	1	0,000	0,250	0,224	0,279
	[Household_Inc_Quintiles=R4 023-R222 000]	0 ^b			0				
	[Geo_Location=Metro]	0,090	0,065	1,940	1	0,164	1,094	0,964	1,242
	[Geo_Location=Urban]	0,066	0,065	1,030	1	0,310	1,068	0,940	1,214
	[Geo_Location=Rural]	0 ^b			0				
	[Household_Expenses=R 0-R399]	-4,467	0,224	396,329	1	0,000	0,011	0,007	0,018
	[Household_Expenses=R 400-R1 799]	-3,972	0,094	1799,021	1	0,000	0,019	0,016	0,023
	[Household_Expenses=R 1 800-R9 999]	-2,021	0,074	750,651	1	0,000	0,133	0,115	0,153
	[Household_Expenses=R 10 000 or more]	0 ^b			0				
a. The reference category is: .00.									
b. This parameter is set to zero because it is redundant.									

The Wald Statistics and the significance value (p-value) are used to test the significance of each independent variable. In this model, they test if the independent variables can significantly distinguish the number of vehicles owned in a household (1, 2+ vehicles) against the reference category, which is households with 0 vehicle ownership.

The estimated coefficients are a measure of change in the logit for a one-unit change in the predictor variable, with all other predictors remaining constant. A positive estimated coefficient indicates an increase in the probability that a household will own 1, 2 or more vehicles, and a negative estimated parameter indicates that there is less probability that ownership of 1 or 2+ vehicles per household will occur.

The p-value (Sig.) shows whether a change in the predictor variable significantly alters the logit at the acceptance level (95%). If the p-value of an estimate is higher than the accepted confidence level, then there is unsatisfactory evidence that a change in the predictor affects the response category regarding the reference category.

The parameter estimates table found in table 6.29 shows that significant predictor variables at 95% (p-value < 0.05) include the following: all four income quintiles on all categories 1 and 2+ vehicle ownership households, while main dwelling shows no significance in any of the dependent variable categories of household vehicle ownership. All total household expenditure ranges are significant in all categories 1 and 2+ vehicle ownership households, whereas geographical location Metro, Urban and Rural categories are significant only for 1 vehicle households. Therefore, the interpretation of the model will focus on these significant variables.

The Exp (B), also known as the odds ratio is the predicted change in odds for a unit increase in the independent variables. The "Exp" refers to exponential value of B. When the value of Exp (B) is less than 1, it indicates that the likelihood of an event occurring is less likely for the response category (dependent variable) than the selected reference category of an independent variable. When the value Exp (B) is bigger than 1, it indicates that the likelihood of an event occurring is more likely for the response category (dependent variable) than the selected reference category of an independent variable.

Household monthly income quintiles:

Analysing household monthly income quintile 1 (R0-R395) it can be deduced that, when holding other variables constant, the odds for households with a monthly income within quintile 1 (R0-R395) owning 1 vehicle rather than owning 0 vehicles, is 0,161 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 1

vehicle increase by one unit, the probability of the household falling within quintile 1 decreases by 0.161.

Again the odds of households with a monthly income within quintile 1, owning 2 vehicles rather than 0 vehicles, is 0.066 times less likely than households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 2 vehicle increase by one unit, the probability of the household falling within quintile 1 decreases by 0.066.

The Exp (B) of household monthly income quintile 2 (R395.11-R828) is 0.189. It can be assumed that, when holding other variables constant, the odds for households with a monthly income within quintile 2 (R395.11-R828) owning 1 vehicle rather than owning 0 vehicles, is 0,189 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within quintile 2 decreases by 0.189.

The odds ratio of household monthly income quintile 2 (R395.11-R828) is 0.189. It can be assumed that, when holding other variables constant, the odds for households with a monthly income within quintile 2 (R395.11-R828) owning 2 or more vehicles rather than owning 0 vehicles, is 0,078 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 2 or more vehicles increase by one unit, the probability of the household falling within quintile 2 decreases by 0.078.

The value of Exp (B) of household monthly income quintile 3 (R828.33-R1 600) is 0.264. It can be assumed that, when holding other variables constant, the odds for households with a monthly income within quintile 3 (R828.33-R1 600) owning 1 vehicle rather than owning 0 vehicles, is 0,264 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within quintile 2 decreases by 0.264.

Whereas, again holding other variables constant, the odds for households with a monthly income within quintile 3 (R828.33-R1 600) owning 2 or more vehicles rather than owning 0 vehicles, is 0,111 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 2 or more vehicles increase by one unit, the probability of the household falling within quintile 2 decreases by 0.111.

The Exp (B) of household monthly income quintile 4 (R1 600.01-R4 017) is 0.450. When holding other variables constant, the odds for households with a monthly income within quintile 4 (R1 600.01-R4 017) owning 1 vehicle rather than owning 0 vehicles, is 0,450 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within quintile 2 decreases by 0.450.

Holding other variables constant, the odds for households with a monthly income within quintile 4 (R1 600.01-R4 017) owning 2 or more vehicles rather than owning 0 vehicles, is 0,250 times less likely than that of households with a monthly income within quintile 5 (R4 023-R222 000). The regression coefficient is negative and therefore, that means as South African households owning 2 or more vehicles increase by one unit, the probability of the household falling within quintile 2 decreases by 0.250.

In conclusion, the analysis on the odds ratio values of the significant household income quintiles shows that households within household income quintile 1 (R0-R395), have the least likelihood of owning 1 vehicle, with an odds of 0.161, and the least likelihood of owning 2 or more vehicles with an odds ratio of 0.066, compared to owning 0 vehicles. This conclusion is supported by the graphs in Chapter 3 on the data. There it is illustrated that 5.6% of the households within the lowest quintile own 1 vehicle, this being the lowest percentage of households owning 1 vehicle compared to the other quintiles. The more income available to a household, the more capable the household is of owning a vehicle because of their increased buying power.

Total household monthly expenditure:

The total household monthly expenditure 1 (R0-R399) Exp (B) is 0.056. It can be deduced that, when holding other variables constant, the odds for households with total monthly expenditure 1(R0-R399) owning 1 vehicle rather than owning 0 vehicles, is 0,056 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.056.

Holding other variables constant, the odds for households with total monthly expenditure 1 (R0-R399) owning 2 or more vehicles rather than owning 0 vehicles, is 0,011 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households

owning 2 or more vehicles increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.011.

The total household monthly expenditure 2 (R400-R1 799) odds ratio is 0.104. It can be deduced that, when holding other variables constant, the odds for households with total monthly expenditure 2 (R400-R1 799) owning 1 vehicle rather than owning 0 vehicles, is 0,104 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.104.

Holding other variables constant, the odds for households with total monthly expenditure 2 (R400-R1 799) owning 2 or more vehicles rather than owning 0 vehicles, is 0,019 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households owning 2 or more vehicles increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.019.

The total household monthly expenditure 3 (R1 800-R9 999) odds ratio is 0.0.345. Holding other variables constant, the odds for households with total monthly expenditure 3 (R1 800-R9 999) owning 1 vehicle rather than owning 0 vehicles, is 0,345 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.345.

Holding other variables constant, the odds for households with total monthly expenditure 3 (R1 800-R9 999) owning 2 or more vehicles rather than owning 0 vehicles, is 0,133 times less likely than that of households with a total monthly household expenditure 4 (R10 000 or more). The regression coefficient is negative and therefore, that means as South African households owning 2 or more vehicles increase by one unit, the probability of the household falling within total monthly expenditure 1 decreases by 0.133.

From the above analysis on the odds ratio values of the total household expenditure of the households in the survey, it becomes clear that the more vehicles a household owns the higher the total household expenditure. It can be concluded that households with total household expenditure 3 (R1 800-R9 999) are most likely to own 1 vehicle and 2 or more vehicles with odd ratios 0.345 and 0.133 respectively, compared to households falling within other expenditure categories. This supports the data analysis in Chapter 3 where households within

total households expenditure 3 owning 2 or more vehicles made up 15.77% and 28.20% households in that category of owning 1 vehicle.

Geographical location:

The parameter estimates table shows that the Exp (B) value of the geographical location 1 (Metro) is 1.182. Holding other variables constant, the odds for households residing in a metro geographical location owning 1 vehicle rather than owning 0 vehicles, is 1.182 times more likely than that of households residing in a rural geographical location. The regression coefficient is positive and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household residing in a metro geographical location increase by 1.182.

While the Exp (B) value of the geographical location 2 (urban) is 1.158. Holding other variables constant, the odds for households residing in a urban geographical location owning 1 vehicle rather than owning 0 vehicles, is 1.158 times more likely than that of households residing in a rural geographical location. The regression coefficient is positive and therefore, that means as South African households owning 1 vehicle increase by one unit, the probability of the household residing in an urban geographical location increase by 1.158.

The above is an analysis of the odds ratio values of the significant geographical location of the households in the conducted survey. All geographical location types for households owning 2 or more vehicles were statistically insignificant in the study model. In conclusion, households residing within metropolitan areas are most likely to have 1 vehicle as opposed to owning no vehicles. The graphs in Chapter 3 concerning geographical location support this conclusion, as 20.70% of households in metropolitan areas own 1 vehicle. This is the highest household percentage share owning 1 vehicle in the three stated geographical area types.

7.2. Summary and conclusion

The analysis and discussion above conclude that households are less likely to own 1 vehicle as compared to owning 0 vehicles, with regards to the independent variables included in the scope of this study. Results show that the accuracy of the model used is 79.5% accurate and thus acceptable.

The same pattern income displayed graphically in chapter 3, is observed again in the multinomial logistics regression model, observed through the monthly income quintiles. As household monthly income increases the odds of owning 1, 2 or more vehicles increases notably. Income creates accessibility and ability to purchase vehicles for many households. This indicates the sensitivity of vehicle ownership in South Africa to the change in income.

Household income quintiles revealed that households earning an income within the household income quintile 1 (R0-R395) were the less likely than households in higher income quintiles to own 1 vehicle compared to owning 0 vehicles. Households within household income quintile 1 are the least likely to own 2 or more vehicles compared to owning 0 vehicles. These results reveal that income

Total household expenditure shows that households with expenditure falling within the R0-R395 range compared with other households within other expenditure ranges are less likely to own 1 vehicle or 2 or more vehicles compared to owning 0 vehicles.

Similar to income, as households move from rural geographical locations to metro locations, the likelihood of owning 1, 2 or more vehicles increases. There is a strong influence by geographical location moves on the household vehicle ownership, these new locations offer better road networks to gain access to areas work and opportunities. The safety offered by these geographical location is another encourager for households to increase vehicle ownership.

Therefore, as in the analysis of geographical location, households located in Metropolitan areas are more likely than households in other geographical locations to own 1 vehicle with an odds ratio of 1.182, compared to owning 0 vehicles. Vehicle infrastructure has an impact on the choice of ownership, thus households in urban areas have a higher likelihood of owning vehicles. This is to be expected as infrastructure in such areas is better than that in rural or urban areas.

CHAPTER 8: CONCLUSIONS

This study examined the future number of vehicles in South Africa and further investigated factors that influence household vehicle ownership. This was done to better understand what impacts the increase in number of vehicles owned by a household and the probability of those households owning 0, 1, 2 or more vehicles given the identified factors (independent factors). The investigation on the probability of ownership, was undertaken using the multinomial logistics regression model.

The main questions addressed within the study are; (1) what socio-economic factors influence household vehicle ownership? and (2) What will the growth in vehicles be for South Africa for the period 2018-2038? The study answers these research questions by also focusing on investigating what the main factors influencing household vehicle ownership are and what the probabilities of these factors are on households owning 0, 1, 2 or more vehicles.

Literature indicated that a number of researchers have developed and conducted forecasting models concerning number of vehicles. However, literature in the South African context was found to be limited. Literature than showed that vehicle ownership forecasting models are vital for the business of local and national governments and manufacturing industries. The ability to forecast provides industries with the capacity to conduct efficient demand planning.

Various models have been developed for the forecasting of vehicle ownership, providing alternatives for different purposes of the required forecast. All the models are used for forecasting; however, the choice of a model depends on the type of forecasting that needs to be done. Different models of forecasting were used in public and in private sector, with most developing countries discovered to make use of aggregated time series models.

Income and geographical location have been used frequently as influencing factors for households' probability of owning a certain number of vehicles. Models applied in the study use exponential smoothing on the data during the forecasting process for the prediction of number of vehicles in the period 2018-2038. The data used in the analysis has no seasonality present, and thus the models focus only on trends when conducting forecasts. When using SPSS as a forecasting tool, it is important to note the three assumptions made in its construction. First, it was assumed that the dependent variable and independent variable are time series data. Second, the data is assumed stationary. Third, the independent variable does not have missing values in the period used for forecasting.

The main findings show that the forecasted number of vehicles in South Africa are approximated at an average of 16 000 627 vehicles in 2038. This being the average between the three predicted number of vehicles given by the forecast models used. Furthermore, the

population is predicted at an average of 71 372 788 people for 2038, average of the predicted totals of the three models. With, number of vehicles growing at an approximated average annual rate of 1.75% and the human population at an approximated average annual growth rate of 1.08% in the analysis period 2018-2038.

This predicted number of vehicles indicate that, in the year 2038 there will be approximately 224 vehicles per 1000 of the population. Compared to the 165 vehicles per 1000 of the population in 2014, South Africa with all countries held constant will move up to number 65 out of the 191 countries included in the Figure 8.1. The rapid increase in the number of vehicles in South Africa is a result of the growing population.

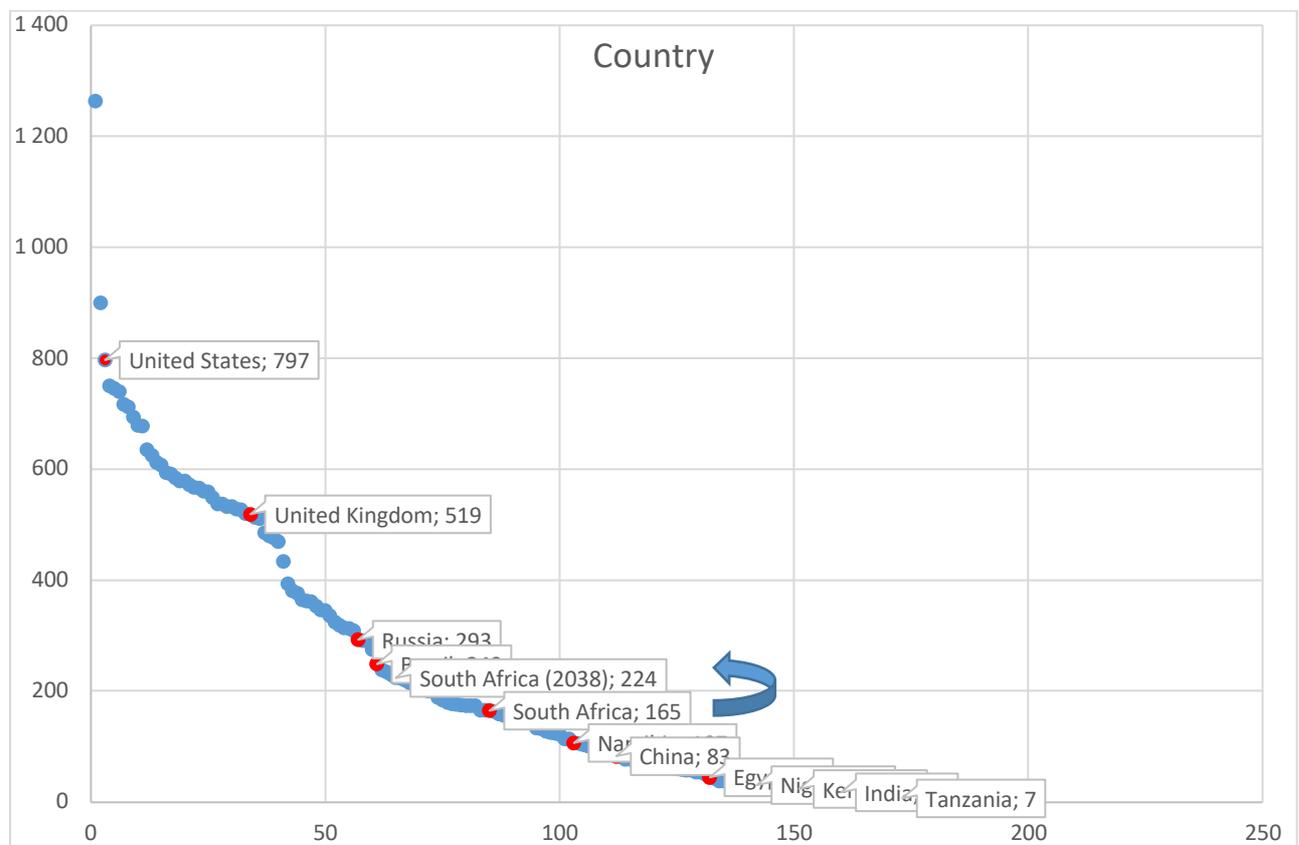


Figure 8.1: Vehicles per 1000 of the population including South Africa 2038.

The growing vehicle fleet will lead to an increase in congestion on South African roads. This congestion is an inescapable result of the community's behaviour, the desires of the household is to now chase certain goals that overload roads on a daily basis. Congestion is the result of too many commuters wanting to travel at the same time, as work, school and certain errands demand that most households be on the road at the same time.

South Africa's fuel consumption is expected to show an increase. Conventional vehicles primarily consume liquid fuel and thus, with an expected vehicle increase, this implies that the demand in energy (fuel) will experience an increase. This would mean South Africa is to expect

an increase in air pollution, resulting from the increased carbon dioxide emissions from the additional vehicles. South African government will need to investigate better carbon dioxide emission regulations and try reduce private vehicle use by households.

In order to predict the odds of a household owning 0, 1, 2 or more vehicles the multinomial logistics regression model is applied. This is the most fitted model for the type of data used in the study of household vehicle ownership.

Geographical location of household, totally monthly household expenditure, monthly household income quintiles and main dwelling were identified as the important and most significant factors influencing household vehicle ownership. With geographical location revealed to have the strongest likelihood in influencing households to owning 1, 2 or more vehicles as compared to owning no vehicles.

The analysis and discussion concludes that households are less likely to own 1 vehicle as compared to owning 0 vehicles, with regards to the independent variables included in the scope of this study. Results show that the accuracy of the model used is 79.5% accurate and thus acceptable.

As anticipated, the study concluded that a positive relationship exists between household location and household vehicle ownership. Households in geographical locations with vehicle infrastructure, such as a metropolitan area are expected to have a higher vehicle ownership compared to households in rural areas. In the study 15.7% of households residing within metropolitan areas own 2 or more vehicles, while only 3.5% of the households residing in rural areas own 2 or more vehicles and 87.1% of households in this type of area (rural) have no vehicles. The odds of households within metropolitan areas owning 1 vehicle rather than owning 0 vehicles, is 1.182 times more likely than that of households residing in a rural geographical location.

Vehicle infrastructure has an impact on the choice of ownership, thus households in urban areas have a higher likelihood of owning vehicles. This is to be expected as infrastructure in such areas is better than that in rural or urban areas. Main areas of land use that effect household vehicle ownership include retail location, business location and housing location. Overall, outlook towards urban density has a large impact on vehicle ownership probability.

The multinomial logistics regression results indicate that as households move from rural geographical locations to metro locations, the likelihood of owning 1, 2 or more vehicles increases. There is a strong influence by geographical location moves on the household vehicle ownership, these new locations offer better road networks to gain access to areas of work and opportunities. The safety offered by these geographical location is another encourager for households to increase vehicle ownership.

The study also confirms that the higher total expenditure a household has, the higher the chances that that household will own a vehicle. Figure 3.13 in Chapter 3 displays this, in this figure 56.7% of households with a total expenditure of R10 000 or more own 2 or more vehicles, and only 0.83% of households with an expenditure of R0-R399 own 2 or more vehicles, whereas 93.87% of households in this expenditure category own no vehicles.

The odds of a household within a low expenditure range owning 1, 2 or more vehicles is lower compared to those households within a higher expenditure range. Therefore, the higher the number of vehicles owned within a South African household, the higher the expenditure. The negative regression coefficient means, as South African households owning 1 vehicle or the probability of owning 2 or more vehicles increase by one unit, the probability of the household falling within the stated total monthly expenditure decreases by the odds ratio.

The same pattern income displayed graphically in chapter 3, is observed again in the multinomial logistics regression model, observed through the monthly income quintiles. As household monthly income increases the odds of owning 1, 2 or more vehicles increases notably. Income creates accessibility and ability to purchase vehicles for many households. This indicates the sensitivity of vehicle ownership in South Africa to the change in income.

The positive from the increase in vehicle population for the South African government would be the potential increase in income. Vehicle fleet is connected to various income sources, such as tax from the fuel levy of which is expected to increase. As vehicles increase, fuel consumption will increase. Road accident fund is another income source that will experience an income, value added tax on the predicted vehicle sales and income collected from vehicle licensing

The predicted growth in number vehicles will increasingly become a problem for the government of South Africa if strategies are not put in place to attempt reducing the number of private vehicles on national roads. The 'stop and go' traffic flow that will be caused by the congestion of these additional vehicles increase travel time and potential accidents and air pollution. There are potential strategies the government can implement to lessen costs associated with increasing vehicle ownership.

A short term solution would be to expend road network capacity, specifically road networks used to access central business districts where all work opportunities are situated. This capacity would be built to manage drivers traveling in peak hours at the same time with minimal delays. However, it has been noted in the past that the extension of road networks encourages more individuals to travel in private vehicles because there is now 'more' space on the roads and roads would be underutilized during off peak hours. This solution would be very expensive for the government and it would cure the problem only for a while.

To decrease the pressure from road networks, governments need to encourage commuters to shift from private vehicle use, to public transport. However, South Africa's public transport network doesn't have a good reputation and thus the government would have to work on safety concerns associated with public transport, ensure that the existing public transport is accessible, efficiency, affordability and convenient in terms of location and ensure that it runs on time services. Public transport has great potential of reducing congestion to a certain level.

South African government can introduce a lane congestion toll, targeted at peak-hours. The charge would aid in decreasing congestion during peak hours, on major commuting roads. A lane or two allowing those willing to pay to drive at a higher speed during these peak periods and without forcing low-income drivers off the same roads during these peak periods.

The implementation of flexible work hours by companies can help in reducing congestion, by allowing employees to come into work after peak hours and leaving before afternoon peak. Allowing the flexibility of also working from home. This solution could potentially lead to people getting rid of private vehicles and opting to use e-hailing as a way to getting to work and meetings when needing to.

Urban transport policy framework, involves establishing that socio-spatial structure of South African towns is an inherited structure of the apartheid era, which placed large percentages of underprivileged African, Coloured and Indian population in outlying locations. These locations were a significant distance from areas with employment opportunities and higher standard commercial and social facilities. Thus, government needs to establish structures that facilitate integrated public transport in these areas. Transport planning needs to address the underlying areas and work towards better infrastructures that will allow households in these areas to access employment opportunities.

8.1. Future research

The study has revealed the prospective for qualitative approaches to provide explanations to the continuous increase in number of vehicles within South Africa, attending to some of the knowledge gaps within the industry regarding forecasting. Though the sample used in the study was limited to the sample size used in the national household travel survey conducted in 2013 and not a particular representation of the current South African households the intention has been to obtain a general framework. Nevertheless, further variables are required to project the probabilities of household ownership on a wider scale.

The study can be taken further and split the vehicles into types of vehicles i.e. heavy vehicles, busses, private vehicles and light vehicles, forecasting the number of vehicles in terms of vehicles and looking at the amount of road in square meters that these vehicles will need and

occupy. Then looking into the forecasted demand these vehicles will bring to the fuel industry, forecast petrol and diesel demands. Electrical vehicles can also be a small section of the research, identifying the impact on the numbers of vehicles, whether there will be significant decrease in the demand of conventional vehicles or demand will remain stagnant.

There is potential for a further study, similar to this study and including more variables with a stronger correlation to vehicle ownership. The researcher can take it a step further and look at the relationship between household vehicle ownership and life events. For example the research can investigate the motivation behind the relocation of households and identify how this impacts the decision to own a vehicle. Otherwise, also investigating the influence of the birth of a child in a household to acquire a vehicle.

With the rise of technology in the transport space, it would make sense to also explore the impact of shared ownership on household vehicle ownership. Some households may not afford owning a car alone but still require the benefits of owning one at their convenience. The likes of Uber, bolt have become players in the transport space and that can impact vehicle ownership but how much of an impact is it. Will there be a decrease in ownership in some household groups that opt to use these services than own a vehicle.

To conclude it is important to note, the decision to purchase a vehicle as a household can be influenced by many variables and not limited to those used. Changes in the number of vehicles owned by a household should be considered as the outcome of a continuous process of adjustments to life events and with a better access to employment opportunities through improved road infrastructure in urban and rural areas, household vehicle ownership will display a better spread within South Africa.

REFERENCES

- Adam. (2018). Forecasting in Excel for Analyzing and Predicting Future Results. Retrieved June 26, 2019, from Spreadsheetweb website:
<https://www.spreadsheetweb.com/forecasting-in-excel/>
- Bayaga, A. (2010). Multinomial Logistic Regression: Usage and Application in Risk Analysis. *Journal of Applied Quantitative Methods*, 5(2), 288–297.
- BBCNews. (2018). South Africa profile-Timeline. Retrieved September 29, 2019, from BBC website: <https://www.bbc.com/news/world-africa-14094918>
- Beca Carter Hollings, & Ferner Ltd. (2003). *Smartgrowth - Car ownership forecasting*.
- Ben-Akiva, M., & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In R. W. Hall (Ed.), *Handbook of Transportation Science* (Second, pp. 7–37). https://doi.org/10.1007/978-1-4615-5203-1_2
- Bhat, C. R., & Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B*, 32(1), 61–75.
- Bjørner, T. B., & Leth-Petersen, S. (2004). *The effect on car ownership of changes in household size and location - Descriptive analyses base on panel household data*. Retrieved from
http://www.akf.dk/workingpaper/2004/pdf/5_2004_effect_car_ownership.pdf
- Boogaard, K. (2018). Convergent Thinking vs. Divergent Thinking: Why Planning Isn't Always the Right Thing to Do. Retrieved October 1, 2019, from
<https://www.wrike.com/blog/convergent-thinking-vs-divergent-thinking/>
- Button, K., Ngoe, N., & Hine, J. (1993). *Modelling Vehicle ownership and use in low income countries*. 27(1), 51–67.
- Button, K., Pearman, A. ., & Fowkes, A. . (1982a). *Car ownership modelling and forecasting*.
- Button, K., Pearman, A. ., & Fowkes, A. . (1982b). *Car ownership modelling and forecasting*.
- Button, K., Pearman, A. ., & Fowkes, A. . (1982c). *Car ownership modelling and forecasting*.
- Button, K., Pearman, A. ., & Fowkes, A. . (1982d). *Car ownership modelling and forecasting*.
- Cheusheva, S. (2019). Excel forecast and related functions with formula examples. Retrieved June 7, 2019, from Ablebits.com website: <https://www.ablebits.com/office-addins-blog/2019/03/13/excel-forecast-function-formula-examples/>
- Cirillo, C. (2010). *Automobile Ownership Model*.

- Cole, N. L. (2018). Secondary Data and Secondary Analysis - An Overview. Retrieved May 7, 2019, from ThoughtCo. website: <https://www.thoughtco.com/secondary-analysis-3026573>
- Cundill, M. A. (1986). *Car ownership and use in Kenya*.
- Dargay, J., & Gately, D. (1999). Income's effect on car and vehicle ownership, worldwide: 1960-2015. *Transportation Research Part A: Policy and Practice*, 33(2), 101–138. [https://doi.org/10.1016/S0965-8564\(98\)00026-3](https://doi.org/10.1016/S0965-8564(98)00026-3)
- De Jong, G., Fox, J., Daly, A., Pieters, M., & Smit, R. (2004). Comparison of car ownership models. *Transport Reviews*, 24(4), 379–408. <https://doi.org/10.1080/0144164032000138733>
- Discrete choice analysis. (2010). Retrieved October 1, 2019, from <https://www.mailman.columbia.edu/research/population-health-methods/discrete-choice-analysis>
- Explorable.com. (2009). Quantitative and Qualitative Research - Objective or Subjective? Retrieved June 26, 2019, from Explorable.com website: <https://explorable.com/quantitative-and-qualitative-research>
- Foley, B. (2018). What is SPSS and How Does it Benefit Survey Data Analysis? Retrieved May 20, 2019, from Surveygizmo.com website: <https://www.surveygizmo.com/resources/blog/what-is-spss/>
- Forecasting853. (2018). Lecture 3: Moving Averages and Exponential Smoothing. *Sunlearn*.
- ForecastIT. (2010). ForecastIT 4. Holt's Exponential Smoothing. Retrieved June 7, 2019, from SlideShare.net website: <https://www.slideshare.net/ForecastIT/forecast-it-4-holts-exponential-smoothing-4370645>
- Fridstrøm, L., Østli, V., & Johansen, K. W. (2016). A stock-flow cohort model of the national car fleet. *European Transport Research Review*, 8(3), 1–15. <https://doi.org/10.1007/s12544-016-0210-z>
- Gakenheimer, R. (1999). Urban mobility in the developing world. *Transportation Research Part A: Policy and Practice*, 33(7–8), 671–689. [https://doi.org/10.1016/S0965-8564\(99\)00005-1](https://doi.org/10.1016/S0965-8564(99)00005-1)
- Gould, P. R. (1969). Spatial Diffusion. In *Association of American Geographers*.
- Heggie, I. G. (1979). Economics and road programme. *Transport Economics and Policy*, 13(1), 52–67.
- Hyndman, R. J., & Athanasopoulos, G. (2018). Holt's Trend Methods. In *Forecasting*:

- Principles and Practice* (second, pp. 191–201). Retrieved from http://course1.winona.edu/bdeppa/FIN335/Handouts/Exponential_Smoothing__part_2_.html
- IBM. (2019). Multinomial Logistic Regression. Retrieved June 26, 2019, from IBM Knowledge center website: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_sub/statistics_mainhelp_dita/spss/regression/idh_mnlr.html#idh_mnlr
- Ingram, G., & Liu, Z. (1997). Motorization and the Provision of Roads in Countries and Cities. In *Policy Research working paper series* (Vol. 1842).
- Kim, S.-H. (2011). *An analysis of household vehicle type acquisition using multinomial logit model* (University of Wisconsin-madison). Retrieved from https://minds.wisconsin.edu/bitstream/handle/1793/54503/Sangheun_Kim_report.pdf?sequence=1&isAllowed=y
- Krygsman, S., & Dijst, M. (2001). Multimodal Trips In The Netherlands: Conceptual Clarification, Micro-Level Individual Attributes And Residential Context. *Transportation Research Record*, 1753(01–3102), 11–19.
- LaerdStatistics. (2018). How to perform a Multinomial Logistic Regression in SPSS Statistics | Laerd Statistics. Retrieved December 11, 2019, from <https://statistics.laerd.com/spss-tutorials/multinomial-logistic-regression-using-spss-statistics.php>
- Letshwiti, V., Stanway, R. A., & Mokonyama, M. (2003). Strategic Overview and Analysis of Vehicle Ownership Trends in South Africa With an Emphasis on Light Passenger Vehicles. *22nd Southern African Transport Conference*, 1–9. Retrieved from [http://repository.up.ac.za/dspace/bitstream/handle/2263/7448/Letshwiti_Strategic\(2003\).pdf;jsessionid=3A2D0CFFFD30AB33B659C55208F916F5?sequence=1](http://repository.up.ac.za/dspace/bitstream/handle/2263/7448/Letshwiti_Strategic(2003).pdf;jsessionid=3A2D0CFFFD30AB33B659C55208F916F5?sequence=1)
- Lombard, M., Cameron, B., Mokonyama, M., & Shaw, A. (2007). *Report on Trends in Passenger: Transport in South Africa (Development Paper No. 174)*.
- Manski, C. F. (1983). Analysis of equilibrium automobile holdings in Israel with aggregate discrete choice models. *Transportation Research Part B*, 17(5), 373–389. [https://doi.org/10.1016/0191-2615\(83\)90004-8](https://doi.org/10.1016/0191-2615(83)90004-8)
- Marks, A. ., & Brown, R. . (1981). A new approach to car ownership forecasting in South Africa. *South African Institution of Civil Engineering*, 23(12), 575–582.
- Microsoft. (2019). Create a forecast in Excel for Windows - Excel. Retrieved May 16, 2019, from Support.office.com website: <https://support.office.com/en-us/article/create-a->

forecast-in-excel-for-windows-22c500da-6da7-45e5-bfdc-60a7062329fd

- Mokonyama, M., & Venter, C. (2007). Forecasting household car ownership in South Africa : alternative models and future trends. *South African Institution of Civil Engineering*, 49(3), 2–10.
- National Reserach Council. (2010). Persistent Forecasting of Disruptive Technologies. In *Persistent Forecasting of Disruptive Technologies* (pp. 8–16). <https://doi.org/10.17226/12557>
- Öğüt, K. S. (2004). S-Curve models to determine the car ownership in Turkey. *ARI Bulletin of the Istanbul Technical University*, 54(2), 65–69.
- Power BI Team. (2014). Describing the forecasting models in Power View | Microsoft Power BI Blog | Microsoft Power BI. Retrieved June 7, 2019, from Microsoft.com website: <https://powerbi.microsoft.com/en-us/blog/describing-the-forecasting-models-in-power-view/#ETSAAA>
- Prevedouros, P. D., & An, P. (1998). Automobile ownership in Asian countries: Historical trends and forecasts. *ITE Journal (Institute of Transportation Engineers)*, 68(4), 24–29.
- Romilly, P., Song, H., & Liu, X. (1998). Modelling and forecasting car ownership in Britain: A cointegration and general to specific approach. *Journal of Transport Economics and Policy*, 32(2), 165–185.
- RTMC. (2018). Vehicle population. Retrieved May 3, 2018, from http://www.rtmc.co.za/images/rtmc/docs/traffic_reports/population/Website Vehicle Population.xls
- SAHO. (2019a). Timeline Interim Constitution 1991 to 1993. Retrieved September 29, 2019, from South African History Online website: <https://www.sahistory.org.za/article/timeline-interim-constitution-1991-1993>
- SAHO. (2019b). Timelines | South African History Online. Retrieved June 11, 2019, from [sahistory.org.za](https://www.sahistory.org.za/timelines) website: <https://www.sahistory.org.za/timelines>
- Serag, M. S. (2014). Modeling car ownership in egypt. *The Egytian International Journal of Engineering, Science and Technology*, 17(2), 1812–1819.
- Shaygan, M., Mamdoohi, A., & Masoumi, H. E. (2017). Car ownership models in Iran: a review of methods and determinants. *Transport and Telecommunication*, 18(1), 45–59. <https://doi.org/10.1515/ttj-2017-0005>
- SPSS. (2015). IBM SPSS Forecasting. Retrieved May 16, 2019, from SPSS.com.hk website: <http://www.spss.com.hk/software/statistics/forecasting/>

- Statistics South Africa. (2014a). National Household Travel Survey: Statistical release. In *Statistics South Africa*. <https://doi.org/10.4135/9781483346526.n341>
- Statistics South Africa. (2014b). National Household Travel Survey 2013. In *Statistics South Africa*. <https://doi.org/10.4135/9781483346526.n341>
- Statistics South Africa. (2019). The Count & How it was Done. Retrieved June 12, 2019, from Statssa.gov.za website:
<https://apps.statssa.gov.za/census01/Census96/HTML/Metadata/Docs/count.html>
- The World Bank. (2017). GDP (constant LCU) | Data. Retrieved June 26, 2018, from worldbank.org website:
<https://data.worldbank.org/indicator/NY.GDP.MKTP.KN?locations=ZA>
- Train, K., & Lave, C. A. (1978). *A disaggregate model of auto-type choice*. 13(1), 1–9.
- Van Rensburg, J., & Krygsman, S. (2015). Running Dry: Assessing the fuel levy as a long term economically efficient road use fund. *34th Annual South African Transport Conference - Transport: Working Together to Deliver*. Pretoria.
- Van Rensburg, JA, & Krygsman, S. (2019). Funding for roads in South Africa: Understanding the principles of fair and efficient road user charges. *World Conference on Transportation Research - WCTR 2019 Mumbai 26-31 May 2019*.
- Venter, C. J., & Mohammed, S. O. (2013). Estimating car ownership and transport energy consumption: A disaggregate study in Nelson Mandela Bay. *Journal of the South African Institution of Civil Engineering*, 55(1), 2–10.
- Whelan, G. (2007). Modelling car ownership in Great Britain. *Transportation Research Part A: Policy and Practice*, 41(3), 205–219. <https://doi.org/10.1016/j.tra.2006.09.013>
- Wong, K. I., & Lin, H. (2011). Modeling Household Car and Motorcycle Ownership : A Case of Macao. *Eastern Asia Society for Transportation Studies*, 8.
- Worldbymap. (n.d.). Roadways / Countries of the World. Retrieved December 11, 2019, from <http://world.bymap.org/Roadways.html>
- Yang, Z., Jia, P., Liu, W., & Yin, H. (2017). Car ownership and urban development in Chinese cities: A panel data analysis. *Journal of Transport Geography*, 58, 127–134. <https://doi.org/10.1016/j.jtrangeo.2016.11.015>

APPENDIX A: Questions extracted from NHTS 2013

Province (Pr_code)

(@47 1.)

South African provinces according to the provincial boundaries as demarcated in December 2005.

Final code list

- 1 = Western Cape
- 2 = Eastern Cape
- 3 = Northern Cape
- 4 = Free State
- 5 = KwaZulu-Natal
- 6 = North West
- 7 = Gauteng
- 8 = Mpumalanga
- 9 = Limpopo

Geographical location (Type)

(@471 1.)

Derived variable:

Note to users

This variable is based on the Census 2001 typology. This typology was used to classify all enumeration areas (EA) and per implication the NHTS sampled PSUs into one of four classes, namely: urban formal, urban informal, tribal area and rural formal areas. Firstly these four classes were combined into two classes, namely urban (urban formal and urban informal) and rural (tribal area and rural formal).

Once classified into two groups, a set of new classifications based on a combination of the Municipal Demarcation Board's categorisation of metros and non-metros were applied to the urban category, resulting in three distinct categories: metro, urban (all non-metro urban) and rural.

Final code list

- 1 = Metro
- 2 = Urban
- 3 = Rural

Main Dwelling type (Q71MAIND)

(@28 2.)

7.1	<p>Indicate the type of main dwelling that the household occupies</p> <p>01 = Dwelling/house or brick/concrete block structure on a separate stand or yard or on farm</p> <p>02 = Traditional dwelling/hut/structure made of traditional materials</p> <p>03 = Flat or apartment in a block of flats</p> <p>04 = Cluster house in complex</p> <p>05 = Town house (semi-detached house in complex)</p> <p>06 = Semi-detached house</p> <p>07 = Dwelling/house/flat/room in backyard</p> <p>08 = Informal dwelling/shack in backyard</p> <p>09 = Informal dwelling/shack not in backyard, e.g. in an informal/squatter settlement or on farm</p> <p>10 = Room/flatlet on a property or a larger dwelling/servants' quarters/granny flat</p> <p>11 = Caravan/tent</p> <p>12 = Other</p>
-----	---

Notes to users

This question is about the dwelling type which the household occupies. The focus is on the main house the household occupies if there is more than one dwelling.

Universe

The respondent who is responding on behalf of the household.

Household income quintiles (quintile_rev)

(@263 1.)

Derived variable:

Note to users

Total monthly household income per capita is used to calculate income quintiles, i.e. the 20% of individuals with the lowest incomes (quintile 1), those between 20% and 40% (quintile 2), individuals between 40% and 60% (quintile 3), those between 60% and 80% (quintile 4) and finally the 20% of individuals who earn the highest incomes (quintile 5).

Final code list

- 1 = Lowest income quintile
- 2 = Quintile 2
- 3 = Quintile 3
- 4 = Quintile 4
- 5 = Highest income quintile

Total Household Expenditure (Q72EXP)

(@30 2.)

7.2	<p>What was the total household expenditure in the last month?</p> <p><i>Include money spent on food, clothing, transport, rent and rates, alcohol and tobacco, school fees, entertainment and any other expenses.</i></p> <p>01 = R0 02 = R1 – R199 03 = R200 – R399 04 = R400 – R799 05 = R800 – R1 199 06 = R1 200 – R1 799 07 = R1 800 – R2 499 08 = R2 500 – R4 999 09 = R5 000 – R9 999 10 = R10 000 or more 11 = Do not know 12 = Refuse</p>
------------	---

Notes to users

This refers to the total amount spent by the household in the last month. This includes money spent on food, clothing, transport, rent and rates, alcohol and tobacco, school fees, entertainment and any other expenses.

Universe

The respondent who is responding on behalf of the household.

Population group (RACE)

(@53 1.)

E	<p>What population group does belong to?</p> <p>1 = African/Black 2 = Coloured 3 = Indian/Asian 4 = White 5 = Other</p>
----------	--

Note to users

This question is asked to determine the population group of persons from the selected dwelling. The respondent must answer for each member without any assumptions. In this instance the enumerator is also instructed not to make any conclusions which may be influenced by his observation or using people's names during the interview. This question is important to measure progress in moving away from the effects of past discrimination. It is a sensitive question and some respondents might think it is racist and inappropriate in post-apartheid South Africa. If there is resistance to this question, explain to the respondent the importance for policy purposes of monitoring progress in the new political dispensation.

Universe

Every person who normally resides in this household at least four nights per week.

Income category (Q48CAT)

(@289 2.)

4.8 <i>Only if 'None', 'Refuse' or 'Do not know' in Q 4.6. Show the categories. Make sure the respondent points at the correct income column (weekly, monthly, annually) on prompt card 3 and mark the applicable code.</i>				
		Weekly	Monthly	Annually
	01	NONE	NONE	NONE
	02	R1–R46	R1–R200	R1–R2 400
	03	R47–R115	R201–R500	R2 401–R6 000
	04	R116–R231	R501–R1 000	R6 001–R12 000
	05	R232–R346	R1 001–R1 500	R12 001–R18 000
	06	R347–R577	R1 501–R2 500	R18 001–R30 000
	07	R578–R808	R2 501–R3 500	R30 001–R42 000
	08	R809–R1 039	R3 501–R4 500	R42 001–R54 000
	09	R1 040–R1 386	R4 501–R6 000	R54 001–R72 000
	10	R1 387–R1 848	R6 001–R8 000	R72 001–R96 000
	11	R1 849–R2 540	R8 001–R11 000	R96 001–R132 000
	12	R2 541–R3 695	R11 001–R16 000	R132 001–R192 000
	13	R3 696–R6 928	R16 001–R30 000	R192 001–R360 000
	14	R6 929 OR MORE	R30 001 OR MORE	R360 001 OR MORE
	15	DO NOT KNOW	DO NOT KNOW	DO NOT KNOW
	16	REFUSE	REFUSE	REFUSE

Note to users

This question is only applicable to those who answered with a 'None', 'Refuse' or 'Do not know' in Q4.6. The aim of this question is to establish the income of all household members who are economically active. Since this kind of information is very personal, the enumerator is instructed to inform the respondent that there is a Statistics Act which protects them against any disclosure. If a respondent works from home, the instruction is to skip to section 5 after answering Q4.8.

Educational Institution (Q33EDUII)

(@167 2.)

3.3	<p>Which of the following educational institutions does attend? <i>Read all the options, if more than one, record the MAIN institution.</i></p> <p>1 = Pre-school (including day care, crèche, pre-primary, ECD centre, nursery school) 2 = School (including Grade R/Grade 0 learners who attend a formal school) 3 = Adult Basic Education and Training Learning Centre (ABET Centre) 4 = Literacy classes (e.g. Kha Ri Gude) 5 = Higher educational institution (University/University of Technology) 6 = Further Education and Training College (FET) 7 = Other college 8 = Home-based education/home schooling 9 = Other than any of the above</p>
------------	--

Note to users

This question is only applicable to those who answered 'Yes' in Q3.2. It is also instructed from the questionnaire that both distance and correspondence education must be considered in this question. There is always a tendency of taking people's ages and relating them to the level of education. The enumerator is therefore instructed to probe for such information when given by the respondent, especially in cases of proxies.

Universe

All household members who are currently attending educational institutions.

Highest Grade (Q31HIEDU)

3.1	<p>What is the highest level of education that has successfully completed?</p> <p><i>Diplomas or certificates must be of six months plus study duration full-time (or equivalent) to be included</i></p> <p>98 = No schooling 00 = Grade R/00 01 = Grade 1/Sub A/Class 1 02 = Grade 2/Sub B/Class 2 03 = Grade 3/Standard 1/ABET 1 (Kha Ri Gude, Sanli) 04 = Grade 4/Standard 2 05 = Grade 5/Standard 3/ABET 2 06 = Grade 6/Standard 4 07 = Grade 7/Standard 5/ABET 3 08 = Grade 8/Standard 6/Form 1 09 = Grade 9/Standard 7/Form 2/ABET 4 10 = Grade 10/Standard 8/Form 3 11 = Grade 11/Standard 9/Form 4 12 = Grade 12/Standard 10/Form 5/Matric (No Exemption) 13 = Grade 12/Standard 10/Form 5/Matric (Exemption *) 14 = NTC 1/N1/NC (V) Level 2 15 = NTC 2/N2/NC (V) Level 3 16 = NTC 3/N3/NC (V) Level 4 17 = N4/NTC 4 18 = N5/NTC 5 19 = N6/NTC 6 20 = Certificate with less than Grade 12/Std 10 21 = Diploma with less than Grade 12/Std 10 22 = Certificate with Grade 12/Std 10 23 = Diploma with Grade 12/Std 10 24 = Higher Diploma (Technikon) 25 = Post Higher Diploma (Technikon Master's, Doctoral) 26 = Bachelor's Degree 27 = Bachelor's Degree and post-graduate diploma 28 = Honours Degree 29 = Higher degree (Master's, Doctorate) 30 = Other (specify in the box below) 31 = Do not know</p>
------------	---

Note to users

This question is applicable to all household members. The enumerators are instructed that only those qualifications already obtained must be considered for household members. That means the current level, whereby a person is still busy with the studies, is not applicable. It is very important to complete each record even if the person has not attended school. Moreover, the enumerators are instructed that diplomas and certificates must be of at least six months' duration.

Universe

All household members in the selected dwelling unit.