APPLICATION OF SMALL AREA ESTIMATION TECHNIQUES IN MODELLING ACCESSIBILITY OF WATER, SANITATION AND ELECTRICITY IN SOUTH AFRICA: THE CASE OF CAPRICORN DISTRICT

by

RESHOKETSWE MOKOBANE

THESIS

Submitted in fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS

in the

FACULTY OF SCIENCE AND AGRICULTURE
(School of Mathematical and Computer Sciences)

at the

UNIVERSITY OF LIMPOPO

PROMOTER: Prof. M. Lesaoana
CO-PROMOTER: Prof. P. Lahiri (University of Maryland, USA)

2019

Declaration

I, **Reshoketswe Mokobane**, declare that the thesis which is hereby submitted for the qualification of Doctor of Philosophy in Statistics at the University of Limpopo, is my own independent work and has not been handed in before for a qualification at/in another University/Faculty/School. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references. I further cede copyright of the thesis to the University of Limpopo.

Signature..

Date: 10 April 2019

Abstract

This study presents the application of Direct and Indirect methods of Small Area Estimation (SAE) techniques. The study is aimed at estimating the trends and the proportions of households accessing water, sanitation, and electricity for lighting at small areas of the Limpopo Province, South Africa. The study modified Statistics South Africa's General Household Survey series 2009-2015 and Census 2011 data. The option categories of three variables: Water, Sanitation and Electricity for lighting, were re-coded. Empirical Bayes and Hierarchical Bayes models known as Markov Chain Monte Carlo (MCMC) methods were used to refine estimates in SAS. The Census 2011 data aggregated in 'Supercross' was used to validate the results obtained from the models. The SAE methods were applied to account for the census undercoverage counts and rates. It was found that the electricity services were more prioritised than water and sanitation in the Capricorn District of the Limpopo Province. The greatest challenge, however, lies with the poor provision of sanitation services in the country, particularly in the small rural areas. The key point is to suggest policy considerations to the South African government for future equitable provisioning of water, sanitation and electricity services across the country.

Keywords: Small Area Estimation, Basic services, Water, Sanitation, Electricity, Household, Accessibility, Census data, Hierarchical Bayes.

Dedication

I dedicate this work to my parents, Mokgadi and Mmotle Nkadimeng

Acknowledgements

It is a genuine pleasure to express my deep sense of appreciation and gratitude to my supervisor, Prof. 'Maseka Lesaoana. Her timely and thorough scrutiny, her scientific instructive advices, which were never missed, have helped me to a great extent to accomplish this task. Her scholarly guidance for my thesis led to the completion of the work; not excluding my co-supervisor, Prof Partha Lahiri. I am extremely thankful to Mr. Olebogeng Mfoloe who assisted me with SAS manipulation techniques; and Dr. Alexander Boateng who taught me how to compile my thesis into Latex. I also wish to extend my thanks to my friend, Rolly Lerutla, who frequently uttered religious messages of encouragement – not forgetting all who contributed to the completion of this thesis. I acknowledge with deep sense of gratitude, the love, inspiration and moral support received from my husband, Phuti, and my daughter, Amogelang.

Contents

List of Figures

List of Tables

List of Special Symbols

p_{ij}	Probability of selecting individual i in area j
w_{ij}	Weight of individual j in area i
S_i^2	Variance of the sample obtained from area i
X	Covariates matrix
β	Regression coefficients of vector X
Z	Known structure matrix of area random effects
ε	Random effects vector due to small area
e_{ij}	Random errors vector associated with sampling error
$\tilde{\beta}$	Best linear unbiased estimator (BLUE) of β
$\hat{\theta}_i^B$	Bayes Estimator
$\hat{v}$	Best Linear Unbiased Prediction estimator
$k(\theta\|x)$	Conditional probability distribution function of Θ given the sample X
N_i	Number of individuals in area i
y_{ij}	Unit responses assumed to be related to the auxiliary values x_{ij}
$\bar{X}$	Vector with the area mean values of the covariate
$\hat{\bar{Y}}_{i,COMPOSITE}$	Weighted sum of the Direct and Synthetic estimators
$\hat{\bar{Y}}_{i,GREG}$	Generalised regression(GREG) estimator produced on a linear predictor
$\hat{\bar{Y}}_{i,SYNTHETIC}$	Synthetic estimate in area i

List of Acronyms

AEMSE	Average Empirical Mean Square Error
ANC	African National Congress
BLUP	Best Linear Unbiased Prediction
CCD	Census Collection District
CD	Capricorn District
COSATU	Congress of South African Trade Union
DWAF	Department of Water and Forestry
EA	Enumeration Area
EB	Empirical Bayes
EBLUP	Empirical Best Linear Unbiased Prediction
EMSE	Empirical Mean Square Error
GHS	General Household Survey
GLM	General Linear Mixed models
GREG	Generalised Regression Estimator
HB	Hierarchical Bayes
HH	Household
HT	Horvitz Thompson estimator
LED	Living Environment Deprivation
LP	Limpopo Province
MCMC	Markov Chain Monte Carlo
MDGs	Millennium Development Goals
ML	Maximum Likelihood

WSE	Water, Sanitation and Electricity for lighting
MSA	Municipal System Act
MSE	Mean Square Error
NGDP	National Growth and Development Plan
PCI	Per Capita Income
PIMD	Provincial Indices of Multiple Deprivations
PSU	Primary Sampling Unit Number
RDP	Reconstruction and Development Programme
RELM	Residual Maximum Likelihood
SA	South Africa
SACP	South African Communist Party
SAE	Small Area Estimation
SAIMD	South African Index of Multiple Deprivations
SANCO	South African National Civic Association
SAS	Statistical Analysis System
SEIFA	Socio-Economic Indicator for Areas
SNA	State of the Nation Address
SOWETO	South Western Townships
Stats SA	Statistics South Africa
TBVC	Transkei, Bophuthatswana, Venda and Ciskei
UDM	United Democratic Front
US	United States

Research Outputs

The following sections give a list of research outputs from this thesis.

Peer Review Journal Publications

1. Mokobane, R., Lesaoana, M. and Lahiri, P. (2017). Exploring service delivery of water, sanitation and electricity by South African households using Empirical Bayes and Hierarchical models. (Under review).

2. Mokobane, R., Lesaoana, M. and Lahiri, P. (2017). Water and sanitation poverty: Application of small area estimation (SAE) (Under Construction).

3. Mokobane, R., Lesaoana, M. and Lahiri, P. (2017). Small Estimation Area Models (SAE): Provincial sanitation in South Africa (Under Construction).

Conferences and Related Events

1. World Social Science Forum. DST Funded event to interact and collaborate with international experts. 13-17 September 2015.

2. Training in Small Area Estimation methods, Statistical Packages, R and WinBugs. VLIR funded trip to Belgium to interact and collaborate with

international experts, hosted by Universiteit Hasselt, 07-27 June 2015, Brussels, Belgium.

3. Mokobane, R., Lesaoana, M. and Lahiri, P. (2015). Application of small area estimation methods in modelling service delivery at ward level in South Africa: Water, sanitation and electricity. Faculty of Science and Agriculture Postgraduate Research Conference (FSA-PGD 2015), 1-2 October 2015, Bolivia Lodge, Polokwane, South Africa.

4. Mokobane, R., Lesaoana, M. and Lahiri, P. (2013). Application of small area estimation methods in modelling service delivery at ward level in South Africa: Water, sanitation and electricity. Statistics South Africa Conference, hosted by University of Limpopo (SASA 2013), 28 November to 2 December 2013, Polokwane, South Africa.

5. Mokobane, R., Lesaoana, M. and Lahiri, P. (2016). Application of small area estimation methods in modelling service delivery at ward level in South Africa: Water, sanitation and electricity. Statistics South Africa Conference, hosted by University of Cape Town (SASA 2016), 28 November to 2 December 2016, Polokwane, South Africa.

6. Mokobane, R., Lesaoana, M. and Lahiri, P. (2017). Application of small area estimation methods in modelling service delivery at Municipality level in South Africa: Water, sanitation and electricity. STEM Conference, hosted by Howard University, 24- 27 October 2017, Washington DC, USA.

Chapter 1

Introduction

Chapter 1 presents an overview profile of South Africa disaggregated by province, focusing on small areas of the Capricorn District in Limpopo Province, South Africa. Facts related to service delivery on water, sanitation and electricity for lighting; and the rationale behind the study are included. The aim and objectives of the study are also specified in this chapter.

1.1 Background

1.1.1 Delivery of Basic Services in South Africa

The study applied Small Area Estimation (SAE) techniques to explore the trends and proportions of households accessing water, sanitation and electricity for lighting at small area levels in South Africa (SA) over a period of seven years: 2009-2015. Farley (1984); and Danzinger and Weinberg (1987) established that the South African government was experiencing challenges in delivering basic services such as water, sanitation and electricity to its citizens.

Hacker (1992) established that failure to deliver basic services in SA influenced poor people to demand services through protests in order to bridge the inequality gap between Blacks and Whites (Nattrass and Seekings, 2001).

Several scholars, including Cotter et al. (1999); Hulme and Shepherd (2003); Mattes et al. (2003); and Noble et al. (2009) found that the geographical position of a household plays a major role in the accessibility of resources in SA (Hacker, 1992). Christopher (2001) argued that the evolution of service delivery protests in SA was demarcated by segregation of former homelands. Similarly, Botting et al. (2010) found that the delivery of basic services in SA focused mainly on households in urban locations, excluding small rural areas (Earle et al., 2005).

It is, therefore, critical that research from different perspectives, be undertaken to study characteristics among smaller communities in order to seek solutions that would eventually alleviate the problems associated with the delivery of basic services in SA. This study is intended to apply statistical tools, specifically small area estimation techniques to investigate the accessibility of three key basic services, i.e., water, sanitation and electricity for lighting.

Escalating protests in the democratic SA, according to Adelzadeh and Padayachee (1994), are believed to have been sparked by the Reconstruction and Development Programme (RDP) that has raised expectations on the provision of jobs, equitable distribution of income, wealth and basic services. Similarly, Adelzadeh (1996) concluded that protests in SA signified a deviation from the RDP goals, while Roux et al. (1996) found that the release of Growth, Employment and Redistribution (GEAR) deviated from the RDP. Beside the RDP, the supply of clean water, sanitation and electricity is prioritised in the South African Constitution Act No. 108 of 1996 (Roux et al., 1996).

Ntsala and Mahlatji (2016) found that service delivery protests in SA led to burning of educational institutions such as libraries and schools – as evidenced from protests in Vuwani, Limpopo Province (LP), where communities burnt at least 20 schools in the area in 2016. In agreement with other scholars, Alexander et al. (2013) found that service delivery protests have increased in the post-Apartheid era, i.e. since 1994. To the contrary, the State of the Nation Address in SA claimed that citizens protest because they are impatient to wait for their turn of service delivery (Surender, 2014).

1.1.2 Water, Sanitation and Electricity Access in SA

Throughout the study, the accessibility of the three basic services (variables) by households in SA is defined differently from that of Statistics South Africa (Stats SA) as follows: Water refers to the treated drinkable tap water accessed from the tap inside the house, within the yard and outside the yard, i.e., healthy drinkable water; Sanitation means a flushing toilet connected to the sewerage system which uses water to dispose human waste, i.e., a flushing toilet is considered a toilet and a non-flushing toilet (pit latrine, bucket and other) as No toilet; Electricity means the electricity for lighting.

This study assumes that South Africans who regularly use electricity for cooking are likely to afford electricity for other services like heating and/or business, but that poor people will restrict the usage of electricity to lighting only. The research contributes by closing the gap on the lack of data at small geographical level in SA, and considered modified SAE models to fit the concepts of Big Data as outlined in Marchetti et al. (2015).

The study considered Solar energy as electricity for lighting. The challenge is that Solar panels are expensive. Donev et al. (2012) found that thermal demand of Solar water heating in SA for 20 years was estimated to 2.2 EJ, its implementation would provide 1.3 EJ clean energy till 2030 at 369 billion but would save 231 billion rand and reduced Carbon dioxide emissions by 297 metric tons. Ideally, this cannot be applicable to most of households considered in this study, since most of them use electricity for lighting only. Martinot et al. (2001) established that Solar marketing campaigns consumed high costs and time in rural areas of developing countries. The Republic of Kiribati, formerly known as the Gilbert Islands and other countries in the pacific used Solar photovoltaic (PV) technologies for lighting and electricity. It was established that it made insignificant (less than 1%) contribution to the total annual primary energy supply (Mala et al., 2009).

Devenish (1998) and Botting et al. (2010) viewed clean water and proper sanitation as basic services as defined in the South African Constitution Act No. 108 of 1996. Since there are several services which should be rendered to households by municipality authorities, this study focused on the accessibility of water, sanitation (toilet) and electricity for lighting. These services are globally pronounced number 6 and 7 in the Sustainable Development Goals (SDGs), and Le (2015) relates them to SDG 3.

The accessibility of clean water and improved sanitation is internationally associated with prevention of adverse health outcomes related to different diseases caused by unhygienic conditions. This study will assist the South African government in guiding them when rendering services equitably to citizens; for example, delivery of free basic services, such as, housing, water, sanitation, electricity, etc., to indigent households.

1.1.3 Water, Sanitation and Electricity Prescripts in SA

Adelzadeh and Padayachee (1994) found that the RDP pinned the South African government to the goal of supplying every citizen with clean water. After 1994, Malzbender et al. (2005) established that citizens are entitled to 25 litres of water per person per day or 6000 litres per household a month, but free of charge for indigent households only. However, Earle et al. (2005) established that communities spent several months without water and proper sanitation.

During the year 2000, the Department of Water Affairs (DWAF), was tasked with ensuring that all South Africans have equitable sanitation and supply of 20-30 litres of water. In complying, DWAF reviewed the basic water supply in 2002 as 25 litres per person per day within 200 metres from the household. The Regulation 3(b) of Gazette No. 22355 of 8 June 2001, states that there should be water supply of at least 25 litres per person per day or 6000 litres (6 kilolitres) per household per month, flowing at the rate of at least 10 litres per minute within 200 metres of the household. Gazette No.22355 emphasises that consumers should not spend seven days in succession without water supply in a year.

The right to sanitation was less prominent in the South African Constitution and the 1994 White Paper than in the documented prescripts of water supply. Craythorne (2006) established that sanitation was prioritised in Section 73 of the Municipal System Act (MSA) of 2001, giving the right to basic municipal services. In the South African Constitution, sanitation is also protected by Regulation 2 of the Water Service Act. The acceptable sanitation is a toilet that is harmless, clean and hygienically safe to prevent swam of flies and other disease-carrying pests. The 2001 MSA encourages municipalities to update the indigent register for proper allocation of free basic services to the deserving households.

Electricity in this study has been added as a basic household service. The essential services can be referred from the South African Constitution Act No. 108 of 1996. The essential services include Housing, Education, Health care, Social welfare, Transport, Water, Electricity and Energy, Sanitation, and Refuse and Waste removal.

There is a gap in specifying the right to electricity in the Constitution of SA. The Constitution specifies that electricity should be sufficient to provide for lighting, basic media access, water heating and other basic domestic services for poor households. Municipalities realised while rendering these basic services that poverty and unemployment contributed to an increase in unpaid electricity bills. Free access to electricity was restricted because many people could not afford the costs. Lodge (2003) estimated that there were 19000 schools and 4000 clinics among Black/African communities without electricity for lighting in rural places of South Africa.

The above realisations point to communities among mostly rural and impoverished areas in South Africa. These communities continue to be deprived from the delivery of vital basic services, hence the need for research in similar matters.

1.2 SAE Definition and its Applications

Small Area Estimation (SAE) is one of several statistical techniques for the estimation of parameters for small sub-populations. This technique is applied when the sub-population of interest is included in a larger survey, i.e., the small area in this study refers to a small geographical area such as a county, municipality or group of villages. A domain is a specific statistical study of population within an area.

The small area refers to a small domain, a small territory or region as a subset of the big area. For example, when national surveys are conducted for the whole population across the country, in some cases, there exists a specific small area whose sample size is too small to generate precise estimates from the country data. This problem is solved by using additional data such as administrative records that are available for these small areas.

As the need for reliable estimates at local places arose, the evolvement of SAE methods increased. Old and new developments in SAE methods are outlined in Pfeffermann (2002) and Pfeffermann (2013). A comprehensive SAE book by Rao and Molina (2015) that discusses the advantages and limitations in the application of various SAE approaches, emphasises the importance of real data. Guadarrama et al. (2016) applied SAE methods to poverty mappings. This study learnt from Battese et al. (1988) to compute estimates of Mean Squared Error considering the uncertainty involved in estimating the variance. The insight laid by Rao (2003) motivated this study to apply GHS and Census data, which have different sampling designs, through Horvitz and Thompson (1952) in estimating basic household services in SA.

1.2.1 SAE Evolvement Over Years

The literature indicates that the SAE techniques has usefully evolved over years. Fay and Herriot (1979) applied it to estimate the Per Capita Income (PCI) for several small places. In the same year, Holt et al. (1979) obtained Best Linear Unbiased Predictor (EBLUP) estimator of the finite population. Ericksen and Kadane (1985) applied weighted averages and Synthetic regression using the Fay and Herriot (1979) undercounts to adjust population counts from U.S. Census 1980 and beyond. As more interest and need for reliable estimates arose, Census was conducted in ten year periods in U.S.

Battese et al. (1988) obtained EBLUB estimates of Corn and Soybeans, adding strength to data using LANDSAT satellite data, applying nested error regression models involving random small area effects. Later, Ericksen et al. (1989) recommended the application of regression model for areas without sampled data; while Cressie (1989) applied the SAE methods on cancer mortality and undercounts in USA, and showed how standard data analytic techniques can be modified through spatial link. Sarndal and Hidiroglou (1989) applied SAE methods for estimation of wages and salaries of units for each census division in a province using business income as the auxiliary variable with known population mean.

Datta and Ghosh (1991) applied the Hierarchical Bayes approach to Battese et al. (1988)'s research and got similar results. Freedman and Navidi (1992) applied SAE methods in critiquing the work by Ericksen et al. (1989) for poor SAE model validation and large sampling errors in the estimates. Ghosh and Rao (1994) cited an example of SAE application to the Federal State Cooperative program initiated by the U.S. Bureau of Census in 1967 for high quality and consistent series of county population estimates, which led to the production of sub-county estimates.

Pfeffermann and Barnard (1991) derived estimators for the prediction variances by replacing unknown variances by maximum likelihood estimates in the regression formulae that account for State effects and nested domain effects. This was computed through SAS iterations for the assessment of farm values. The SAE techniques were also applied in Canada to produce monthly estimates of unemployment rates (You et al., 2003) and youth smoking patterns (Pickett et al., 2000) at national, provincial and to smaller area levels. Minot and Baulch (2005) used census and budget (borrowing strength) data to establish poverty distribution in Vietnam for food policy; and found that their poverty estimates were not closely correlated to estimates applied by government.

Having learned how SAE techniques have evolved over the years, we note that several scholars around the globe bought into these models and applied them from different research perspectives, yet little was done in modelling the delivery of basic services to communities. In addition, SAE methods were applied minimally in some developing countries, especially in South Africa and other African countries. We next review applications of SAE models in some African countries.

Elbers et al. (2003) and Levine and Roberts (2013) estimated poverty and inequality in South Africa and Namibia, respectively. Kipruto et al. (2015) applied SAE spatial modelling to estimate Tuberculosis in Kenya. The SAE methods were also applied by Mercer et al. (2015) on household data to construct the sub-national estimates of child mortality in Tanzania through space-time smoothing method.

1.2.2 Types of Models Considered

It is assumed that:

- This study is the first in applying SAE models to explore service delivery change at municipality level in SA.
- There is little literature on service delivery change at small areas in South Africa.

- Applying SAE to estimate the accessibility and provision of three basic services at small places makes this study unique in SA.

In terms of models, we assumed normality of the estimates as this was not as restrictive as the normality of the random effects computed in the study. The study applied the central limit theorem on estimates as was considered by Ghosh and Rao (1994) for a similar challenge, and Freedman and Navidi (1986) for the adjustment of under-cover (under-enumeration). It was traditionally assumed that the sampling variance is known. However, specific assumptions are elaborated per model approach in Chapter 3.

SAE methods are categorised into direct and indirect approaches. Table 1.1 lists the types of model approaches applied in this study.

Table 1.1: Direct and indirect SAE techniques

Direct estimation	**Indirect estimation**	
Horvitz-Thompson (H-T) estimator	Explicit models	Implicit models
GREG estimator	-	-
-	Unit level	Synthetic estimator
-	Area Level	Composite estimator
-	GL Mixed	Demographic estimator
-	E-BLUP estimator	
-	EB estimator	
-	HB estimator	

1.3 Historical Developments

The democratic South Africa is demarcated into nine provinces as shown in Table 1.2. The table also provides an overview of the South African population by province, according to the three democratic population censuses undertaken to date (1996, 2001 and 2011). The South African population has increased from

Table 1.2: SA population by province, Censuses 1996, 2001 and 2011

Province	**Census 1996**	**Census 2001**	**Census 2011**
Western Cape	3956875	4524335	5822734
Eastern Cape	6147244	6278651	6562053
Northern Cape	1011864	991919	1145861
Free State	2633504	2706775	2745590
KwaZulu Natal	8572302	9584129	10267300
North West	2727223	2984098	3509953
Gauteng	7834125	9388854	12272263
Mpumalanga	3123869	3365554	4039939
Limpopo	4576566	4995462	5404868
South Africa	40583573	44819778	51770560

Source: Statistics South Africa Census 2011.

40.58m in 1996 to 51.77m in 2011. The Gauteng Province (with the smallest area size) overtook KwaZulu-Natal in 2011 to record the highest population. The Northern Cape (with the largest area size) remains with the smallest population, and it is the province that showed a decline in population growth (from 1996 to 2001). While the population of Limpopo was the 4th highest in 1996 and 2001, it was overtaken by the Western Cape in 2011.

1.3.1 Definition of Small Areas: South African Context

Each one of the nine provinces in SA is divided into district municipalities (or simply districts), and each district has a number of local municipalities (or simply municipalities). Each municipality is divided into a number of wards. Small places known as villages comprise a ward. The area demarcations are mutually exclusive and constitutionally dependent.

This study selected the Capricorn District (CD) as a case study because it has a mixture of urban, semi-urban, rural and slum settlements. The CD is also included in the former homeland of Lebowa. It is named after the Tropic of Capricorn which runs through it. The CD is situated in the centre of the LP, and consists of five municipalities, namely: Aganang, Blouberg, Lepelle-Nkumpi, Molemole and Polokwane. There are 113 wards in the CD. Polokwane, one of the five municipalities of the CD, is the capital city of the Limpopo Province, and has the highest number of wards (37), followed by Lepelle-Nkumpi (27), Aganang and Blouberg with 18 wards each. Molemole has the least number of wards (13). Polokwane city is situated at the core of economic development in the LP. The CD shares borders with all the four other district municipalities in the Province: Mopani on the east, Sekhukhune on the south, Vhembe on the north and Waterberg on the west. This latter feature makes the CD truly central and representative of the Province.

1.3.2 Small Areas in the Limpopo Province

Table 1.3: Population of Limpopo Province by district and gender

District	Male	Female	Total	% population
Capricorn	618709	699776	1318485	23.0
Mopani	532778	613794	1146572	20.0
Sekhukhune	539921	614742	1154663	20.2
Vhembe	633730	740727	1374457	24.0
Waterberg	374388	358327	732615	12.8
Total	2699426	3027366	5726792	100.0

Source: Statistics South Africa, Census 2011.

Table 1.3 shows the population distribution of the LP across all the five districts by sex. Vhembe district accounts for 24.0% of the total population of the LP, followed closely by the Capricorn District with 23.0%, then Mopani and Sekhukhune with about 20% each. Waterberg is the only district which has relatively small population and fewer numbers of females than males. Waterberg district is predominantly farmland. Vhembe shares the border with Zimbabwe.

1.4 Small Areas: GHS Context

Table 1.4: Population and Area Size of the CD by Municipality

Mpal	**Area**	**L.Size**	**Wards**	**M**	**F**	**Tot. Pop.**	**% Mpal**
Aganang	185222	10.9	18	75474	88546	164020	12.8
Blouberg	454084	26.8	18	81937	97206	179143	13.9
Lepelle-Nkumpi	345478	20.4	27	115901	137261	253162	19.8
Molemole	334725	19.7	13	56472	65093	121565	9.5
Polokwane	377521	22.2	37	268391	296446	564837	44.0
Capricorn	1697030	100.0	113	598175	684552	1282727	100.0

Source: CD Spatial Development Framework, Statistics South Africa 2011.
Note: F and M denote number of females and males, respectively; Tot.Pop denotes total population; L.Size is the % of occupied land; Area is specified in $1000km^2$; % Mpal is the % of population in each municipality.

Table 1.4 shows the number of wards and population in each municipality of the CD by sex. The table shows that Polokwane municipality has the largest number of wards and the largest population (44.0%); and covers the second largest land area $377521Km^2$ after Blouberg ($454084Km^2$). Thus, Polokwane is densely populated and has the smallest ratio of males to females, which is typical of most urban and capital cities. Molemole municipality has the least population in the CD.

1.5 Problem Statement

1.5.1 Ideal World

Ideally, national surveys should contain enough population information of any geographical domain. Researchers and data agencies would produce reliable estimates that are precise and would require no further modelling. There would be no data problems such as less sampled Primary Sampling Unit (PSUs) in each small area which results in direct variance estimates that are not credible due to lack of precision. Raghunathan et al. (2007) argued that although statistics modelling is mostly used, the difference made by survey researchers among model-based, model-assisted, and design-based techniques does not necessarily help.

The major problem is that there is limited research on social statistics in South Africa for the application of SAE methods, especially in modelling the accessibility of basic services. The SAE methods were initially used by international farmers to model corn or soybean (Battese et al., 1988) and in order to compute the allocation of resources by the states (Smith et al., 2002).

To our knowledge, this study is likely to be the first to apply SAE methods to service delivery challenges in South Africa. This challenge is currently evidenced by lack of direct references to similar applications. The main question is: how can the South African Government estimate, monitor and evaluate the proportions of households that are accessing/not accessing basic services such as water, sanitation and electricity, for planning? It is the purpose of this study to address some of these challenges.

1.5.2 Some Realities About SAE Techniques

Usually, in SAE application, one has to use either the direct estimates or the model-based estimates. Little (2006) argued that hierarchical models are a solution for these problems because they provide links between the direct estimate from the saturated model and the model-based estimate from the unsaturated model. In this study Bayes modelling has been used to solved the problem of variance components.

The other reality is that the model-based estimates in small areas turn to compromise confidentiality and local individuals who know and reside in a specific small area may challenge estimates if there is a lack of precision (Brackstone, 2003).

This study has presented the application of direct and indirect techniques of SAE methods to establish service delivery in water, sanitation and electricity for lighting. The research focused on the Capricorn District in the Limpopo Province, one of the nine provinces of SA.

Empirical Bayes and Hierarchical Bayes models known as Markov Chain Monte Carlo (MCMC) methods were estimated using a statistical package for data analysis, SAS. A system called Supercross was used to validate the results at small geographical areas. The Census 2011 undercounts of the Capricorn District, one of five districts of the Limpopo Province (LP), were established.

1.5.3 Challenges on the SAE Techniques

The difficulty in producing precise direct estimates for small areas due to inadequate or missing information about the population in small areas, poses a challenge in case of planning and provision of services within the boundaries. For example, if a type of disease breaks in one small area, the out-break could be quickly controlled prior to its spread to other adjacent areas if adequate information is available.

Since national surveys require costly resources, thorough planning and time, the study modified the South African General Household Survey (GHS) data for the years 2009-2015 and Census 2011 by re-coding the option categories of

Water, Toilet and Electricity for lighting as three variables, while Household (HH) has been considered as the unit of analysis. The GHS and Census 2011 data sets were obtained from Statistics South Africa (Stats SA).

The study was triggered by the fact that Stats SA does not disseminate the GHS information at small areas such as ward level. The ultimate key point is to suggest policy considerations to the South African government for future equitable provisioning of water, sanitation and electricity services across the country.

1.5.4 Proposed Statement

This study noted that modelling framework is an important tool to explore estimates in the application of SAE methods. The study, therefore proposes that policy makers and researchers need to apply quality spatial data. Kalton (1983) found that the main problem arises when applying SAE models on data from small areas.

1.6 Rationale

The study has applied small area estimation methods to assess the service delivery of three basic services to South African households. In particular, to the estimation of proportions of households that are accessing/not accessing water, sanitation and electricity for lighting at small areas for planning purposes, i.e., the study has used the South African Census 2011 data to compute unbiased estimators. The type of sanitation (pit toilet or flush toilet), the type of water and its source and access to electricity for lighting, were of interest in this study.

Small area estimation methods became useful census data in SA contain more detailed information on limited variables, and a census is conducted once in ten years – which is long enough for the demographic dynamics to change. In mitigating this challenge, surveys are conducted with limited costs. It is also the aim of this study to address these shortcomings.

Surveys fail to cover small areas, or indeed small areas lack enough data. In SAE techniques, sample surveys are used to provide estimates for large areas or domains (Jiang and Lahiri, 2006). The rationale of this study was to suggest policy considerations to the South African government for future equitable provisioning of proper water, modern sanitation and electricity services.

1.7 Aim and Objectives of the Study

1.7.1 Aim

The aim of the study is to explore the accessibility of water, sanitation and electricity for lighting at small area level in South Africa using the Capricorn District of the Limpopo Province as a case study.

1.7.2 Objectives

The objectives of the study are to:

1. Evaluate small area estimation models.
2. Re-code the data to establish the definition of water, sanitation (toilet) and electricity for lighting, relevant to the study conducted.
3. Apply small area estimation models to compute Empirical Best Linear Unbiased Predictor of household-size accessing/not accessing services.

4. Estimate proportions of households accessing water, sanitation (toilet) and electricity at municipality level.
5. Determine the access variability of water, sanitation and electricity for lighting at municipality level.
6. Establish which service among water, sanitation (toilet) and electricity is mostly accessed by households in the Capricorn District.
7. Compute the Census 2011 undercoverage rates and counts that have not been accounted for in the data collection for CD municipalities.
8. Compare the undercoverage rates of CD municipalities by rate type (Direct, Hierarchical, standard and covariate rates).
9. Suggest policy considerations to the South African government for future provisioning of water, sanitation and electricity services.

1.8 Scope of the Study

The study considered services rendered to households in the Limpopo Province. It has focused on exploring the accessibility of water, sanitation and electricity for lighting at small area level of the Capricorn District, one of the five districts in the Limpopo Province.

1.9 Limitations of the Study

Demarcations of the land in SA are inter-overlapping:

- Currently, demarcations among urban, rural, semi-urban and semi-rural areas are intertwined in SA.

- The demarcations by wards for the 2009-2015 data sets may not be consistent throughout due to administrative re-demarcations during this period.
- The indigent register that should be kept and updated by municipalities is still in a fragmented state. Such register would have been useful in borrowing strength (technique in SAE methods) and validating some of the results obtained by this study.
- Unavailability of documented basic service delivery information was a challenge in rural areas across the Capricorn District.

1.10 Significance of the Study

The study modified small area estimation techniques to explore the provision of water, sanitation and electricity at municipality level, which can be replicated to any big area such as provincial or state.

The South African government will be convinced of the usefulness of SAE models by scholars who would have applied these models for equitable provisioning of services at small areas. The estimated proportions of households accessing or not accessing water, sanitation (toilet) and electricity at a small geographical area, will directly assist the Capricorn District in service management and developmental planning.

More importantly, the study would also raise awareness to public data agencies to apply SAE techniques when processing their data. These agencies would start to avail data at point/small area level. The results of the study would contribute to the policy formulation aligned to social rights entitled to communities which are deprived of basic services.

1.11 Organisation of the Work

The study addresses the gap of inadequate research in the area of SAE application in estimating service delivery at local municipality and sub-place levels in South Africa. This study is the first to modify Census 2011 data released by Stats SA to service delivery at household level.

In Chapter 1, the study presents the prescripts and factual background on the accessibility of water, sanitation and electricity for lighting (3 variables). The definition and evolvement of SAE methods were presented in relation to the objectives and the problem statements which were achieved through the application of different types of SAE models listed in Chapter 1. The hierarchy levels of domains in SA to the focal point level, i.e., the Capricorn District, were also presented in Chapter 1.

Chapter 2 outlines sampling design and how the data was computed. This is the data description of the Census 2011 and GHS series. Modification of Stats SA data by re-coding and coding into SAS was added. The exercise assisted this study in defining water, sanitation and electricity for lighting properly in chapter 2. The literature review on SAE methods are presented in Chapter 3. The evaluation, assumptions and the application of SAE models are also presented in Chapter 3.

Chapter 4 presents the descriptive proportions of households accessing or not accessing services relating to the three variables of interest at country level and district municipality (CD) level. The results derived from the GHS series, Geotype and Census data are presented in this chapter. Trends on Geotype data were presented for the Limpopo Province to explore the service delivery change in the three variables, only for the years 2009, 2014 and 2015. The Piped (tap) water and the Watersource variables were added to establish the

distance travelled by households to access water. The watersource type assisted in determining the number of households that consumed water at small areas, that this study does not consider as water. Chapter 4 concludes by presenting the accessibility and lack of accessibility of the three types of services at small geographical (village) level in the CD. The summary of descriptive results are also discussed in Chapter 4.//
Chapter 5 links the results thus obtained with the theoretical models evaluated in Chapter 3. The SAE models on water, sanitation and electricity for lighting were applied at municipality level of the CD. Only Census 2011 data was used for the application of SAE models. The relevant SAS program codes were selected and included in order to illustrate the application of the SAE techniques in service delivery using SAS.

The computations in SAS assisted the study in achieving objectives 5 and 6. This study established that the CD accesses electricity more than water and sanitation; and the Census 2011 data was used to compute undercoverage rates and counts. All model results are presented in Chapter 5. The Sanitation-Policy considerations were drawn from these computed results through the application of SAE models. Chapter 5 concludes by reviewing data limitations in SAS and the estimated household-size by services. Chapter 6 summarises and concludes the study, and also offers recommendations and possible future studies. Finally, References and Appendices are included to follow Chapter 6.

Chapter 2

Design and Data Transformations

2.1 Introduction

Chapter 2 presents data description and sampling design of the GHS series and the Census 2011. Stats SA outlined the variable design in their 2007 *Master Sample Design and Estimation* document. Appendix A3.10 gives provincial codes and Appendices A3.11, A3.12, A3.13, A3.14, A3.15 and A3.16 provide explanation on variable designs: Stratum number, Primary sampling unit number, Dwelling number, Rotation, Unique number and Sample weights, respectively, as described by Stats SA.

2.1.1 Sampling Design for the GHS 2009-2015 Data

This study used the Stats SA's GHS for the period 2009-2015 and Census 2011 data. The GHS is a sampled annual survey that has been conducted by Stats

SA since 2002. The survey was triggered by the need for precise planning and monitoring of the country's progress on its programmes. The GHS information includes a wide range of variables based on multiple living conditions, service delivery and its quality (directly by government and indirectly by agencies), education, health, employment, etc.

It was noted that the GHS sampling and question design has been changing over time. For example, there were 156 questions for the 2002 GHS, 162 questions for the 2003 GHS, 176 in the 2004 GHS, 179 for the 2005 GHS, 169 for the 2006 GHS and 166 for the 2008 GHS. The 2007 GHS included information on HIV/AIDS and mortality. The 2006 and 2007 GHS questions were reviewed to produce the 2008 GHS questionnaire. This study noted from the design and model-dependent based methods outlined in Pfeffermann (2013) that the 2002-2008 GHS data were not comparable; and this may complicate the application of SAE when exploring the change in service delivery for the period 2002-2015.

In Statistics and Econometrics, cross sectional estimates from repeated surveys form a time series which have serial correlations due to survey error process or sample overlaps. The state-space modelling, X-11-ARIMA [1] approach is applied for the analysis of repeated surveys. It allows combining information from distinct sources such as censuses, administrative records and demographic population counts for model fitting (Feder, 2001). This study excluded the the GHS 2002-2008 series which had inconsistent sample errors compared to GHS 2009–2015. The study did not subject different censuses or administrative records to the SAE models applied.

[1] ARIMA stands for AutoRegressive Integrated Moving Average

2.1.2 Sampling Design for Census 2011

Stats SA does not release data at Enumeration Area (EA) as is but creates a Small Area Layer (SAL) of geography to protect respondents' confidentiality. The Stats SA SAL consists of at least 300 people. For this reason, this study considered the household-size as the unit of analysis. In case of less numbers, one or more neighbouring EAs are combined, provided they conform to population thresholds, area size, geographical constraints and land use type. It should be noted that the Primary Sampling Units (PSUs) were the census EAs.

The principle of selecting the Post Enumeration Survey (PES) sample in Census 2011 was determined in such a way that the sampled EA boundaries should be well defined; and should correspond to those of Census 2011 EAs for item by item comparison between the Census and PES records. It was hoped that the stratification and sampling processes applied would cater for the provision of estimates at national, provincial, urban (geography type = urban) and non-urban (geography type = farm and traditional) levels. Instead, estimates were only reliable at national and provincial levels.

The numbers and percentages on households and hostels in the Census 2001 were adjusted according to the PES findings through the application of weights. These weight were calibrated and adjusted for unique households to account for PSUs that were sub-sampled due to growth or those that were segmented (informal PSUs), non-coverage of very small Census EAs that were excluded at the design phase and unit non-response.

Indeed, data relating to other collective living quarters were not weighted. This version is the 10% sample data set which provides raw and weighted data for a small sample of questionnaires, and data was not adjusted for undercount. Similarly 10% Census 2011 undercounts or overcounts were not adjusted.

2.1.3 Application to Census 2011 data

The Census 2011 data was used in SAS, subjected to Small Area Estimation models for the three variables focusing on 302798 households of the Capricorn District only. During data filtering in SAS, a sample of 299891 households remained. The unit of analysis is the household (dependent variable). The household size is also treated as the dependent variable, assuming that it is the number of household members.

The South African Census 2011 was conducted from 9 to 31 October 2011. Each household was visited only once and all usual members of a household were included. To date SA has conducted three Censuses in 1996, 2001, 2011 and the next one is scheduled in 2021. Practically, some people/households are missed or counted more than once during census or an entire EA can be missed. A post enumeration survey (PES) is conducted to assess the degree of undercount or overcount.

A comparison in questionnaire design relating to the three variables was made for Census 2001, Census 2011 and GHS 2011. It was found that questions differ in verbatim, but the context was consistent (see Appendices A3.1 to A3.9). Although the study planned to conduct the analysis on the three variables only, it was found that water (variable) has two questions, each with its categories for the respondents to choose from. The study included the four variables, namely: Water, Water Source, Sanitation and Energy for Lighting in the analysis. This assisted in clarifying the source from which the household accessed the type of water.

2.2 Data Transformation - Coding

Table 2.1 indicates how questions for water, toilet and electricity for lighting were asked by Stats SA. This study shortened the questions in Tables 2.1 and 2.2 for better presentation of items. The respective questions are: *What is the household's main source of drinking water? What is the type of toilet facility used by this household?* and *What is the main source of energy/fuel for this household?*. The Census 2011 questionnaire provides a reference for shortened questions.

2.2.1 How Data was Re-coded

Original questions

Table 2.1: Option categories of variables: Water, Sanitation and Electricity

Water source	**Toilet facility**	**Energy lighting**
01 = Piped (tap) in dwelling	01= Flush toilet - sewerage	01 = Electricity
02 = Piped (tap) in yard	02 = Flush toilet - septic tank	02 = Electricity - generator
03 = Borehole on site	03 = Chemical toilet	03 = Gas
04 = Rainwater on site	04 = Pit toilet air pipe	04 = Paraffin
05 = Neighbour's tap	05 = Pit toilet no air pipe	05 = Wood
06 = Public tap	06 = Bucket toilet	06 = Coal
07 = Water-tanker	07 = None	07 = Candles
08 = Borehole off site	08 = Other (specify)	08 = Animal dung
09 = Flowing water	09 = Unspecified	09 = Solar energy
10 = Stagnant water	-	10 = Other (specify)
11 = Well	-	11 = None
12 = Spring	-	-
13 = Other	-	-
99 = Unspecified	-	-

Source: Statistics South Africa, GHS 2011.

The new binary variables were created according to how the study defines the three variables. A dummy number 1 is assigned to clean treated drinking water, modern flushing toilet and electricity or solar energy for lighting, otherwise a dummy number 0 for the rest of the options. The binary variables were re-coded the same way in SAS for Census 2011 and GHS 2009-2015. Reference on how binary variables, water, sanitation and electricity for lighting were created can be made in Tables 2.3, 2.4 and 2.5 respectively. Table 2.2 indicates the re-coded options of the three variables.

Table 2.2: Re-coding of option categories of Water, Sanitation and Electricity

Water(GHS 2011)	**Toilet(GHS 2011)**	**Electricity (GHS 2011)**
01 = Piped (tap)in dwelling	01= Flush toilet - sewerage	01 = Electricity
02 = Piped (tap)in yard	02 = Flush toilet - septic tank	02 = Electricity - generator
03 = Borehole on site	03 = Chemical toilet	03 = Gas
03 = Rainwater on site	03 = Pit toilet air pipe	04 = Paraffin
04 = Neighbour's tap	03 = Pit toilet no air pipe	04 = Wood
05 = Public tap	03 = Bucket toilet	04 = Coal
06 = Water-tanker	03 = None	04 = Candles
03 = Borehole off site	03 = Other (specify)	04 = Animal dung
03 = Flowing water	09 = Unspecified	05 = Solar energy
03 = Stagnant water	-	04 = Other (specify)
03 = Well	-	04 = None
03 = Spring	-	-
03 = Other	-	-
99 = Unspecified	-	-

Source: Statistics South Africa, GHS 2011.

2.3 Re-coding into SAS

A new data set was created from both the GHS data of 2009-2015 and Census 2011 in SAS, by extracting only the three variables. As we forge to define our focus variables properly considering health reasons, Categories (options of each variable question), which are regarded by this study as irrelevant were further collapsed through re-coding. For example, option 10 (stagnant water) provided for the water question in column 1 of Table 2.1 is not regarded as healthy drinking water in this study – it is therefore re-coded as 03. The same applied to a pit toilet which is regarded as 'not toilet'. The unspecified and undefined options were not re-coded and were left as in the original data set (see Table 2.1).

An example of how options of questions were re-coded is presented in Table 2.2.

The coding in SAS was used to derive binary variables from the categorical variables, namely: Water, Water Source, Sanitation, Energy for Lighting. The binary variables were created using the number corresponding to the variable question number in the questionnaire. For example, sanitation variable is question 10 in Census 2011 questionnaire, i.e., (H10-Toilet) variable.

2.3.1 Source of Water

The new binary variable water, derived from the water variable, is defined as follows: Water = 1 if H08-watersource =(1, 8), i.e. access to Tap drinking water. Water = 0 if H08-watersource = (2, 3, 4, 5, 6, 7, 9), i.e., no access to Tap drinking water.

For this variable, the results are presented for both coded and un-coded data (see Table 2.3).

Table 2.3: What is the household's main source of drinking water?

Water source (H08-watersource)	**Water source (H08-watersource)**
1.Regional/local water scheme services	6. River/stream
2. Borehole	7. Water vendor
3. Spring	8. Water tanker
4. Rain water tank.	9. Other
5. Dam/pool/stagnant water	-

Source: Stats SA, Census 2011 questionnaire extract.

Table 2.4: What type of toilet is used by this household?

Sanitation (H10-TOILET)	Sanitation (H10-TOILET)
0. None	4. Pit latrine with ventilation (VIP)
1. Flush toilet (connected to sewerage system)	5. Pit latrine without ventilation
2. Flush toilet (with septic tank)	6. Bucket latrine
3. Chemical toilet	7. Other

Source: Stats SA, Census 2011 questionnaire.

2.3.2 Sanitation

The new binary variable *Toilet* was derived from the Sanitation variable for households with access to flush toilet and those with toilets that are not flushing (Table 2.4).

Toilet = 1 (Toilet) if H10-Toilet = (1, 2), i.e., flush toilet with sewage system/septic tank.

Toilet = 0 (No Toilet) if H10-Toilet = (0, 3, 4, 5, 6, 7).

2.3.3 Electricity for Lighting

Table 2.5: What type of energy/fuel does this household mainly use for lighting

Electricity for Lighting (H11-ENERGY-LIGHTING)
1. Electricity
2. Gas
3. Paraffin
4. Wood (not for lighting by Stats SA)
5. Coal (not for lighting by Stats SA)
6. Candles
7. Animal dung (not for lighting by Stats SA)
8. Solar
9. Other
10. None

Source: Stats SA, Census 2011 questionnaire.

The new binary variable *Lighting* was created, derived from Energy/Fuel (Table 2.5):

Lighting = 1 (Access to electricity) if H11-ENERGY-LIGHTING = (1, 8).

Lighting = 0 (No access to electricity) if H11-ENERGY-LIGHTING = (2, 3, 6, 9, 10).

The *Lighting* variable has seven actual categories declared by Stats SA, assuming that households cannot use wood, coal and animal dung for lighting. Hence, categories 4, 5 and 7 were excluded for re-coding and are not available in the Stats SA file. Here, the assumption is that the household with access to electricity can use it for cooking and heating, but the poor households will reserve it for lighting only.

Chapter 3

Literature Review

3.1 Application and Assumptions of SAE

This chapter presents the evaluation of small area estimation models, coupled with the literature review of scholars who applied these models. As indicated in Chapter 1, the phrase *small area* refers to a small geographical portion of a large area such as a county or small domain. For example, a district municipality in the whole of SA, a subgroup of the whole population, a Grade 12 class in a community having many secondary schools, etc.

Rao (2003) outlined several examples in the United States where SAE methods were applied in health care planning at the state and individual levels, in local agricultural decision making, in Bureau of Census for the PCI for small areas, and also in poverty counts.

Guided mainly by the computations in Rao (2003), Rahman (2008), Mukhopadhyay and McDowell (2011), Rao and Molina (2015) and Guadarrama et al.

(2016), this study applied SAE techniques to estimate the accessibility of water, sanitation and electricity for lighting at small areas (municipality level) in the Limpopo Province of South Africa.

3.2 Direct Estimation

Direct estimation requires that all small areas be sampled to produce direct estimators. It relies only on the sample obtained from the survey. Different direct estimators can be produced if the survey samples cover each small area with sufficient data. Since direct estimation produces estimates directly from the locally collected data, Census 2011 data was used assuming that the 10% sample is large enough to provide reliable direct estimates.

The Horvitz-Thompson (H-T) estimator is an unbiased estimator for the population total under unequal probability sampling. It is a method for estimating the total and mean of a super population in a stratified sample (Horvitz and Thompson, 1952). Inverse probability weighting is applied to account for different proportions of observations within strata in a target population. The H-T estimator is applied in survey analysis and can be used to account for missing data. This study used the H-T estimator to produce direct estimates and account for missing data for water, sanitation and electricity for lighting in the CD.

In H-T estimation, when all areas are sampled, the direct π estimator gives the area mean value as:

$$\hat{\bar{Y}}_{i,DIRECT} = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}, \tag{3.1}$$

where y_{ij} represents the value of the study variable in area i and unit j. The weights w_{ij} are used as the inverse of the probability of an individual to be included in the sample. All areas are sampled independently and *with* replacement. The probability of selecting an individual j in area i is given by

$$p_{ij} = \frac{1}{N_i}, \tag{3.2}$$

where N_i is the number of individuals in area i, and $\bar{Y}$ in Equation (3.1) denotes the area level mean of the target variable and n_i is the sample size.

If the sample size in region i is n_i, the probability of selecting an individual at least once is:

$$1 - \left(1 - \frac{1}{N_i}\right)^{n_i}, \tag{3.3}$$

which denotes the inclusion probability. Using weights in Equation (3.3), we obtain:

$$w_{ij}^{-1} = w_i^{-1} = 1 - \left(1 - \frac{1}{N_i}\right)^{n_i}. \tag{3.4}$$

There is a need to obtain the design variance of the direct estimator, which can be estimated and used to assess the uncertainty of the estimates. This is also useful in providing the estimated confidence intervals such as results obtained in the last two columns of Table 5.14. Since this study uses simple random sampling *with* replacement, the design variance of the direct estimator is given by:

$$V[\hat{\bar{Y}}_{i,DIRECT}] = 1 - \left(1 - \frac{1}{N_i}\right)\frac{S_i^2}{n_i}, \tag{3.5}$$

where S_i^2 is the variance of the sample obtained from area i. The variance can be estimated (see Tables 5.2 and 5.3) by substituting the generic sample S_i^2 by the variance of the observed data $\hat{S}_i^2$ to obtain:

$$\hat{V}[\hat{\bar{Y}}_{i,DIRECT}] = 1 - \left(1 - \frac{1}{N_i}\right)\frac{\hat{S}_i^2}{n_i}. \tag{3.6}$$

This relates to objective number 5 in Section 1.7.2 (see Table 5.4 for the results). In using direct estimation technique, Lehtonen and Pahkinen (2004) found that the direct estimator is less efficient, i.e., it produces high variance when sampling *with* replacement than when sampling *without* replacement.

3.2.1 The Generalised Regression Estimator

Sarndal et al. (1992) outlined how the Generalised Regression Estimator (GREG) can be used to link direct information obtained from the sample with aggregated data. The procedure is necessary to improve the quality of the direct estimates. The GREG estimator produced on a linear predictor is

$$\hat{\hat{Y}}_{i,GREG} = \hat{\beta}\bar{X}_i^T + \sum_{j=1}^{n_i} w_{ij}\left(y_{ij} - \hat{\beta}\frac{x_{ij}^T}{N_i}\right) = \hat{\beta}\bar{X}_i^T + \hat{\hat{Y}}_{i,DIRECT} - \sum_{j=1}^{n_i} w_{ij}\hat{\beta}\frac{x_{ij}^T}{N_i}, \quad (3.7)$$

which is calculated through weighted regression of the sample data, the weighted correction terms derived on the sampled units and the difference between the observed and predicted values of the individual. It is noted that in Equation (3.7)

$$\hat{\beta}\bar{X}_i^T, i = 1, ..., d, \quad (3.8)$$

estimates the vector of values; $\bar{X}$ represents the vector with area mean values of the covariates and

$$\sum_{j=1}^{n_i} w_{ij}\left(y_{ij} - \hat{\beta}\frac{x_{ij}^T}{N_i}\right), i = 1, ..., d \quad (3.9)$$

estimates the values. This study expects the Average Empirical Mean Square Error (AEMSE) of the GREG estimate to be lower than that of the direct estimate. AEMSE is used to assess the quality of the estimates. The lower the value of AEMSE, the better the estimates fit the real values.

3.3 Indirect Estimation

Indirect estimation or model-based estimation is approached in two ways: Statistical and Demographic approaches. This study presents the statistical approach which uses explicit and implicit models. In indirect estimation, the model is chosen and its parameters are estimated using the data from the survey. In the indirect estimation, the auxiliary information or covariates are required to produce unbiased estimates.

3.4 Implicit Models

Implicit models provide techniques to relate small areas through additional (sometimes called supplementary) data from census or administrative records. This study has considered two implicit models: Synthetic estimation and Composite estimation.

3.4.1 Synthetic Estimation

Synthetic estimation can be defined as the application of model-based techniques to combine data obtained from the national survey with a set of associated covariates or predictor variable available for all small areas. For example, the proportion of residents who were living as a couple, claiming income support, had limiting long standing illnesses, etc.

Gonzalez et al. (1996) define an estimator to be synthetic if the reliable direct estimator for a large area is successfully used to produce an indirect estimator for a sub-area included in a large area, only if all sub-areas have similar characteristics to those of the large area. The synthetic estimator is generally derived by fitting a regression model to the available data to obtain the predicted values (Saei and Chambers, 2003). The synthetic estimator is justified on assuming the linear model for the data in order to estimate from the model, the values of the areas which were not sampled. In this case, information for available covariates is used. The mean for the synthetic estimator can be written as:

$$\bar{Y}_i = \beta \bar{X}_i + \mu_i, \tag{3.10}$$

where μ_i represents the area-based random effects, which is normally distributed with mean zero and variance σ^2_μ. Here the synthetic estimator is obtained by using the estimate of β from the linear regression of the individual level sample data:

$$\hat{\bar{Y}}_{i,SYNTHETIC} = \hat{\beta}\bar{X}_i, \tag{3.11}$$

which is the synthetic estimate in area i. The challenge here is that the estimator in Equation (3.11) does not incorporate the random effects μ_i, which then causes the area mean estimates to be biased.

3.4.2 Composite Estimator

The weighted sum of the direct and synthetic estimators is combined to produce the composite estimator, defined as follows:

$$\hat{\bar{Y}}_{i,COMPOSITE} = \hat{\gamma}_i \hat{\bar{Y}}_{i,DIRECT} + (1 - \hat{\gamma}_i)\hat{\bar{Y}}_{i,SYNTHETIC}, \tag{3.12}$$

where

$$\hat{\gamma}_i = \frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\mu + \frac{\hat{\sigma}^2_\mu}{n_i}}. \tag{3.13}$$

σ^2_μ is the estimate variance for the random area effects and $\hat{\gamma}_i$ represents the value ranging from 0 to 1, which is used to manage the reduction of the direct estimate and the synthetic estimates, depending on the size of the sample in the small area. If the sample is large, more weight is assigned to the direct estimator than to the synthetic estimator. In the case when the sample is large, but not reliable – extra information is required from other areas, and more weight should be given to the synthetic estimator (Saei and Chambers, 2003). Ghosh and Rao (1994) chose $\hat{\gamma}_i$ in such a way that it minimises the Mean Square Error (MSE) in Equation (3.12) or the average MSE of all synthetic estimators. In principle, the composite estimator should fit better than the synthetic estimator.

3.5 Explicit Models

Explicit models give justification for the small area variations through supplementary data (Rahman, 2008). This study uses multilevel modelling to present the application of unit (individual) and area levels. We are guided by Goldstein (2003) in applying multilevel modelling in SAE to compute models which have different layers that give different effects. The application of explicit models in SAE: Unit level, Area level and General Linear Mixed (GLM) models, are presented.

3.5.1 Unit (Individual) Level

Unit level model is based on unit level auxiliary variables. The linkage exists through unit level response values by nested error linear regression model:

$$y_{ij} = x'_{ij}\beta + \varepsilon_i + e_{ij}, \tag{3.14}$$

where

y_{ij} represents the unit responses assumed to be related to the auxiliary values x_{ij} through the target population estimate at small area i, which is the nested error regression equation (Rahman, 2008).

x_{ij} justifies a unit specific auxiliary information available for areas $i = 1, 2, ..., n$ and individuals $j = 1, 2, ..., N_i$:

N_i is the number of population units in the i^{th} area

β is a vector of regression parameters

ε_i are normal, independent and identically distributed $(i.i.d.)$ with mean 0 and variance σ_ε^2

e_{ij} are normal, independent of ε_i and identically distributed with mean zero and variance σ_ε^2.

The initial layer at which reliable estimates can be obtained from a sample and the additional auxiliary covariates, is at the unit level, which is expressed as follows:

$$y_{ij}|\beta, \mu, \sigma_e^2 \sim N(\beta x_{ij} + z_i\mu_i, \sigma_e^2), \tag{3.15}$$

where

z_i represents the structure of μ_i and it should be equated to 1. Equation (3.15) can be used to model different types of area effects μ_i. It can be used to model the area level effects as follows:

$$\mu_i|\sigma_u^2 \sim N(0, \sigma_u^2). \tag{3.16}$$

Considering different structures for the area random effects by modelling different spatial and temporal effects, the model is written in a matrix form as follows:

$$y|\beta, U, \sigma_e^2 \sim N(\beta x + ZU, \sigma_e^2 I_N), \tag{3.17}$$

$$U|\sigma_\mu^2 \sim N(0, \sigma_\mu^2 I_m),$$

where

$N = \sum_{i=1}^{m} N_i$ and N_i is the number of individuals in region i, m is the total number of regions, and Z serves to model the structure of μ. Equation (3.17) can be properly specified to model more than two levels. For the area effects, the model is written as:

$$y|\beta, \sigma_e^2, \sigma_\mu^2 \sim N(\beta x, \sigma_e^2 I_N + \sigma_\mu^2 Z I_m Z^T). \tag{3.18}$$

3.5.2 Area Level

The model in Equation (3.18) can be used to produce area estimates only if the aggregated data is available and Equation (3.18) can be specified as follows:

$$\bar{Y}_i|\beta, \sigma_e^2, \sigma_\mu^2 \sim N(\beta \bar{X}_i, \frac{\sigma_e^2}{n_i} + \sigma_\mu^2). \tag{3.19}$$

The model in Equation (3.18) is difficult to estimate because there is one observation per area and we will need to estimate two variances: i.e., the random effects σ_μ^2, and of the error term σ_e^2.

3.5.3 General Linear Mixed Model (Unit and Area)

We observe from Equation (3.14), $y_{ij} = x'_{ij}\beta + \varepsilon_i + e_{ij}$, that most of the SAE models are based on the General Linear Mixed (GLM) model defined as follows:

$$y = X\beta + Z\varepsilon + e, \tag{3.20}$$

where

y represents a vector of responses

X denotes a known covariates matrix

β is the regression coefficient vector (called fixed effects)

Z is the known structure matrix of area random effects
ε denotes the random effects vector due to small area
e is the random errors vector associated with sampling error (i.e. the variation of individual or unit level, assuming that $\varepsilon \sim N(0, \sigma_{\varepsilon}^2\Theta)$) while $e \sim N(0, \sigma_{\varepsilon}^2\Omega)$. ($\Theta$ and Ω) are positive definite matrices, ε and e elements are not correlated.

The small area models considered above are regarded as the special cases of linear mixed models involving fixed and random effects. The means or totals of small areas may be represented as linear combination of fixed and random effects. Pratesi and Salvati (2008) illustrates how to compute the Best Linear Unbiased Prediction (BLUP) estimator of such parameters.

3.6 Bayesian Estimates

3.6.1 Empirical BLUP Estimators

It is noted that the BLUP estimator which could be produced in GLM model may not depend on normality. BLUP minimises the model MSE in the group of linear models of unbiased estimators of the quality of interest. The BLUP estimators depend on variances and covariance of random effects which are estimated through fitting constants or moments. If normality is assumed, Maximum Likelihood (ML) or Residual Maximum Likelihood (RML) techniques are used to estimate variances and covariance components. The estimated components are used in the BLUP estimator to obtain Empirical Best Linear Unbiased Prediction (EBLUP) estimator. For this two-stage EBLUP estimator, the variability of the estimated variance and covariance components are taken into consideration.

If δ is known, the BLUP estimator may be defined as:

$$\tilde{\mu}^H = t(\delta, y) = I^T\tilde{\beta} + m^T\tilde{v} = I^T\tilde{\beta} + m^T GZ^T V^{-1}(y - X\tilde{\beta}), \tag{3.21}$$

where

$\tilde{v} = \tilde{v}(\delta) = GZ^T V^{-1}(y - X\tilde{\beta})$,

H represents Henderson (Rao, 2003) and

$\tilde{\beta} = \tilde{\beta}(\delta) = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ is the best linear unbiased estimator of β.

Let us assume that the parameter of interest is:

$$\theta = (\theta_1, ..., \theta_m), \tag{3.22}$$

which can be the vector of totals or area means, m denotes the number of regions and θ may be written as follows:

$$\theta = X\beta + Z\mu, \tag{3.23}$$

where

X denotes the set of covariates

β is the associate coefficients

μ is the area random effects, and

Z is a matrix that models the structure of μ.

The random effects μ are normally distributed with mean 0 and variance δ_μ^2. The computation of the direct estimator $\hat{\theta}$ was derived by Petrucci et al. (2005), applying the Fay-Herriot model to combine Equations (3.22) and (3.23), to obtain:

$$\hat{\theta} = \theta + e, \tag{3.24}$$

where

e represents the sampling error (i.e. variance) and also a diagonal matrix of known constants obtained from the survey design.

μ and e are assumed to be independent

β estimates are obtained through the standard Generalised Least Squares methods

μ estimates are computed using their EBLUP as follows:

$$\hat{\mu} = E[\mu|y], \tag{3.25}$$

where y denotes the sampled data.

ML or REML can be used to compute $\hat{\mu}$ (McCulloch and Neuhaus, 2001; Rao, 2003). Therefore, the EBLUP estimator is defined as:

$$\tilde{\theta} = X\hat{\beta} + Z\hat{\mu}. \tag{3.26}$$

3.6.2 Spatial EBLUP Estimator

The idea of *borrowing strength* by using the information from neighbouring areas when estimating spatially correlated random effects in order to improve estimation in non-sampled areas, was presented by Jiang and Lahiri (2006). The spatial EBLUB estimator is an estimator which considers a simultaneously autoregressive (SAR) specification for the area random effects. If the parameter of interest is θ, then:

$$\theta = X\beta + Zv, \tag{3.27}$$

where $v = \rho W\mu + \mu$, ρ is a special autoregressive coefficient, W is the adjacent matrix and μ denotes random area effects. In fact $v = \mu(1 - \rho W)^{-1}$.

The relationship between the direct estimator and the true value can be written as follows:

$$\hat{\theta} = \theta + e = X\beta + Z(1 - \rho W)^{-1}\mu + e. \tag{3.28}$$

The estimation of β is computed using the generalised least squares methods and the estimates of v are calculated using the EBLUP estimator $\hat{v} - E[v|y]$ (Petrucci et al., 2005; Petrucci and Salvati, 2004; Jiang and Lahiri, 2006).

Then the spatial EBLUP of θ is:

$$\tilde{\theta} = X\tilde{\beta} + Z\hat{v}. \tag{3.29}$$

3.6.3 Empirical Bayes Estimator

Another approach used to estimate the MSE of EBLUP is through Empirical Bayes (EB) estimation. In some particular applications, EB and EBLUP estimators are identical under normality assumptions.

Let the optimal estimator of the computed values of θ_i be given as $f(\theta_i|\hat{\theta}, \beta, \hat{\sigma}_v^2)$. Then the expectation of θ_i is (Rao, 2003):

$$E(\theta_i|\hat{\theta}, \beta, \hat{\sigma}_v^2) = \hat{\theta}_i^B = \gamma_i\hat{\theta}_i + (1 - \gamma_i)z_i^T\beta \tag{3.30}$$

$$E(\theta_i|\hat{\theta}, \beta, \hat{\sigma}_v^2) = \hat{\theta}_i^B = \frac{b_1^2\hat{\sigma}_v^2}{b_1^2\hat{\sigma}_v^2 + \psi_i}\hat{\theta}_i + (1 - \gamma_i)z_i^T\beta,$$

with the assumption that the conditional distribution of θ_i, i.e., $(\theta_i|\hat{\theta}, \beta, \hat{\sigma}_v^2)$ is independent and distributed $N(\hat{\theta}_i^B, g_{1i}\sigma_v^2 = \gamma_i\psi_i)$. The estimator $\hat{\theta}_i^B = \hat{\theta}_i^B(\beta, \sigma_v^2)$ is called the Bayes estimator under squared-error loss, and is regarded optimal because of its MSE. The MSE $\theta_i^B = E(\hat{\theta}_i^B - \theta_i)^2$ is the smallest of all the estimators of θ_i (Rao, 2003). The Bayes estimator $\hat{\theta}_i^B$ depends on the model parameters β and σ_v^2 which are estimated from the marginal distribution: $\hat{\theta}_i \sim i.i.d., N(z_i^T\beta, b_i^2\sigma_v^2 + \psi_i)$, through ML and RELM. It is noted from Jiang and

Lahiri (2006) that $\hat{\theta}_i^B$ was computed from the conditional distribution in Equation (3.30) without assuming a prior distribution on the model parameters, and this makes θ_i^B to be the best predictor of θ_i.

3.6.4 Hierarchical Bayes Estimator

Hierarchical Bayes estimation requires further model specification or assumptions on the model parameters. For example, EBLUP application is restricted to linear mixed models whereas Hierarchical Bayes (HB) and Empirical Bayes (EB) models are valid to general application. EB and HB are applicable to models for binary and count data, including normal linear mixed models. In Bayesian inference, the parameter is considered a random variable, and the Bayesian approaches rely in any prior knowledge of the experiment/project under consideration.

The derivation of the Prior and Posterior distribution is outlined in Hogg et al. (2005) as follows:
Let X be a random variable (r.v.) with the distribution of the probability that depends on θ, where θ represents an element of a defined Ω set. Suppose that θ is the mean of a normal distribution, then Ω may be regarded as a line. In addition, let Θ be a random variable with the distribution of probability over the set Ω. This will imply that x is a possible value of a r.v. X and θ a possible value of a r.v. Θ. The pdf of Θ is denoted by $h(\theta)$, where $h(\theta) = 0$ if $\theta \notin \Omega$. $h(\theta)$ is called the prior pdf of Θ. Then the conditional pdf of X is denoted by $f(x|\theta)$. This model can be written as: $X|\theta \sim f(x|\theta)$, i.e., $\Theta \sim h(\theta)$.
Let $X_1, X_2, ..., X_n$ be a random sample from the conditional distribution X given $\Theta = \theta$ with pdf $h(\theta)$, i.e., we let the vector $X' = X_1, X_2, ..., X_n$ and $\mathbf{x}'$. Then the

joint conditional pdf of $\Theta = \theta$ is

$$L(x|\theta) = f(x_1|\theta)f(x_2|\theta), ..., f(x_n|\theta).$$

The joint pdf of X and Θ is $g(x,\theta) = L(x|\theta)h(\theta)$. If Θ is continuous, then the joint marginal pdf of X is written as[1]:

$$g_1(x) = \int_{-\infty}^{\infty} g(x,\theta)d\theta. \tag{3.31}$$

This is called the predictive distribution of X given the likelihood and the prior $h(\theta)$. The conditional pdf of Θ given the sample X is:

$$k(\theta|x) = \frac{g(x,\theta)}{g_1(x)} = \frac{L(x|\theta)h(\theta)}{g(x)}. \tag{3.32}$$

$k(\theta|x)$ is called the posterior pdf and any distribution defined from it is called the posterior distribution. It is noted here that the prior distribution is based on the subjective influence of Θ, while the posterior is the conditional distribution of Θ derived from the data. If additional data is collected beyond $x_1, x_2, ..., x_n$, then the posterior distribution computed from $x_1, x_2, ..., x_n$ will be the new prior distribution, and additional information assists in producing a new posterior distribution from which inferences are generated. The repetition of this process is called Bayesian sequential procedure. The *non-informative prior* is the prior that treats all values of θ the same, as if all values are uniform.

[1]The integral sign is replaced by Σ if Θ is discrete

3.7 Markov Chain Monte Carlo Methods

Solving the posterior distribution analytically is often difficult because it requires multidimensional integration methods to determine the integration constant. The practical alternative way is to compute the integral using the numerical integration methods only if few parameters are involved. This problem is solved by sampling with Markov Chain Monte Carlo Methods (MCMC) methods called Gibbs Sampler and Metropolis algorithm. The the Gibbs Sampler was introduced by Geman and Geman (1983) to estimate image processing parameters of *Gibbs distribution*. Gelfand and Smith (1990) introduced Gibbs Sampling in order to tackle complex estimation problems through Bayesian methods.

The MCMC method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. In principle, the MCMC procedure is designed specifically for this purpose (Chen, 2009).

Suppose the random sample is drawn from the $N(\theta, \sigma^2)$, where θ is known. Let $Y = \bar{X}$ be a sufficient statistics. The Bayes model:

$$Y|\theta \sim N(\theta, \frac{\sigma^2}{n}),$$

$$\Theta \sim h(\theta) \propto \exp \frac{(-\frac{\theta - a}{b})}{(1 + \exp(-[\frac{\theta - a}{b}])^2)}, -\infty < \theta < \infty, a > 0, b > 0, \tag{3.33}$$

is the prior logistic distribution, where a and b are known.
The inverse of the logistic cdf can be written as:

$$a + b \log\left(\frac{\mu}{1-\mu}\right), 0 < \mu < 1 \tag{3.34}$$

and the posterior pdf can be written as:

$$k(\theta|y) = \frac{\frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} \exp\left(1 - \frac{1}{2}\frac{(y-\theta)^2}{\frac{\sigma^2}{n}}\right) \frac{b^{-1}e^{\frac{-(\theta-a)}{b}}}{(1+e^{[\frac{(\theta-a)}{b}]^2})}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} \exp\left(1 - \frac{1}{2}\frac{(y-\theta)^2}{\frac{\sigma^2}{n}}\right) \frac{b^{-1}e^{\frac{-(\theta-a)}{b}}}{(1+e^{[\frac{(\theta-a)}{b}]^2})} d\theta}. \tag{3.35}$$

If the error loss is assumed, the Bayes estimate will be the mean of the posterior in the distribution given by Equation (3.35). Its computations require double integral which cannot yield any results in a closed form. Alternatively, let us consider the likelihood $f(y|\theta)$ to be the function of θ:

$$f(y|\theta) = w(\theta) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} \exp\left(1 - \frac{1}{2}\frac{(y-\theta)^2}{\frac{\sigma^2}{n}}\right).$$

The Bayes estimate could be written as:

$$\delta(y) = \frac{\int_{-\infty}^{\infty} \theta w(\theta) \frac{b^{-1}e^{\frac{-(\theta-a)}{b}}}{(1+e^{[\frac{(\theta-a)}{b}]^2})} d\theta}{\int_{-\infty}^{\infty} w(\theta) \frac{b^{-1}e^{\frac{-(\theta-a)}{b}}}{(1+e^{[\frac{(\theta-a)}{b}]^2})} d\theta} = \frac{E[\Theta w(\Theta)]}{E[w(\Theta)]}. \tag{3.36}$$

Here the expectation is taken with Θ in the prior logistic of Equation (3.33). The estimation is computed using Monte Carlo techniques as follows: The prior logistic equation $\Theta \sim h(\theta)$ is used to generate $\Theta_1, \Theta_2, ..., \Theta_m$ for a random variable:

$$T_m = \frac{m^{-1}\sum_{i=1}^{m} \Theta_i w(\Theta_i)}{m^{-1}\sum_{i=1}^{m} w(\Theta_i)}. \tag{3.37}$$

The value of m can be too large. Applying the weak law of large numbers, $T_m \to \delta(y)$.

Using the Monte Carlo method, the sample can be bootstrap to compute the confidence interval for the estimate:

$$\frac{E[\Theta w(\Theta)]}{E[w(\Theta)]}.$$

Suppose that we need to generate two streams of $i.i.d.$ variables from X and the other from Y, i.e., (X, Y) has pdf $f(x, y)$. We generate $Y_i|X_{i-1} \sim f(y|x)$ and $X_i|Y_i \sim f(x|y)$.

Let x_{i-1} represent the observed value of X_{i-1} and y_i the observed value of Y_i. We use the value x_{i-1} to generate sequentially the new Y_i from the pdf $f(y|x_{i-1})$, and then draw a new X_i from the pdf $f(x|y_i)$. It should be noted that X_{i-1} would have been generated prior to the ith step of the algorithm. This is known as the Gibbs sampler. Consider the sequence of generated pairs as follows:

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), ..., (X_k, Y_k), (X_{k+1}, Y_{k+1}).$$

All previous pairs included are not necessary, except for (X_k, Y_k). This means that if the present state of the sequence is given, the future of the sequence is independent of the past. Stochastically, such sequences are called Markov chains. In general, these chains reach stability as the chain increases. One way of controlling the Prior is to model it in terms of another random variable through Hierarchical Bayes model – known as MCMC methods.

Chapter 4

Results

4.1 Introduction

This chapter presents the descriptive analysis for the results of the re-coded GHS series data from 2009 to 2015. The proportions of households accessing /not accessing water, sanitation and electricity for lighting are presented for the nine provinces of South Africa.

4.2 Access to services: SA, GHS

This section analyses the re-coded GHS data at the national level (SA).

4.2.1 Access to Drinking Water: SA

Throughout the entire study period, access to drinking water is lowest in the Eastern Cape and Limpopo provinces (see Figure 4.1). These are the two provinces that have been considered the poorest in the country across all major economic and social categories such as employment, education, poverty, infrastructure and access to basic services, in general. The third lowest province with regard to access to tap drinking water is KwaZulu-Natal. Relatively few households in the Western Cape and Gauteng provinces (less than 1.0% and 2.5% respectively) were disadvantaged on access to tap drinking water in 2015 (see Figure 4.1).

Estimates from different GHS.

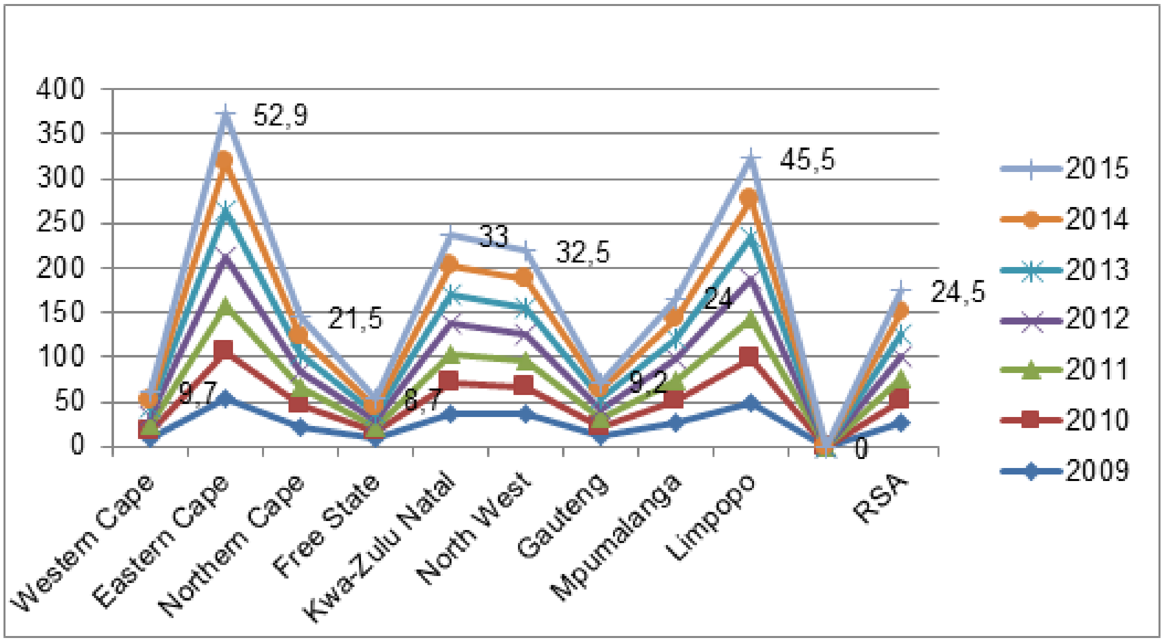

Figure 4.1: Proportion of HHs with No Tap drinking water by province, GHS

For more information on the accessibility of clean drinking water by province, the reader is referred to Figures 6.1 and 6.2 in Appendices A4.1. and A4.2, respectively.

4.2.2 Access to Sanitation: SA

Figure 4.2 presents the proportions of households with no access to flushing toilets, by province, applying the GHS series. Although Limpopo has the least percentage of households accessing flush toilets, the percentage of households with flushing toilets increased steadily from 2009 to 2015. In Limpopo, the change in accessing flush toilet was not significant between 2009 and 2011 [2009 (19.6%), 2010 (19.2%) and 2011 (19.4%)]. Figure 4.2 shows that at least 40.0% of households in the Eastern Cape had flushing toilets between 2010 and 2015. Figure 4.2 also reveals that households in KwaZulu-Natal and North West have similar access to flushing toilet throughout the entire period from 2009 to 2015.

Limpopo is the most deprived province in SA concerning access to proper sanitation facilities. The proportion of HHs without flush toilets in the LP slightly decreased from 80.4% in 2009 to 73.0% in 2015 (see Figure 4.2). The best sanitation system in the country is enjoyed by the residents of the Western Cape province, where only 3.6%, 3.9% and 5.3% of HHs did not have access to flush toilets in 2010, 2014 and 2015, respectively.

Reference to the accessibility of flushing toilet by province also appears in Figure 6.3, Appendix A4.2.

4.2.3 Access to Electricity for Lighting: SA

South Africa generally provides its people with reasonable electricity services across all its nine provinces (Table 4.1). By 2015, the Eastern Cape (with the least services in this regard), had nearly 87% of its HHs accessing electricity for lighting, an improvement from 69.3% in 2009. It is in the provision of electricity for lighting services where the people of Limpopo, not only ranked

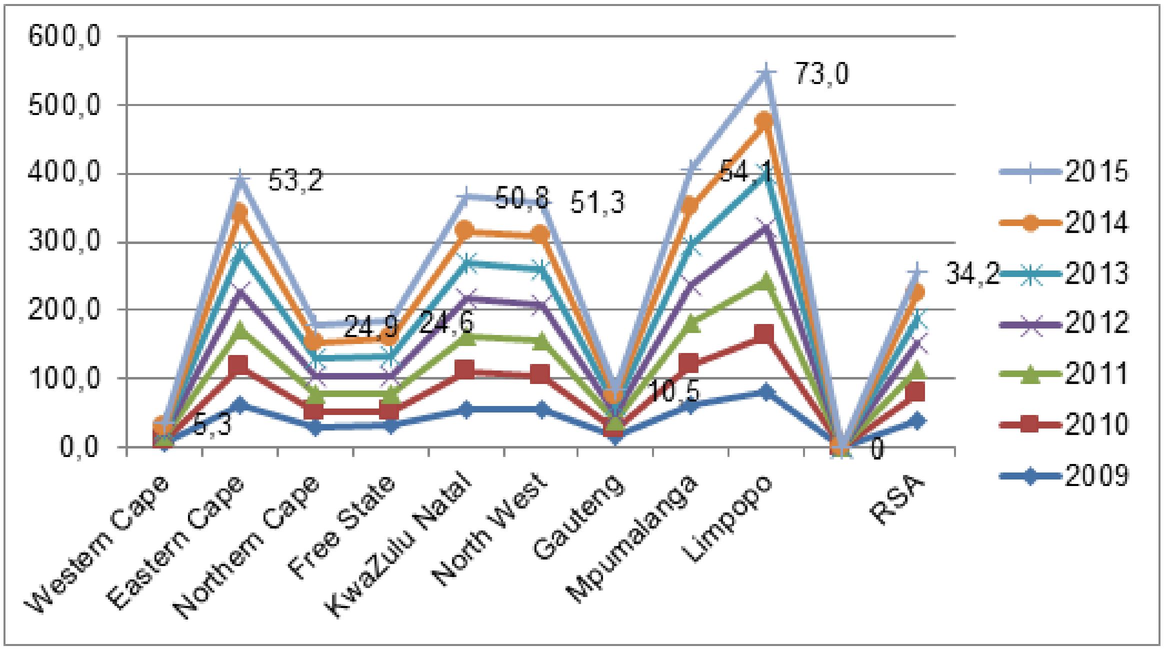

Figure 4.2: Proportion of HHs with No Flush toilet by province, GHS

Table 4.1: Proportion of HHs with access to electricity for lighting by province, GHS

Province	2009	2010	2011	2012	2013	2014	2015
Western Cape	91.43	90.63	96.00	96.22	96.80	97.60	97.91
Eastern Cape	69.33	72.87	76.15	82.58	83.45	86.86	86.88
Northern Cape	89.42	88.29	90.94	93.13	91.29	91.40	93.58
Free State	91.14	93.17	95.09	93.16	94.17	94.93	93.01
KwaZulu Natal	76.17	76.96	78.64	82.40	84.89	87.34	88.01
North West	82.78	83.19	85.62	86.84	89.73	89.08	91.11
Gauteng	88.31	87.06	81.99	86.77	87.77	90.79	90.92
Mpumalanga	85.89	87.18	89.11	90.19	91.65	91.95	90.84
Limpopo	84.49	87.52	91.13	91.12	92.44	93.31	94.52
RSA	83.61	84.27	84.97	87.77	89.08	90.96	91.32

Source: CD Spatial Development Framework, Statistics South Africa, 2011.

above the national average of 91.3%, but also came second to the Western Cape on the proportion of HHs accessing the service. HHs in the Western Cape and the Free State have been enjoying the provision of electricity for lighting since 2009. Only HHs in the Eastern Cape and KwaZulu-Natal remained below 90%

by 2015. The provinces that more than 1% increase in accessing electricity for lighting from 2014 to 2015 are Northern Cape (2.18%), North West (2.03%) and Limpopo (1.21%).

Additional information on the accessibility of electricity for lighting by province can be obtained from Figure 6.4 in Appendix A4.2.

4.3 Geotype Data: LP, GHS

In this section, the re-coded GHS data is analysed at the Limpopo provincial level (i.e., LP).

4.3.1 Access to Tap Drinking Water: LP

This section presents the proportions of households which do not have access to water, sanitation and electricity for lighting services in the LP by Geotype. Trends are compared for change during the years 2009, 2014 and 2015. The data were also re-coded for consistency of definitions of water, sanitation and electricity indicated in this study.

Table 4.2: Proportions of HHs with No Tap drinking water in LP, GHS 2009 and 2014

Area Type	**2009**	**2014**
1. Urban Formal	7	4
2. Urban Informal	28	0
3. Tribal Areas	32	54
4. Rural Formal	33	42
5. Total	100	100

Source: Stats SA, Census 2011 questionnaire.
Note: Farm and Traditional were used in 2015 only.
Note: Rural Formal was used in 2009 and 2014.

The definition of geotype was changed in 2015, with categories Farm and Traditional introduced in place of Rural and Tribal areas, respectively. Again in 2014 and 2015, the Urban Informal category did not exist. The only consistent geotype category across all the years is Urban Formal. Small proportions of HHs in Limpopo (not more than 10%), had access to clean water, i.e. Tap drinking water (Table 4.2). Thus, an overwhelming majority ($\geq$90%) of HHs in the LP reside in rural formal, farm and traditional areas, or in urban informal settlements without access to Tap drinking water. This is consistent with the findings in Section 4.2.1, where Limpopo was found to be the most unfavourable province in the country with regards to Tap drinking water over the entire study period.

4.3.2 Access to Flush Toilet: LP

Table 4.3: Proportions of HHs with No Flush Toilet in LP, GHS 2009 and 2014

Area Type	**2009**	**2014**
1. Urban Formal	8	3
2. Urban Informal	23	8
3. Tribal Areas	45	51
4. Rural Formal	24	38
5. Total	100	100

Source: Stats SA, Census 2011 questionnaire.
Note: Farm and Traditional were used in 2015 only.
Note: Rural Formal was used in 2009 and 2014.

About 8% and 3% of HHs in Limpopo did not have flush toilets in 2009 and 2014, respectively (see Table 4.3). Thus in 2009 and 2014, most HHs without flush toilets in the LP reside in Tribal Areas and Rural Formal. Over the six-year period under study, no improvements have been realised on proper sanitation in the province. The government / municipalities need to put an extra effort to improve this situation.

4.3.3 Access to Electricity for Lighting: LP

Table 4.4: Proportions of HHs with No access to electricity for lighting, GHS 2009 and 2014

Area Type	2009	2014
1. Urban Formal	11	5
2. Urban Informal	44	59
3. Tribal Areas	12	8
4. Rural Formal	33	28
5. Total	100	100

Source: Stats SA, Census 2011 questionnaire.
Note: Farm and Traditional were used in 2015 only.
Note: Rural Formal was used in 2009 and 2014.

Table 4.4 indicates that only 11% of HHs without access to electricity for lighting in Limpopo lived in the Urban Formal settlements in 2009. This proportion decreased to 5% in 2014. Table 4.4 shows large proportions of HHs with no access to electricity for lighting in 2009 (44%) and 2014 (59%) in the Limpopo Province.

Table 4.5: Proportions of HHs with No access to services, Geotype 2015

Area Type	Water	Toilet	Electricity for lighting
1. Urban Formal	10	8	17
2. Traditional	44	56	12
3. Farm	46	36	71
4. Total	100	100	100

Source: Stats SA, Census 2011 questionnaire.
Note: Farm and Traditional were used in 2015 only.
Note: Services denotes water, sanitation and electricity for lighting.

The definition of Geotype variables changed in 2015 as indicated in Table 4.5. It is observed in 2015 from Table 4.5 that 71.0% of Farm households had no access to electricity for lighting. This validates the findings by Noble et al. (2009)

that in SA, the geographical set-up plays a major role in determining the accessibility of resources by communities.

4.4 Analysis of Geotype Data: CD, Census 2011

The Census 2011 data in this section, is analysed at the Capricorn District level (CD).

4.4.1 Access to Tap Drinking Water: CD

The distribution of piped (tap) water by municipality in the CD was closely studied, taking into consideration the distance travelled to access water (Figure 4.3). The results revealed that the highest proportion of HHs in Polokwane (35.3%) access piped water inside the dwelling. This is followed by Lepele-Nkumpi (20.3%). In fact, 89.0% of HHs in Polokwane access tap water inside the dwelling, or inside the yard or within 200m from the dwelling.

Access to piped (tap) water gives interesting results. Less than 10% of HHs in each of the remaining three municipalities in the CD (Blouberg, Aganang and Molemole) access piped (tap) water inside the dwelling, 85.7% of HHs in Aganang access piped water inside the dwelling, or inside the yard or within 200m from the dwelling. The people of Polokwane and Lepelle-Nkumpi municipalities reside predominantly within Urban Formal settlements.

Despite Lepelle-Nkumpi ranking second highest with its people accessing piped (tap) water inside the dwelling, it has the highest proportion of its residents (23.7%) with no access to piped (tap) water, followed by Molemole with 21.0%, and Blouberg with 16.9%. Blouberg further has 8.5% (the highest) of its residents accessing piped water from community stand beyond 0.5km from dwelling.

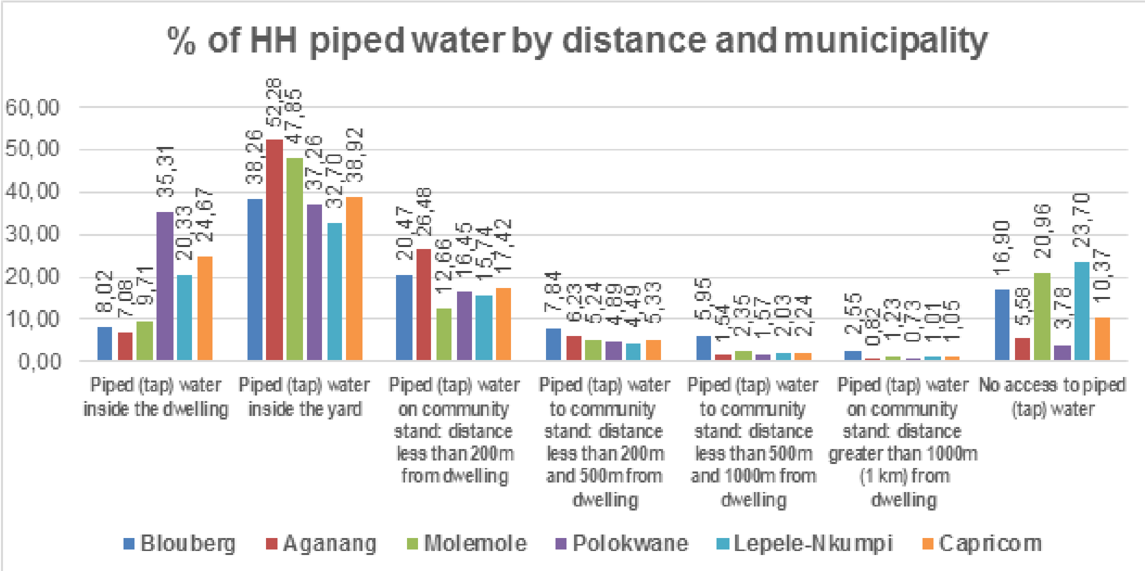

Figure 4.3: Proportion of HHs by piped (tap) water and municipality in the CD, Census 2011

Lepelle-Nkumpi has the largest proportion of HHs (26.7%) either accessing piped (tap) water from community stand beyond 0.5km from dwelling, or having no access to clean water, followed closely by Molemole (25.4%) in this regard. The analysis in this section raises serious concern of discriminatory delivery of clean water within the same municipality (which could most likely spark service delivery protests).

The proportion of HHs in the CD per municipality was studied, according to source of water (see Figure 4.3). The figure indicates that Polokwane (82.2%) and Aganang (75.3%) had their majority of households benefiting from regional/local water scheme and borehole as their alternative water sources than other municipalities. Molemole had the highest proportion (18.0%) of households that sourced water from the water vendors. In Blouberg, 25.7% of households source water from the boreholes. Blouberg had the highest proportion of households sourcing water from the spring (2.1%) and dam/pool/stagnant

(5.9%) than all other municipalities.

Figure 4.3 also indicates that Lepelle-Nkumpi had the highest proportion of households sourcing water from the rain-water tank (1.3%) and the river/stream (4.7%). Water tanker is water that is supplied at small scale by District Authority, and driven cars along the streets in small areas to ease the water shortage. Less than 10.0% of households across all municipalities access this type of service.

4.4.2 Access to Flush Toilet: CD

Table 4.6: Proportions of HHs accessing Type of Toilet by municipality, Census 2011

Toilet Type	Blouberg	Aganang	Molemole	Polokwane	Lepelle-Nkumpi
1. PLVIP	62.6	77.6	58.6	42.8	57.4
2. PLWVIP	13.5	13.1	17.8	6.3	17.2
3. FTSS	6.7	1.7	14.4	43.2	19.8
4. FTST	1.6	0.6	0.7	3.1	0.8
5. None	11.5	5.0	4.0	2,7	3.0
6. Sum(types)	8.3	2.3	15.1	46.3	20.6

Source: Stats SA, Census 2011 questionnaire.
Note: PLVIP denotes Pit latrine with ventilation, PLWVIP denotes Pit latrine without ventilation, FTSS denotes Flush toilet connected to sewerage system, FTST denotes Flush toilet with septic tank and Sum (types) denote Chemical toilet, Bucket and other.

Table 4.5 shows that flush toilet (either with sewerage system or septic tank) features prominently in Polokwane municipality (46.3%). Lepelle-Nkumpi comes second with 20.6% of its HHs having access to flush toilets. Access to flush toilets is most deprived in Aganang (2.3%), followed by Blouberg (8.3%), and then Molemole (15.1%). This realisation shows how unevenly sanitation services are provided in the CD.

Polokwane and Lepelle-Nkumpi being in the urban formal settlements receive much decent sanitation services than the remaining three municipalities. We note, however, that it is less than 50% of HHs enjoying flush toilets in Polokwane and Lepelle-Nkumpi municipalities (46.3% and 20.6%, respectively).

Government authorities in the LP, in particular, the CD, need to speed up the provision of sanitation services. It is inhumane to observe that 11.5% of HHs in Blouberg do not have any kind of sanitation services.

By far the majority of HHs in the CD depends on pit latrine (with or without VIP), with Aganang taking a lead at 90.7%, followed by Molemole, Blouberg and Lepelle-Nkumpi at 76.4%, 76.1% and 74.6%, respectively. The remaining municipality, i.e. Polokwane, has 49.1% of its HHs relying on pit latrine.

4.4.3 Access to Electricity for Lighting: CD

Table 4.7: Proportions of HHs by source of energy / electricity for lighting and municipality, Census 2011

Energy	**Capr**	**Blou**	**Agan**	**Mol**	**Pol**	**Lep**
1. Electricity	87.2	87.2	94.7	95.3	83.1	91.8
2. Gas	0.1	0.1	0.0	0.1	0.1	0.1
3. Paraffin	1.1	0.6	0.3	0.3	1.6	0.7
4. Candles	11.0	11.5	4.6	4.0	14.3	7.1
5. Solar	0.4	0.2	0.2	0.2	0.6	0.2
6. None	0.2	0.3	0.2	0.2	0.2	0.2

Source: Stats SA, Census 2011.
Note: Energy denotes source of electricity for lighting, Capr, Blou, Agan, Mol, pol and Lep denote Capricorn, Blouberg, Aganang, Molemole, Polokwane and Lepelle-Nkumpi respectively.

Table 4.6 shows that there are no major challenges in accessing electricity as the main source of energy for lighting in the CD. All municipalities in Capricorn have over 80.0% of households accessing electricity as source of energy for lighting. Molemole (95.3%) and Aganang (94.7%) municipalities have the highest proportions of households accessing electricity as source of energy for lighting compared to Polokwane (83.1%) in 2011. It should be noted that Polokwane city has several small rural areas such as Dikgale which form part of Polokwane municipality. Table 4.6 indicates that municipalities in the CD also use candles (11.0%) as source of energy for lighting more than paraffin (1.1%) , solar (0.4%) and gas (0.1%). There are some HHs in the CD, albeit a few (0.2%), that have no means of energy for lighting at all.

4.5 Analysis of Small Areas of the CD: Census 2011

In this section, the small areas within the Capricorn District (CD) of the Limpopo Province are selected and analysed, using the Census 2011 data.

4.5.1 Water Source: Small Areas

Tables 6.2 to 6.5 indicate the estimated number of households in a small area by source of water. The study reveals a serious concern based on the small areas with more than 100 households accessing water from *dam/pool/stagnant water* or *river/stream*. Tables 6.2 and 6.3 indicate that Polokwane had 3072 households sourcing water from *dam/pool/stagnant water*, followed by Blouberg with 2417, Lepelle-Nkumpi with 708, Aganang with 501 and Molemole with 257. Thirteen small areas in Polokwane have more than 100 households sourcing water from *dam/pool/stagnant water*, followed by Blouberg with seven small areas.

A small area, Sebayeng, has the highest number (528) of households sourcing

water from a *dam / pool / stagnant water* than other small areas in 2011. Table 6.3 shows that Lepelle-Nkumpi has only one small area, Mphahlele, with more than 100 households (125) sourcing water from *dam / pool / stagnant water*. However most of its small areas source water from the river/stream. Table 6.2 indicates that Aganang and Molemole, each has one small area with more than 100 households sourcing water from a *dam / pool / stagnant*, i.e. Ga-Modikana (131) and Ramokgopa (128).
There is no household sourcing water from a *dam / pool / stagnant water* in the small area called Tshitale, in Molemole municipality, but 139 households in this municipality source water from the *river / stream*. Ga-Rammutla in Blouberg lacked water services in 2011. It is also shown in Table 6.3 that a small area, Lenting in Lepelle-Nkumpi, did not source water from *rain*, *dam / pool / stagnant water* and *river / stream*.

4.5.2 Flush Toilet: Small Areas

Figure 4.4 presents the proportions of households accessing or not accessing flush sanitation after re-coding the data in the Capricorn District. There is a large disparity between toilet accessibility and non-toilet services in Blouberg, Aganang and Molemole.

Figure 4.4 shows a dire shortage of hygienic toilets in Aganang (97.6%), Blouberg (91.6%) and Molemole (82.9%). In all the three municipalities in the CD, more than 80% of HHs are without flush toilets. Polokwane municipality has the highest proportion of households with flush toilets (46.1%), followed by Lepelle-Nkumpi (20.5%). With (53.9%) of HHs without flush toilets, Polokwane municipality is the most advantaged of all other municipalities in the CD. The other municipalities have more than 79.4% of the HHs without clean (flushing) toilets.

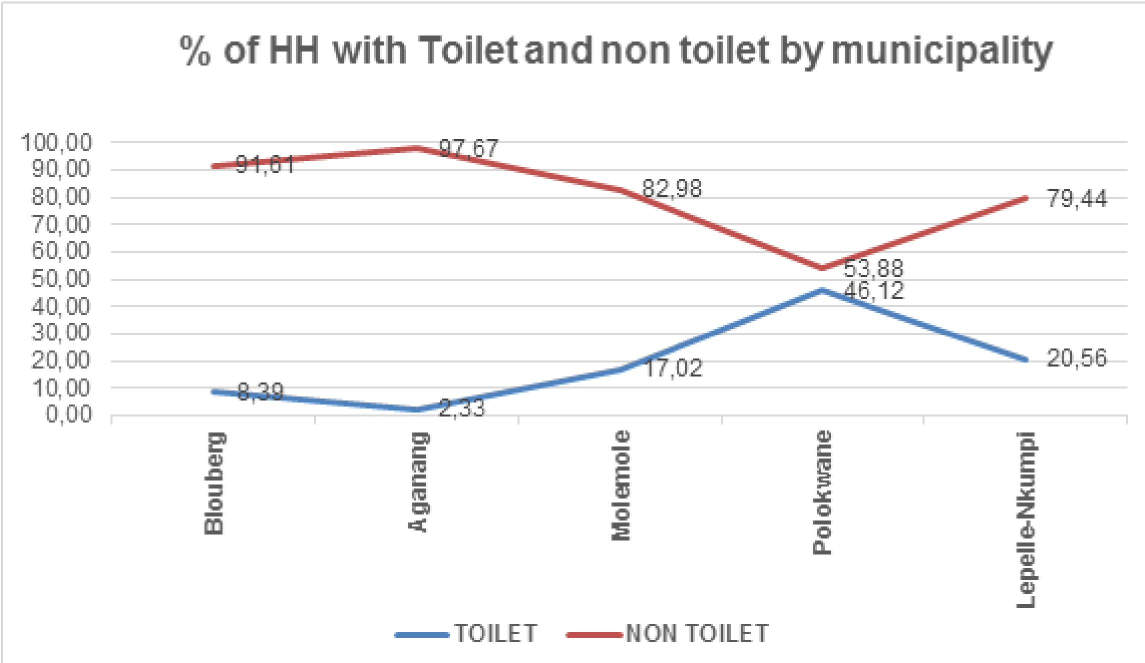

Figure 4.4: Proportion of HHs with toilet and non-toilet in the CD, Census 2011

It should be noted that flush toilets are mainly in white suburbs and townships of the Polokwane City. Figure 4.4 also indicates that there is a huge difference between HHs with flushing toilets and non-toilets in Aganang (95.3%), followed by Blouberg (82.6%), Molemole (66.0%) and Lepelle-Nkumpi (58.9%).

4.5.3 Electricity for Lighting: Small Areas

Table 4.8 and Figure 4.5 show that there are no wards in Aganang and Molemole which have less than 10 households with access to electricity. Polokwane has small areas which have less than 5 households accessing electricity: Makgoba, GaMailula, Sekgweng and Makatiane. It is also noted from Table 4.8 that there is no household that has electricity in a small area, Masenya, which has the highest (36) number of households using Solar energy than all small areas.

Table 4.8: Small areas with less than 10 HHs accessing Electricity, CD

Village	Electricity	Gas	Paraffin	Solar
Blouberg Selowe	10	-	-	-
Blouberg Vienen	4	1	1	-
Blouberg Lekiting	9	-	2	-
Polokwane Vierhoek	5	-	4	2
Polokwane Makgoba	1	1	3	1
Polokwane Masenya	-	-	2	36
Polokwane GaMailula	1	-	-	12
Polokwane GaMakgobathe	6	-	-	3
Polokwane Gamalahlela	5	-	1	6
Polokwane Masekwameng	7	2	4	17
Polokwane Sekgweng	3	1	-	-
Polokwane Makatiane	3	1	26	-
Lepelle-Nkumpi Osterd	-	-	-	1

Note: Village represents Geography by small place name, source of energy for lighting.

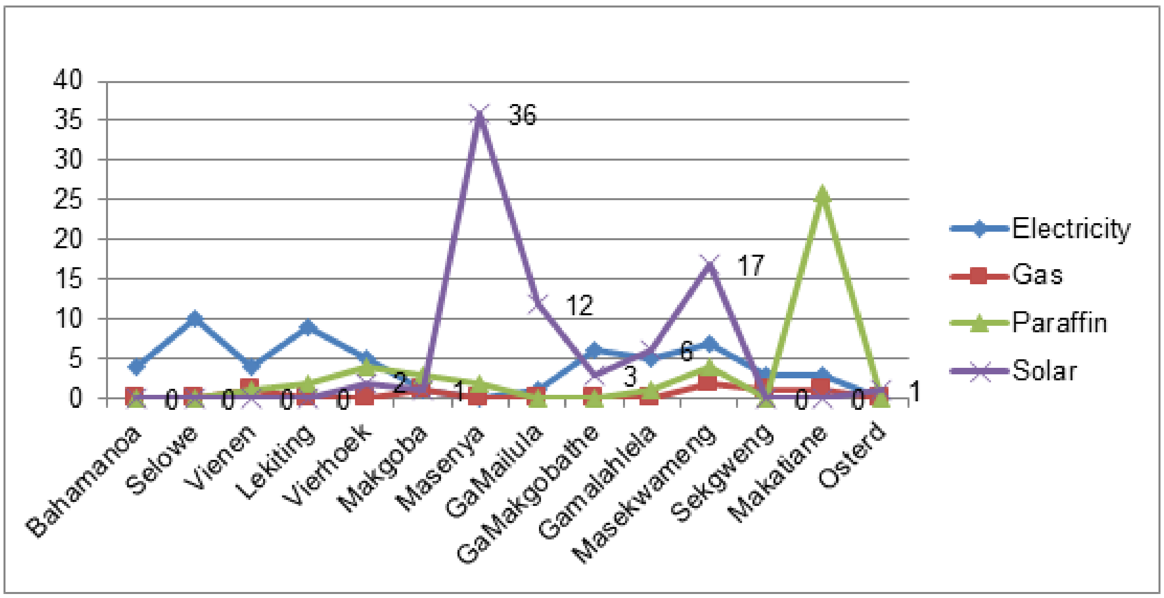

Figure 4.5: Villages with less than 10 HHs accessing electricity by small area

4.6 Summary of Results

In this section, we provide a summary of the results analysed in Chapter 4. Both the Census 2011 and the GHS 2009-2015 data sets that were re-coded in

Chapter 2, have been considered in the analysis. Access to water by households was based only on tap drinking water, access to sanitation was based on flush toilet and access to electricity was bases solely on electricity and solar.

Nationally, the percentages of households accessing flush toilets are far lower than the percentages for water. Table 4.1 indicates that there is no crisis in terms of electricity for lighting across SA. The households in the Eastern Cape and Limpopo were found to be most deprived in the accessibility of basic services focused in this chapter.

The provinces were ranked, for reference, in three categorical order of least, middle and high accessibility of three basic services at country level in Table 4.7. It should be noted that the order of High, middled and least is maintained within the categories.

Table 4.9: Accessibility rank of basic services by province

Province	**Tap water**	**Flush sanitation**	**Electricity for lighting**
Least accessibility	-	LP	-
Middle access	EC, LP, KZN	NC, FS	EC, KZN
High access	WC, GP	WC, GP	WC, FS, NC

Source: Stats SA Census 2011 and GHS 2009-2015.

The summary in Table 4.7 shows that the households in the Eastern Cape, Limpopo and KwaZulu-Natal are ranked at the middle in accessibility of water, where Eastern Cape is the highest among the three provinces. The table shows that Western Cape and Gauteng are ranked high in the accessibility of tap drinking water, and the Western Cape is higher than Gauteng in the accessibility of tap drinking water.

The Limpopo Province is ranked the least in the accessibility of flush sanitation among all provinces. It is observed that the Northern Cape and FS are ranked at the middle while Western Cape and Gauteng are accessing flush sanitation services more than other seven provinces.

In terms of the electricity, no province is ranked the least. However, the percentages of Eastern Cape and KwaZulu-Natal in the accessibility of electricity for lighting is less than the ones for other provinces.

The absence of Gauteng is observed from the high category, whereby the Free State and Northern Cape are ranked at high category with Western Cape in the accessibility of electricity for lighting.

In general, the accessibility of tap drinking water services is ranked at the middle category and the electricity for lighting in the high category across South Africa.

Chapter 5

Model Procedure and Results

5.1 Unit-Level Small Area Model: The Kenward-Roger Estimate Method

For a unit-level small area model, the mixed program in SAS Procedure 5.1 was used to estimate the regression parameters and the variance parameters. The $method =$ option in the Procedure specifies that the Type 1 estimation method should be used. The Type 1 estimation method uses a method of moments estimator which produces an unbiased estimate of the residual variance. The $asycov =$ option requests the asymptotic covariance matrix for the variance parameters.

SAS Procedure 5.1

- $Proc\ mixed\ data = census2011\ method = type1\ asycov\ order = data;$
- $classHMUNIC;$

- $model\ derhhsize = water1toilet1lighting1/;$
- $ddfm = kenwardroger\ solution\ covb\ outp = pred;$
- $randomHMUNIC/cl;$
- $weighthhwgt;$
- $formathmunicMunicipality;$
- $run;$

The $ddfm = kenwardroger$ option in the MODEL statement computes the Mean Squared Error of Prediction (MSEP) and the degrees-of-freedom calculations derived by Kenward and Roger (1997). The procedure is based on considering the true nonlinearity of the mixed model estimates in order to achieve a higher order of accuracy for the estimated covariance of effects. The inclusion of a county-level random effect in the model is specified by The RANDOM statement. Table 5.1 provides model information.

Table 5.1: Unit level Model Information

Model specification	**Model specification**
Data Set	WORK.CENSUS2011
Dependent Variable	DERH-HSIZE
Weight Variable	HH-WGT
Covariance Structure	Variance Components
Estimation Method	Type 1
Residual Variance Method	Factor
Fixed Effects SE Method	Kenward-Roger
Degrees of Freedom Method	Kenward-Roger

Source: Stats SA Census 2011, model information.

A data file of 299891 household observations was used for this analysis with all the missing observations removed from the data. The aim here is to compare the accessibility of the basic services of water, sanitation and electricity for lighting within the local municipalities of the Capricorn District. The household size is used as the dependent variable, assumed to be the number of people in the household. It represents the population densities that have access to the three basic services. Results from the model presented in Table 5.1 are shown in Table 5.2.

Table 5.2: Analysis of variance on water, sanitation and electricity in the CD

Source	**DF**	**SS**	**MS**	**EMS**
Water1	1	2443.455	2443.455	$Var(R) + 6950.3Var(M) + Q(w,t,l)$
toilet1	1	69241	69241	$Var(R) + 11006.0Var(M) + Q(w,t,l)$
lighting1	1	29048	29048	$Var(R) + 3593.1Var(M) + Q(w,t,l)$
H-MUNIC	5	2922.274	584.454	$Var(R) + 41603.0Var(M) + Q((w,t,l)$
Residual	299882	1830604	6.104	$Var(R)$

Note: Letters *w, t, l* and *R* denote: water1, toilet, lighting1 and Residual respectively; MS, EMS and DF denote Mean Square, Expected Mean Square and Degree of Freedom respectively.

Table 5.3: Analysis of variance by water, sanitation and electricity in the CD (Cont).

Source	**Error**	**E.DF**	**F value**	$Pr > F$
Water1	$0.1671MS(M) + 0.8329MS(R)$	5.53	23.79	0.00
toilet1	$0.2646MS(M) + 0.7354(R)$	5.29	435.17	< 0.00
lighting1	$0.0864MS(M) + 0.9136MS(R)$	6.17	518.22	< 0.00
H-MUNIC	$MS(R)$	299882	95.74	< 0.00
Residual	-	-	-	-

Note: E.DF denotes Error Degree of Freedom.

Tables 5.2, 5.3 and 5.4 show the strong evidence that the accessibility of services differ based on the Type 1 F tests. The variance component of the local municipalities is 6.1044 while the No access to water, sanitation and electricity services (within municipalities) variance component is 0.0139, as shown in Table 5.4. The variance components provide methods to estimate by applying the *Intraclass Correlation Coefficient* (ICC).

Table 5.4: Results on Covariance Parameter Estimates

Cov Parm	Estimate
H-MUNIC	0.0139
Residual	6.1044

Source: Stats SA Census 2011.

Based on the covariance estimates in Table 5.4, the ICC is computed as:
$0.0139/(0.0139 + 6.1044) = 0.002272$,
which accounts for the portion of the total variance that occurs within municipalities of the Capricorn District.

Table 5.5: Asymptotic Covariance Matrix of Estimates

Row	Cov Parml	CovP1	CovP2
1	H-MUNIC	0.000058	1.10E-07
2	Residual	1.10E-07	0.000249

Source: Stats SA Census 2011. Model calculations.

The standard errors for the estimated covariance parameters are the square root of the diagonals of the estimated asymptotic covariance matrix. Therefore, Table 5.5 shows that the standard error of the estimate for σ_μ^2 is 0.0076 (which is derived from $(\sqrt{0.000058})$), while the standard error of the estimate for σ_e^2 gives 0.01578 (which is $(\sqrt{0.000249})$).

The intercept (2.9074) in Table 5.6 represents the estimated mean for all variables: water, sanitation and electricity for lighting. The estimated toilet mean (-1.0761) of access to flush toilet, indicates that the accessibility of flush toilet is far less than the accessibility of water (0.06016) and electricity for lighting (0.8918) in the Capricorn District.

The study also notes that the estimated access to water is less than that of electricity. The intercept, 2.9074, was added to the estimate of each variable to compute the estimated accessibility mean for each variable (see Table 5.6). Therefore, the estimated mean for water1 is 2.96756, for toilet1 is 1.8313 and for lighting1 is 3.7992. Related to objective number 6, this indicates that on average, services for electricity for lighting are more accessible than services for water and sanitation in the CD. This seems to agree with the results in the previous chapters.

The magnitude of the variation among municipalities in their mean access to tap drinking water, flush toilet and electricity for lighting, can be used to calculate the range of plausible values for these means, based on the between variance obtained from the model:
$2.9074 \pm 1.96 * \sqrt{0.0139} = (0.1179, 0.3728)$. (Refer to Table 5.4).

Table 5.6: Solution for Fixed effects: water, toilet and electricity

Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	$Pr > \|t\|$
Intercept	2.9074	0.05561	8.21	52.28	0.0001
water1	0.06016	0.01453	300000	4.14	0.0001
toilet1	-1.0761	0.01026	290000	-104.88	0.0001
lighting1	0.8918	0.01325	300000	67.29	0.0001

Source: Stats SA Census 2011. Model calculations.

Table 5.7: Covariance matrix for fixed effects by water, toilet and electricity

Row	**Effect**	**Col1**	**Col2**	**Col3**	**Col4**
1	Intercept	0.003093	-0.00018	0.000024	-0.00015
2	water1	-0.00018	0.000211	-0.00002	-0.00000263
3	toilet1	0.000024	-0.00002	0.000105	-0.00003
4	lighting1	-0.00015	-0.00000263	-0.00003	0.000176

Source: Stats SA Census 2011. Model calculations.

The structures in Table 5.8 result from the use of random effects parameters, which are additional unknown random variables assumed to influence the variability of the data. The variances of the random effects parameters, commonly known as variance components, become the covariance parameters for the particular structure such as in Table 5.8.

Table 5.8: Solution for Random effects by local municipality in the CD

Munic.	**Est.**	**SE**	**DF**	**t.val**	$Pr > \|t\|$	α	**L**	**U**
Blouberg	0.1192	0.05382	7.19	2.21	0.0613	0.05	-0.0074	0.2458
Aganang	-0.00671	0.05403	7.30	-0.12	0.9045	0.05	-0.1334	0.1200
Molemole	-0.2111	0.05413	7.36	-3.90	0.0054	0.05	-0.3378	-0.0843
Polokwane	-0.02364	0.05327	6.91	-0.44	0.6708	0.05	-0.1499	0.1027
Lepelle-Nkum	0.1222	0.05357	7.07	2.28	0.0562	0.05	-0.0042	0.2486

Note: Munic, Est., SE, L, U denote Municipality, the Parameter Estimates, Standard Error, Lower and Upper confidence Intervals, respectively.

The solution for random effects in Table 5.8 provides the EBLUP estimates related to objectives 3 and 5. The EBLUPs in Table 5.8 are generated from the solution of fixed effects in Table 5.7 which gives the intercept, slope, and nested errors. Table 5.8 indicates the estimates predictor, standard error, the confidence interval per local municipality of the CD and other parameters as indicated in the table.

It is shown that the confidence interval (CI) for Blouberg is (-0.0074;0.2458) and the corresponding parameter estimate for Blouberg is 0.119 which lies within the CI. Similarly for Aganang, the CI is (-0.1334;0.2), hence -0,00671 lies within the CI, for Polokwane the CI is (-0.1499;0.1027) and -0.02364 lies within its CI; and for Lepelle-Nkumpi the CI is (-0.004;0.2486) and 0.1222 lies within its CI but the estimate for Molemole (-0.2111) behaves differently from the estimates of other municipalities since its CI (-0.3378;-0.0843). This indicates that all local municipalities are not significant at 5.0% level, except for

Molemole with an estimate (-0.2111), which is not outside its CI but too closer to the lower bound of its CI.

Table 5.9: Type 3 Tests of Fixed effects for water, toilet and electricity

dy**Effect**	**DF**	**Den DF**	**F-value**	$Pr > F$
water1	1	300000	17.14	0.0001
toilet1	1	290000	11000.5	0.0001
lighting1	1	300000	4528.33	0.0001

Source: Stats SA Census 2011. Model calculations.

The Type 3 tests of fixed effects in Table 5.9 display tests for all of the fixed effects. These tests are partial in the sense that they account for all of the other fixed effects in the model. The table shows that the test for water1, toilet1 and lighting1, are all significant at the 5.0% level (reference can also be made to Table 5.6).

5.2 Area-Level (Small Area Model)

In SAS Procedure 5.2, $Proc\ mixed \ldots = reml$ estimates the regression parameters and the covariance parameters for the area-level model. The $method = reml$ option in the $model$ statement specifies that the residual (restricted) maximum likelihood method should be used in the procedure to estimate the covariance parameters. The class statement confirms the variable $H\ MUNIC$ to be a class variable. The model statement specifies Y as the dependent variable and $derhhsize$ as the only independent variable in the model. The solution option computes a solution for the fixed-effects parameters, and the $covb$ option produces the approximate variance-covariance matrix of the fixed-effects parameter estimates $\hat{\beta}$.

SAS Procedure 5.2

- $Proc\ mixed\ data = census2011asycovmethod = reml;$
- $classHMUNIC;$
- $modelderhhsize = water1toilet1lighting1/solutioncovb;$
- $randomHMUNIC/cl;$
- $ods\ output\ covb = covbeta$
- $solutionF = beta$
- $covparms = sigma2$
- $asycov = aCovSigma2;$
- $weight\ hh\ wgt;$
- $formath\ municMunicipality;$
- $Run;$

The random statement defines the random effects and specifies that a local municipality $(H\ MUNIC)$ random effect be included in the model. The $ods\ output$ statement specifies that the covariance matrix of fixed-effects parameter estimates, the fixed-effects solution vector, the estimated covariance parameters, and the asymptotic covariance matrix of covariance parameters be saved in the SAS procedure $data\ sets\ covbeta, beta, sigma2$ and $covsigma2$, respectively. These data sets will be used to compute the EBLUPs and their standard errors. Table 5.10 gives model information. The covariance parameter estimates in Table 5.11 show similar results to those presented in Table 5.5. However, a negative estimate change of -0.1208 in the Municipality variable indicates the decline in three services (water, toilet and electricity variables) across municipalities

Table 5.10: Small Area Model Information

Model specification	Model specification
Data Set	WORK.CENSUS2011
Dependent Variable	DERH-HSIZE
Weight Variable	HH-WGT
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Source: Stats SA Census 2011. Model calculations.

Table 5.11: Results on Covariance Parameter Estimates

Covariance Parameter	Estimate
H-MUNIC	0.01817
Residual	6.1044

Source: Stats SA Census 2011. Model calculations.

Table 5.12: Asymptotic Covariance Matrix of Estimates

Row	Covariance Parameter	Covariance P1	Covariance P2
1	H-MUNIC	0.000058	1.10E-08
2	Residual	1.10E-08	0.000249

Source: Stats SA Census 2011.

(H-MUNIC variable).

The standard errors for the estimated covariance parameters are the square root of the diagonals of the estimated asymptotic covariance matrix. Therefore, the standard error of the estimate for σ^2_μ is 0.0076 (which is the $(\sqrt{0.000058})$). The one for σ^2_e is 0.01578 (the $(\sqrt{0.000249})$), which increased slightly compared to the one in Table 5.5. The slight difference is seen between the residual estimates in Table 5.5 and Table 5.12.

Table 5.13: Solution for Fixed Effects

Effect	Estimate	SE	DF	t-Val.	$Pr > \|t\|$
Intercept	2.9073	0.06281	5	46.29	0.0001
water1	0.06016	0.01453	300000	4.14	0.0001
toilet1	-1.0761	0.01026	300000	-104.9	0.0001
lighting1	0.8918	0.01325	300000	67.3	0.0001

Source: Stats SA Census 2011. Model calculations.

Table 5.13 validates the results shown in Table 5.6 that the flush toilet mean estimate (-1.0761) is less than the mean estimate for water and electricity for lighting. This implies that people in the Capricorn District are more deprived from the accessibility of flush toilet than of water and electricity. However, it does not mean that there is no problem of service delivery in terms of water and electricity. Table 5.13 indicates that electricity services are more accessed than water and sanitation.

Table 5.14: Solution for Random effects by municipality in the CD

Munic.	Est.	SE	DF	t.val	$Pr > \|t\|$	α	L	U
Aganang	-0.006	0.061	3.00E+05	-0.11	0.913	0.05	-0.127	0.1137
Blouberg	0.119	0.061	3.00E+05	1.95	0.051	0.05	-0.00	0.2396
Lepelle-Nkumpi	0.123	0.061	3.00E+05	2.01	0.045	0.05	0.002	0.242
Molemole	-0.212	0.062	3.00E+05	-3.44	0.001	0.05	-0.332	-0.09119

Source: Stats SA Census 2011. Model calculations.
Note: Munic., Est., SE, L, U denote Municipality, Parameter Estimates, Standard Error, Lower and Upper confidence Intervals respectively.

The solution for random effects in Table 5.14 indicates that Blouberg, Lepelle-Nkumpi and Molemole are significant at the 5.0% level compared to the results of the previous model shown in Table 5.8, where only one municipality (Molemole) is significant. This implies that households in these municipalities are severely affected in terms of accessing water, flush toilets and electricity for lighting than in Aganang and Molemole municipalities. Results in Table 5.8 are modified: the residual (restricted) maximum likelihood method is used in the program to estimate the covariance parameters. The $H\ MUNIC$ was specified as class variable and the model excludes outliers such as Polokwane.

Table 5.15: Type 3 Tests of Fixed effects for water, toilet and electricity in the CD

dy**Effect**	**Num DF**	**Den DF**	**F-val.**	$Pr > F$
water1	1	3.00E+05	17.14	0.0001
toilet1	1	3.00E+05	11003.6	0.0001
lighting1	1	3.00E+05	4529.03	0.0001

Source: Stats SA Census 2011. Model calculations.

Type 3 tests of fixed effects in Table 5.15 refine the Type 3 tests of fixed effects computed in Table 5.9. The study performed the tests noting that fixed effects account for all of the other fixed effects in the model. Although there is a slight change in the results between Tables 5.5 and 5.12, water1, toilet1 and lighting1, are significant at the 5.0% level.

In SAS Procedure 5.3 *Proc iml* was used in conjunction with other model procedures for both **unit** and **area** level models of SAE. The procedure provides outputs which compare all model-based estimates discussed theoretically in Chapter 3 of this study, in particular, to produce Hierarchical Bayes estimates[1]:

SAS Procedure 5.3

- $Proc\ iml$
- $proc\ iml\ symsize = 3000000000\ worksize = 3000000000;$
- $use\ beta; read\ allvarestimateintobet;$
- $use\ covbeta;\ readallvar\ num\ intocovb;$

[1]The $\sharp$ in the SAS code has been replaced with no.

- $use\ sigma2;\ readall\ var\ estimate\ into\ der\ sigma2;$
- $use\ aCovSigma2;\ read\ all\ var\ CovP1into\ acSigma2;$
- $use\ census2011;\ readallvar$
  $derhhsize\ water1\ toilet1\ lighting1intodat;$
- $nobs = nrow(dat); /*Numberofhhinthecensusdata*/$
- $np = nrow(bet); /*Numberofbeta\ estimates*/$
- $der\ hhsize = dat[, 1]; /*From\ dat\ matrix\ copie\ only\ the\ first\ variable*/$
- $one = J(nobs, 1, 1); /*For\ total\ number\ of\ Observations\ create\ contains1*/$
- $/*Deal\ with\ the\ missing\ values\ fromXI*/$
- $X = J(nobs, 1, 1)||dat[, 2 : np];$
- $d = dat[, np];$
- $sigma2Vec = der\ sigma2 * one;$
- $covb = covb[, 2 : np + 1];$
- $gamma = sigma2Vec/(sigma2Vec + d);$
- $EBLUP = gamma\ no\ der\ hhsize + (one - gamma)\ no\ (XI * bet);$
- $doi = 1tonobs;$
- $ifder\ hhsize[i] = .then$
- $EBLUP[i] = XI[i,] * bet;$
- $end;$
- $g1i = gamma\ no\ d;$
- $XCovBXT = XI * covb * XI`; /*TransposingXIgivesusanerrorof*/$

- $g2i = (one - gamma)no2novecdiag(diag(XCovBXT));$
- $avSigma2 = 1/sum((sigma2Vec + d)\ nono\ (-2));$
- $avSigma2 = 2 * avSigma2;$
- $g3i = ((d\ no\ 2)\ no\ ((d + sigma2Vec)\ no\ (-3))) * acSigma2;$
- $mse = g1i + g2i + 2 * g3i;$
- $doi = 1tonobs;$
- $ifder\ hhsize[i] = .then$
- $mse[i] = XI[i,] * covb * XI[i,]` + der\ sigma2;$
- $end;$
- $Create\ outData1$
- $var/ * betcovbder\ sigma2\ acSigma2\ dat\ nobs\ npder\ hhsizeoneXI$
- $gamma\ d\ sigma2Vec * /EBLUP$
- $/ * XCovBXT * /\ g1i\ g2i\ g3i\ mse\ avSigma2;$
- $append;$
- $close\ outData1;$
- $quit;$

5.3 Unmatched Models

The unmatched models were performed in order to estimate undercoverage in the Capricorn District using Census 2011 data. The variance was calculated through $Proc\ means$ using 10% Census data.

SAS Procedure 5.4

- $Proc\ means\ data = a\ mean\ var$;
- $var\ der\ hhsize$;
- $class\ HMUNIC$;
- $weight\ hh\ wgt$;
- run;

The $Proc\ means$ in SAS Procedure 5.3 was used to calculate variance from the Census 2011 data, which is considered to be exhaustive, but this study detected an existence of missing values. Therefore, missing values were considered as complete missing or undercoverage. The results obtained from previous models exclude missing values. However, this study found it necessary to account for missing values or undercoverage. Two data sets were created: one containing all the missing data and the other called A, containing non-missing data.

SAS Procedure 5.5

- $Data\ A$;
- $set\ census$ 2011;
- $if\ cmiss(of\ ALL) = 0$;
- run;

The Census 2011 data set A containing non-missing data (SAS Procedure 5.5), was used throughout the analysis of this study.

The SAS Procedure 5.6 uses Markov Chain Monte Carlo procedure to estimate the model parameters and the small area undercoverage counts and rates. The input data set named undercoverage, accommodating Municipality length, and contains the variables Index, Census Count, Missing, and *D* which indicates variability. The variable *Missing* contains the direct estimates (refer to Section 3.2.1 of this study) of the undercoverage count, denoted by Mi (Rao, 2003), which represents the difference between the value observed from the full data set and the data set *A* that excludes missing values. The variable *D* contains variances calculated in the SAS Procedure 5.6.

SAS Procedure 5.6

- $data\ undercoverage$;
- $length\ Munic15$;
- $input\ Index\ Munic\ Code\ CensusCount\ Missing\ D@@$;
- $datalines$;

Table 5.16: Census 2011 undercounts by Municipality in the CD

Municipality	Code	Census Count	Missing	D
Aganang	970	30083	1306	12.3558395
Blouberg	969	36254	593	9.0933081
Lepelle-Nkumpi	976	52445	572	12.3250238
Molemole	973	25959	278	14.3151933
Polokwane	974	155150	158	18.338558

Source: Stats SA Census 2011. Model calculations.

Polokwane has the highest Censuscount (155150), highest variability (1833) and lowest missing (158) households compared to all other municipalities in the Capricorn District. Lepelle-Nkumpi (524450) is the second highest from Polokwane in Censuscount, followed by Blouberg (36254), Aganang (30083) and Molemole (25959) in that order. Blouberg has the lowest variability (9.09) among all Capricorn District municipalities. Although Aganang has the highest missed counted households during the Census 2011 enumeration, its variability (12.35) is close to that of Lepelle-Nkumpi (12.32).

The SAS Procedure 5.7 contains statements in $Proc\ mcmc$ model explained as follows: ods graphics directs $Proc\ mcmc$ to produce MCMC diagnostic plots together with the results of the model. The nmc option in the $Proc\ mcmc$ specifies the number of MCMC iterations, excluding the burn-in iterations. The $nthin$ option controls the thinning rate of the simulation, and the nbi option specifies the number of burn-in iterations. The $output\ post$ option gives the output data set for posterior samples of parameters. The $monitor$ option tells $Proc\ mcmc$ to generate output for the specified symbols of interest. The two arrays, m and u respectively denote undercoverage count M_i and the undercoverage rate U_i for each local municipality, which will be created by the $Proc\ mcmc$. The two $parms$ statements specify the parameters of the $Proc\ mcmc$ model. One $parm$ option specifies two regression coefficients, β_0 and β_1; with initial values equal to 1. The other $parm = option$ computes the random effects variance parameter, denoted by S^2.

SAS Procedure 5.7

- $Proc\ mcmc$;
- $ods\ graphics\ on$;
- $procmcmcdata = undercoveragenmc = 45000nthin = 5nbi = 50000seed = 123456$
- $outpost = o1monitor = (parms - mu)$
- $stats = (summaryinterval)diag = none; *(mcseess)$;
- $arraym[5]$;
- $arrayu[5]$;

- $parm(beta0beta1)1;$
- $parm\ s2;$
- $priorbeta:\ general(0);$
- $priors2\ igamma(shape = 0.05, scale = 0.05);$
- $randomgamma\ n(beta0 + beta1 * log(censuscount), var = s2)subject = Munic;$
- $m[Index] = censuscount * exp(-gamma)/(1 - exp(-gamma));$
- $u[Index] = exp(-gamma);$
- $modelmissing\ n(m[Index], var = d);$
- $ods\ output\ post\ summaries = est;$
- $run;$
- $ods\ graphics\ off;$

The theory supporting Equation (3.31) through to Equation (3.33) requires that a prior distribution be specified for each parameter. The $prior\ beta$ option in SAS Procedure 5.7 indicates a general distribution for the regression coefficients, which computes what is known as a *flat* prior. The prior S^2 option for the random effects variance parameter S^2 is specified as an inverse-gamma and its shape and scale parameters are both set to 0.05.

The $random$ statement (SAS Procedure 5.7) defines a random effect and its prior distribution. The $subject$ option identifies our subjects, Municipalities, in the random effects model. These random effects parameters associated with each subject are assumed to be conditionally independent of each other given other parameters in the model. In particular, the random effect is called Gamma, and it is specified to have a normal distribution with a mean equal to

$Beta0+Beta1*log$ (Census 2011 count) and a variance of S^2. The $m[Index]...and\ u[Index]...$ specify equations for M_i and U_i, and results of these computations are saved in arrays *m[5]* and *u[5]*. The $model$ statement specifies the complete Small Area Estimation model, which includes the sampling model for M_i and the linking model for $\log(U_i)$. The final statement, $ods\ output\ post\ summaries = est$, directs the M_i and U_i in the SAS Procedure 5.7 to create a data set named Est in order to save statistics such as the sample size, mean, standard deviation, percentiles and the posterior summaries for each parameter (Mukhopadhyay and McDowell, 2011).

Table 5.17: MCMC Procedure, results on Posterior Summaries

Parameter	N	Mean	SD	25 p.tile	50 p.tile	75 p.tile
beta0	9000	-13.3126	8.6947	-17.5912	-13.5293	-9.4219
beta1	9000	1.6712	0.8098	1.3112	1.6896	2.0654
s2	9000	1.3764	4.0421	0.3739	0.6357	1.2063
m1	9000	1306.0	3.5220	1303.6	1306.0	1308.4
m2	9000	593.0	3.0045	591.0	593.0	595.0
m3	9000	572.0	3.5290	569.6	572.0	574.3
m4	9000	278.0	3.7725	275.4	277.9	280.5
m5	9000	158.0	4.3143	155.1	158.0	160.9
u1	9000	0.0416	0.000108	0.0415	0.0416	0.0417
u2	9000	0.0161	0.000080	0.0160	0.0161	0.0161
u3	9000	0.0108	0.000066	0.0107	0.0108	0.0108
u4	9000	0.0106	0.000142	0.0105	0.0106	0.0107
u5	9000	0.00102	0.000028	0.000999	0.00102	0.00104

Source: Stats SA Census 2011. Model calculations. **Note**: p.tile denotes percentiles.

The posterior summaries output in Table 5.17 shows the number of posterior samples, the posterior mean (Mean), standard deviation (SD) and percentile estimates (p.tile). It is deduced from Table 5.17 that the arithmetic signs of the regression coefficients of $beta0$ and $beta1$ have opposite arithmetic signs. This was influenced by the specification of the parameter Gamma in the SAS Procedure 5.7 explained above. Reversing the sign on Gamma in Equations M_i and

U_i, gives a better range for the estimated quantities of interest.

Table 5.18: MCMC Procedure, results on Posterior Intervals

Parameter	α	**L-Equal-Tail**	**U-Equal-Tail**	**L-HPD**	**U-HPD**
beta0	0.050	-29.3159	3.9296	-30.7049	2.2585
beta1	0.050	0.0840	3.1600	0.1618	3.2116
s2	0.050	0.1694	6.7962	0.0689	4.0465
m1	0.050	1299.1	1312.8	1299.1	1312.8
m2	0.050	587.0	598.9	586.9	598.8
m3	0.050	565.2	578.9	565.2	578.9
m4	0.050	270.5	285.4	270.8	285.7
m5	0.050	149.5	166.4	149.7	166.6
u1	0.050	0.0414	0.0418	0.0414	0.0418
u2	0.050	0.0159	0.0163	0.0159	0.0162
u3	0.050	0.0107	0.0109	0.0107	0.0109
u4	0.050	0.0103	0.0109	0.0103	0.0109
u5	0.050	0.000963	0.00107	0.000964	0.00107

Source: Stats SA Census 2011. Model calculations.
Note: L and U denotes the lower and upper Equal-Tail Intervals; L-HPD and U-HDP denote the Lower and Upper highest posterior densities.

Table 5.18 displays the equal tail interval and the Highest Posterior Density (HPD) interval for each parameter, where *Proc mcmc* automatically generates the trace, autocorrelation, and kernel density plots that are shown in Figures 5.15, A5.1 and 5.17. To produce these plots, *Odds graphics* need to be activated. A trace plot provides researchers with evidence of whether or not the Markov chain has converged to its stationary distribution. Stationarity aspects like constant mean and variance can be recognised from a trace plot. If the distribution of points does not change as the chain progresses, this might imply that a chain has reached stationary state. One can also detect from the trace plot if the chain is mixing well. A chain is said to be mixing well if it traverses its posterior space quickly and it can skip from one remote region of the posterior to another in comparatively few steps.

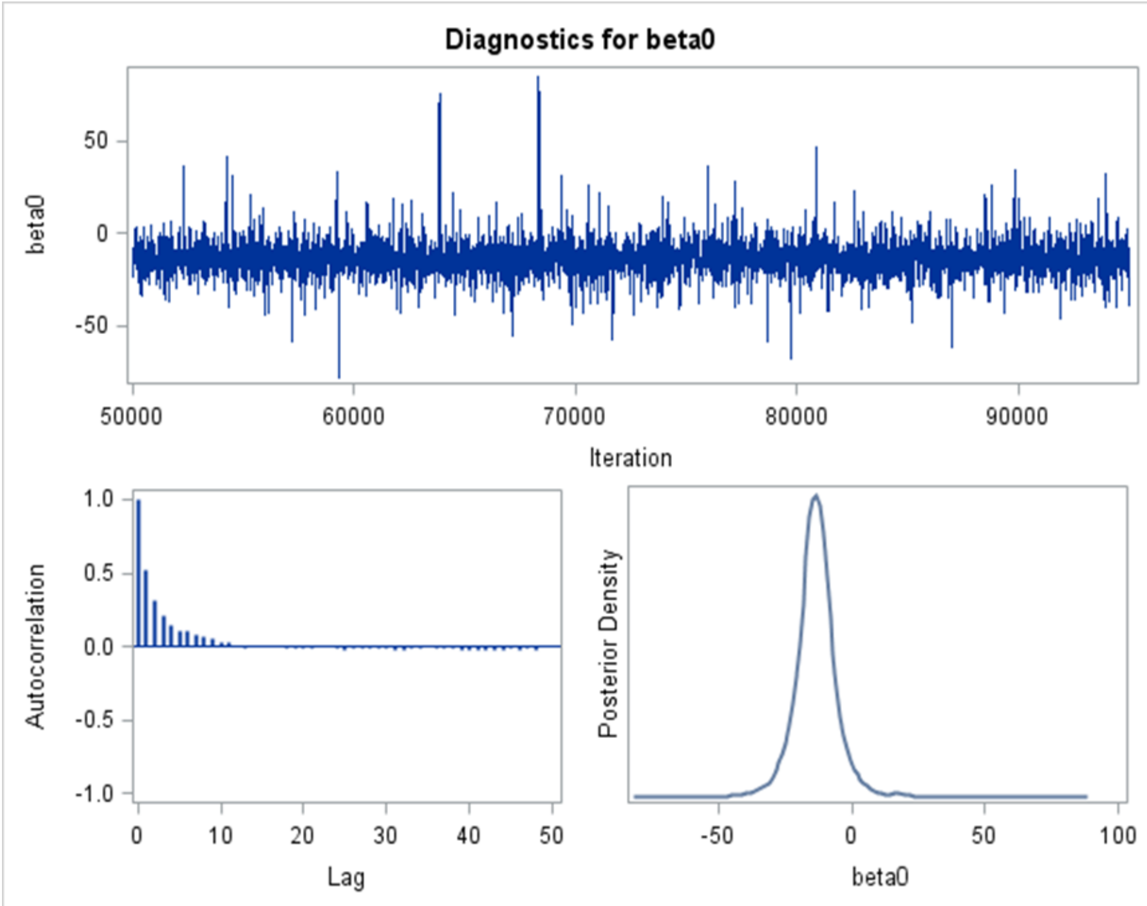

Figure 5.1: MCMC Procedure, results on diagnostic for Beta 0 (zero)

The trace plots in Figures 5.1, 5.2 and 5.3 indicate that the Markov chain has reached stable state and looks constant for all the three variables: $beta1, beta2$ and s2. These three graphs (Figures 5.1, 5.2 and 5.3) show that the Markov chains have mixed well.

The autocorrelation plots in Figures 5.1, 5.2 and 5.3 confirm the tabular information in Tables 5.17 and 5.18. The kernel density plot estimates the posterior marginal distribution. The $nthin$ option in the $Proc\ mcmc$ statement controls the thinning which can control the autocorrelations among the posterior samples. The $nthin = 5$ is sufficient for this study because there are five municipalities in the Capricorn District. The diagnostics plots for M_i and U_i were not shown because they have similar outcomes as in Figure 5.2.

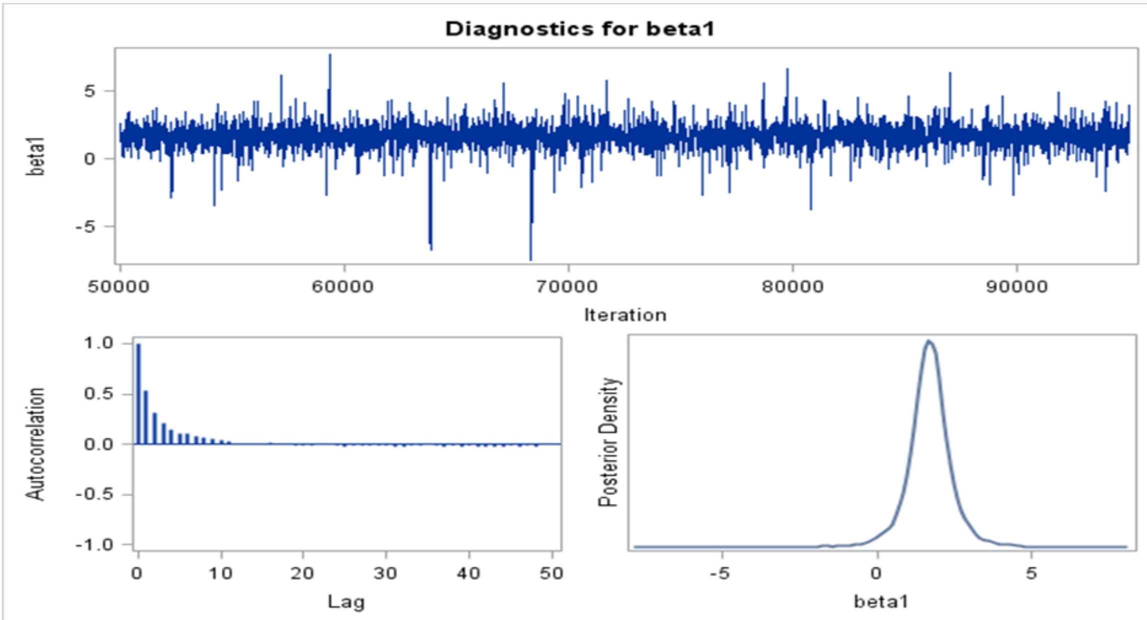

Figure 5.2: MCMC Procedure, results on diagnostic for Beta 1

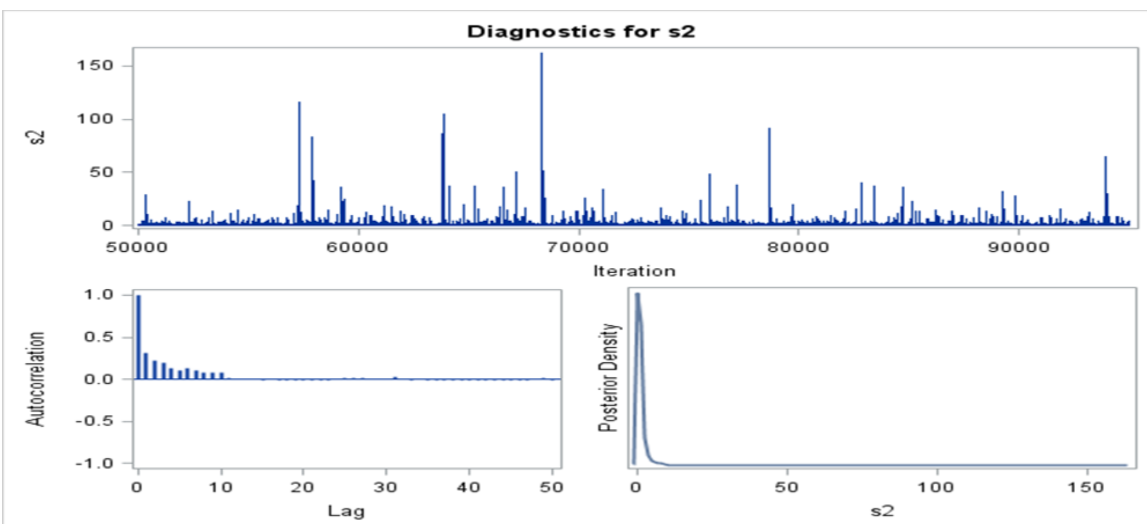

Figure 5.3: MCMC Procedure, results on diagnostic for S^2

The output data est in the SAS Procedure 5.8 was used to derive results shown in Tables 5.19 and 5.20, which are Hierarchical Bayes estimates of the posterior means of the undercoverage Census 2011 count M_i and the coefficients of variation for the estimates.

SAS Procedure 5.8

- $data\ est\ count$;
- $merge\ Undercoverage\ est\ (firstobs = 4\ obs = 8)$;
- $CVHB = StdDev/Mean$;
- $CVD = sqrt(D)/Missing$;
- $labelMissing =' Direct\ Estimate\ for\ Undercount'$
- $Mean =' HB\ Estimate\ for\ Undercount'$
- $StdDev =' Standard\ Deviation\ for\ the\ HB\ Estimator'$
- $CVHB =' CV\ for\ the\ HB\ Estimator'$
- $CVD =' CV\ for\ the\ Direct\ Estimator'$;
- run;
- $proc\ print\ data = estcount\ label\ noobs$;
- $varMunic\ CensusCount\ Missing\ Mean\ StdDev\ CVHB\ CVD$;
- $format\ Mean\ 10.1$
- $CVHB\ 4.6$
- $CVD\ 4.6$;
- run;

Table 5.19: HB Estimates for the undercounts by municipalities in the CD

Munic	Census Count	D.est	HB Est	SD-HB	CV-HB	CV-Direct
Aganang	30083	1306	1306.0	3.5220	0.002697	0.002691
Blouberg	36254	593	593.0	3.0045	0.005067	0.005085
Lepelle-Nkumpi	52445	572	572.0	3.5290	0.006170	0.006138
Molemole	25959	278	278.0	3.7725	0.013572	0.013610
Polokwane	155150	158	158.0	4.3143	0.027310	0.027104

Note: HB denotes Hierarchical Bayes estimates.

Table 5.19 indicates the variability of households that were undercounted in the CD for both direct and indirect SAE. It is observed from Table 5.19 that the undercount estimates are equal for direct and HB estimates; and also, the estimate model CV for the HB estimate does not differ with the design CV for the direct estimates in all municipalities.

SAS Procedure 5.9 uses the output data set est to generate the HB estimates of the posterior means of the undercoverage rate U_i and coefficients of variation for the estimates as shown in the Procedure.

SAS Procedure 5.9

- $data\ est\ rate$;
- $merge\ undercoverage\ est\ (firstobs = 9)$;
- $CVHB = StdDev/Mean$;
- $UR = Missing/(Missing + CensusCount)$;
- $labelUR =' Direct\ Estimate\ for\ Undercoverage\ Rate'$
- $Mean =' HB\ Estimate\ for\ UndercoverageRate'$
- $StdDev =' Standard\ Deviation\ for\ the\ HB\ Estimator'$
- $CVHB =' CV\ \ for\ the\ HB\ Estimator'$;

- run;
- $proc\ print\ data = estrate\ label\ noobs$;
- $var\ Munic\ CensusCount\ UR\ Mean\ StdDev\ CVHB$;
- $format\ UR$ 6.6
- $Mean$ 6.6
- $StdDev$ 6.6
- $CVHB$ 6.6;
- run;

Table 5.20: Hierarchical Bayes Estimates for the Undercoverage Rates

Munic	**Census Count**	**D-rate**	**HB-rate**	**SD-HB est**	**CV-HB**
Aganang	30083	0.041607	0.0416	0.000108	0.002585
Blouberg	36254	0.016094	0.0161	0.000080	0.004985
Lepele-Nkumpi	52445	0.010789	0.0108	0.000066	0.006103
Molemole	25959	0.010596	0.0106	0.000142	0.013428
Polokwane	155150	0.001017	0.00102	0.000028	0.027282

Note: D-rate, HB-rate, SD-HD est., CV-HB respectively denote Direct estimate rates, Hierarchical Bayes estimate for undercoverage Rate, Standard deviation for HB estimates and covariates HB estimates.

Table 5.20 presents the prediction statistics (standard deviations and model CV for the HB estimates) for the undercoverage rate. The model CV for the HB estimates for the undercoverage rate, range from 0.0026% to 0.027%. Table 5.20 shows that the undercoverage rates for the direct estimates are almost equal to the undercoverage rate for the HB estimate in Molemole and Polokwane municipalities.

5.4 Data Limitations

Handling missing values was one of the fundamental issues, especially when dealing with metrics. The case of missing values in numerical data is critical, hence all missing values in each row in the data set were deleted, and subset data of the Capricorn household observations retained 299891 households. A total of 2907 rows/household observations were dropped due to missing values. In contrast, often when using SAS/IML procedure, the first step is to delete rows of the data matrix that contain missing values. Transposing 299891 household observations posed a technical problem of space allocation in SAS. This study could not produce reliable MSE from the models due to this space allocation challenge.

Results emanating from the challenges posed by the space allocation can be referred to Tables 5.21 and 5.22 which have presented only 20 rows per municipality of the Capricorn District. In Appendix A5.3, the MSE value (0.87) is the same for all households in the Capricorn District. The developers of SAS package recommend big computer RAM. Therefore, the transpose was not conducted because it gave error. The alternative way was to run $Proc$ option to assess SAS system memory. The $Proc$ options did not produce solutions. Several solutions were tried in order to increase the system memory size such as the $Procimlsymsize = n1worksize = n2$, but the procedure that transposes matrix signalled an error message.

Tables 5.21 and 5.22 indicate the size of the households in different municipalities of the Capricorn District. The size here refers to the number of household members. The models applied in this study estimated the number of households and the type of services these households are accessing or not accessing. A large number of households in the CD do not have toilets.

Table 5.21: Estimated Household-size by services

Munic	DERH-HSiZE	water1	lighting1	toilet1	EBLUB
Aganang	22	1	1	0	19.44655
Aganang	19	1	1	0	16.86882
Aganang	18	1	1	0	16.00958
Aganang	18	1	1	0	16.00958
Blouberg	2	1	1	0	2.261709
Blouberg	1	1	1	0	1.402467
Blouberg	5	1	1	0	4.839435
Blouberg	3	1	1	0	3.120951
Lepele-Nkum	1	1	1	0	1.402467
Lepele-Nkum	1	1	1	0	1.402467
Lepele-Nkum	13	1	1	0	11.71337
Lepele-Nkum	5	1	1	0	4.839435
Molemole	4	1	0	0	4
Molemole	2	1	1	1	2.110234
Molemole	1	1	1	0	1.402467
Molemole	4	1	1	1	3.828718
Polokwane	2	1	1	0	2.261709
Polokwane	12	1	0	0	12
Polokwane	3	1	0	0	3
Polokwane	1	1	1	0	1.402467

The intensity of the shortage of toilet facilities is also observed in crowded households where more than 10 members share one household. The number of HHs with more than 10 members without a flushing toilet in Aganang, Blouberg, Lepelle-Nkumpi and Polokwane municipalities are respectively, 22, 18, 13, and 12. It should be noted that Tables 5.21 and 5.22 show few rows and columns for each municipality. The other results are presented from Appendix A5.3. for all the five municipalities in the CD.

Table 5.22: Estimated Household-size by services (Cont.)

Munic	MSE	SE-EBLUB	CV-EBLUB
Aganang	0.879055	0.937579	0.048213
Aganang	0.879055	0.937579	0.055581
Aganang	0.879055	0.937579	0.058564
Aganang	0.879055	0.937579	0.058564
Blouberg	0.879055	0.937579	0.414545
Blouberg	0.879055	0.937579	0.668522
Blouberg	0.879055	0.937579	0.193737
Blouberg	0.879055	0.937579	0.300415
Lepele-Nkum	0.879055	0.937579	0.668522
Lepele-Nkum	0.879055	0.937579	0.668522
Lepele-Nkum	0.879055	0.937579	0.080043
Lepele-Nkum	0.879055	0.937579	0.193737
Molemole	0	0	0
Molemole	0.879055	0.937579	0.444301
Molemole	0.879055	0.937579	0.668522
Molemole	0.879055	0.937579	0.244881
Polokwane	0.879055	0.937579	0.414545
Polokwane	0	0	0
Polokwane	0	0	0
Polokwane	0.879055	0.937579	0.668522

Chapter 6

Summary, Conclusion and Recommendations

6.1 Summary of the results

The study presented the application of Small Area Estimation (SAE) methods, focusing at the Capricorn District of the Limpopo Province in South Africa. The application of SAE methods using Census 2011 and the General Household Survey (GHS) series 2009-2015 data sets assisted in exploring the delivery of basic services accessed by households in the Limpopo Province (LP) at municipality and lower levels. The SAE methods assisted this study to document the geographical disparity in the accessibility of water, sanitation and electricity for lighting (WSE) in the Capricorn District at small area (village) level. The study demonstrated the value of small area analysis in producing estimates of the delivery of WSE services at Unit level (household) and Area level (municipality) of the Capricorn District (CD).

Since the Area level model can be used to produce area estimates only if the aggregated data is available, the study used Census 2011 data which is aggregated. It should be noted that only the 10% sample of Census data in SA gets released for academic and related purposes. The weighted sum of the Direct and the Synthetic estimators was combined to produce the Composite estimator. Scholars apply Composite estimator to manage the reduction of the Direct and the Synthetic estimates, depending on the size of the sample in the small area. If the sample is large but not reliable, extra information is borrowed from other close areas, and more weight should be given to the Synthetic estimator in such a way that it minimises the Mean Square Error of the Composite estimator.

The study assumed that the sample of 299891 households in the CD was reliable and large enough. For the areas with large samples in the CD, more weight was assigned to the Direct estimator than to the Synthetic estimator in the SAS computations. The combination of several SAE models assisted in minimising the MSE obtained in this study. A trial to use extra information from administration records was attempted (borrowing strength), but generated a larger MSE of the Composite estimator due to lack of reliability in the data.

The Unit and Area level models were combined in SAS $Proc\ mixed$, to form a special case of the General Linear Mixed (GLM) model, which requires a combination of fixed and random effects. These are advantages in applying Direct, Implicit and Explicit models in SAE. The results on solution-of-fixed-effects show that the intercept value, 2.9074 is the access estimated mean for the combined WSE. The computations yield the estimated means for the variables: water1, toilet1 and electricity for lighting1 as 2.96756, 1.8313 and 3.7992 respectively. It is observed that toilet services are least accessed in the CD.

The shortage of flushing toilets is prevalent across all the local municipalities in the CD. The results show that Polokwane municipality is also affected in this regard because it is surrounded by the high proportion of households without flushing toilets.

Estimating the proportion of households accessing/not accessing water, sanitation and electricity services, complicated the Bayesian models such as $Proc\ mcmc$ which were applied in several SAS procedures to obtain the posterior statistics: mean, SD and percentiles (25th to 75th) and the posterior intervals. Since the study did not use additional information, the $Proc\ mcmc$ was performed to address the challenge of computing the posterior distributions analytically (see Equations 3.31 and 3.32).

The study determined the census undercounts and their undercounts rates, i.e., the number of households missed in the Capricorn District during the national census count in 2011. It was found that 1306 households in Aganang municipality were not counted during the Census 2011. On average, the number of missed households in Aganang is twice the number of missed households in Blouberg and Lepelle-Nkumpi municipalities. A total of 2909 households in the CD were not counted in Census 2011. At a small area level, the study found that there was a significant number of households which sourced water from *dam/pool/stagnant* water or *river/stream*.

6.2 Conclusion

The objectives stipulated in this study were carried out through the application of small area estimation methods to achieve the overall goal, i.e., the aim. The models derived in this study can be replicated to any big area to estimate the variable of interest at small area, exploring the accessibility of water, sanitation and electricity. The study established that the water, sanitation and electricity for lighting services are not accessed equitably in the CD, and therefore, the alternative hypothesis is accepted.

Although the results will benefit the governance in the CD, the review on SAE theory and its application reveals, to a minimal extent, that confidentiality may be questioned at unit level in case of a small sample (e.g. small village). It is clear that there is a need for more research in the application of SAE, exploring the variables related to the vulnerability experienced by poor people of South Africa.

It was deduced from the results that:

- The Limpopo governance prioritised electricity services compared to sanitation and water in the Capricorn District.
- On average, there are 78% of households without flushing toilets in the Limpopo Province, in general, but in most small rural areas of the Capricorn District, in particular.
- There is a water crisis that needs intervention in Polokwane and Lepelle-Nkumpi municipalities.
- Undercoverage results indicate that the likelihood of not being counted in census is high in Aganang municipality. Related future research is needed to monitor the situation.

6.3 Recommendations

The study recommends that:

1. The Limpopo Province improve the accessibility of sanitation and water services, particularly in the outskirts of the Polokwane municipality.

2. Current sanitation policies need to be reviewed.

3. New policies be developed with the inclusion of rural small settlements, considering the spatial planning and demarcation challenges in South Africa.

4. The Capricorn District should revisit the South African prescripts for the provision of water, sanitation and electricity to the indigent households. The study was bound to review the literature and the South African prescripts related to the provision of water, sanitation and electricity such as the RDP, GEAR and the Municipal Acts. Comparing the reviewed information with the results obtained, the study learned that the implementation of what is documented into the prescripts remains a challenge, at least in the Capricorn District.

5. The indigent register needs to be frequently updated for efficient planning and monitoring of the provision of basic services.

6. The Capricorn District needs to work closely with Statistics South Africa and involve other relevant data agencies and stakeholders, especially during the national census surveys.

References

ADELZADEH, A. (1996). From the RDP to GEAR: The gradual embracing of neo-liberalism in economic policy. *Transformation*, **31** (4), 66–95.

ADELZADEH, A. AND PADAYACHEE, V. (1994). The RDP white paper: reconstruction of a development vision. *Transformation*, **25** (5), 1–18.

ALEXANDER, P., RUNCIMAN, C., AND NGWANE, T. (2013). Media Briefing: Community Protests 2004-2013: Some research findings. *Social Change, University of Johannesburg, South Africa*.

BAKKER, K. (2010). *Privatizing water: governance failure and the world's urban water crisis*. Cornell University Press.

BATTESE, G. E., HARTER, R. M., AND FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association, Taylor and Francis Group*, **83** (401), 28–36.

BOTTING, M. J., PORBENI, E. O., JOFFRES, M. R., JOHNSTON, B. C., BLACK, R. E., AND MILLS, E. J. (2010). Water and sanitation infrastructure for health: The impact of foreign aid, globalization and health. *Canadian Universities*, **6** (1), 1 –12.

BRACKSTONE, G. (2003). Managing data quality in a statistical agency.

CARTER, M. R. (1987). Risk sharing and incentives in the decollectivisation of agriculture. *Oxford Economic Papers*, **39** (3), 577–595.

CARTER, M. R. AND MAY, J. (1999). Poverty, livelihood and class in rural South Africa. *World Development*, **27** (1), 1–20.

CHAUDHURI, S. AND GHOSH, M. (2011). Empirical likelihood for small area estimation. *Biometrika, JSTOR*, 473–480.

CHEN, F. (2009). *Bayesian Modelling Using the MCMC Procedure in Proceedings of the SAS Global Forum 2008 Conference*. Cary, NC: SAS Institute Inc.

CHRISTOPHER, A. J. (2001). Urban segregation in post-apartheid South Africa. *Urban Studies*, **38** (3), 449–466.

COTTER, D. A., HERMSEN, J. M., AND VANNERMAN, R. (1999). Systems of gender, race, and class: Multilevel analysis. *Social Forces*, **78** (2), 433–460.

CRAYTHORNE, D. L. (2006). *Municipal administration: Gazette No.22355*. Juta and Company Ltd.

CRESSIE, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *Journal of the American Statistical Association, Taylor and Francis Group*, **84** (408), 1033–1044.

DANIELS, R. (2006). Implementation of the right of access to sufficient water through privatization in South Africa. *Penn St. Envtl. L. Rev.*, **15**, 61.

DANZINGER, S. H. AND WEINBERG, D. H. (1987). *Fighting Poverty: What Works and What Doesn't*. Harvard University Press.

DATTA, G. S. AND GHOSH, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics, JSTOR*, 1748–1770.

DEEGAN, H. (2001). *The Politics of the New SA: Apartheid and After*. Pearson Education Ltd, England.

DEVENISH, G. E. (1998). *A Commentary on the South African Constitution, Act No. l08*. Butterworth-Heinemann.

DONEV, G., VAN SARK, W. G., WILFRIED, G. J. H. M., BLOK, K., AND DINTCHEV, O. (2012). Solar water heating potential in South Africa in dynamic energy market conditions. *Renewable and Sustainable Energy Reviews, Elsevier*, **16** (5), 3002–3013.

DRESSEL, P. L. (1988). Gender, race, and class: beyond the feminization of poverty in later life. *The Gerontologist*, **28** (2), 177–180.

EARLE, A., GOLDIN, J., AND PHEMO, K. (2005). Domestic water provision in the Democratic South Africa, changes and challenges. *Uppsala: The Nordic Africa Institute*, 1–40.

EGERO, B. (1991). South African's Bantustans: From dumping grounds to battlefronts. *Nordic Africa Institute*, **4**.

ELBERS, C., LANJOUW, J. O., AND LANJOUW, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, **71** (1), 355–364.

ELMELECH, Y. AND LU, H. H. (2004). Race, ethnicity, and the gender poverty gap. *Social Science Research*, **33** (1), 158–182.

ERICKSEN, E. P. AND KADANE, J. B. (1985). Estimating the Population in a Census year 1980 and beyond. *Journal of the American Statistical Association, Taylor and Francis*, **80** (389), 98–109.

ERICKSEN, E. P., KADANE, J. B., AND TUKEY, J. W. (1989). Adjusting the 1980 census of population and housing. *Journal of the American Statistical Association, Taylor and Francis Group*, **84** (408), 927–944.

FARLEY, R. (1984). *Blacks and Whites: Narrowing the Gap*. Harvard University Press.

FAY, R. E. AND HERRIOT, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association, Taylor and Francis*, **74** (366a), 269–277.

FEDER, M. (2001). Time series analysis of repeated surveys: The state-space approach. *Statistica Neerlandica*, **55** (2), 182–199.

FREEDMAN, D. A. AND NAVIDI, W. C. (1986). Regression models for adjusting the 1980 census. *Statistical Science, Institute of Mathematical Statistics*, **1** (1), 3–11.

FREEDMAN, D. A. AND NAVIDI, W. C. (1992). Should we have adjusted the Census of 1980? *Survey Methodology*.

GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85** (410), 398–409.

GEMAN, D. AND GEMAN, S. (1983). *Parameter estimation for some Markov Fields*. Brown University.

GHOSH, M. AND RAO, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55–76.

GOLDSTEIN, H. (2003). *Multilevel statistical models, 3rd edition*. Arnold Publishers, London.

GONZALEZ, J. F., PLACEK, P. J., AND SCOTT, C. (1996). Synthetic estimation of followback surveys at the national centre for health statistics. *Indirect Estimators in US Federal Programs*, 16–27.

GUADARRAMA, M., MOLINA, I., AND RAO, J. N. K. (2016). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition new series*, **1** (17), 41–66.

HACKER, A. (1992). *Two Nations: Black and White, Separate, Hostile and Unequal*. Charles Scribner's Sons, New York.

HOGG, D. W., BLANTON, M. R., ROWEIS, S. T., AND JOHNSTON, K. V. (2005). Geographic information in small area estimation: Small area models and spatially correlated random areas effects. *Modeling complete distributions with incomplete orbservations: The Velocity Ellipsoid from Hipparcos Data*, **629** (1), 268.

HOLT, D. T. M. F., SMITH, T. M. F., AND TOMBERLIN, T. J. (1979). A model-based approach to estimation for small subgroups of a population. *Journal of the American Statistical Association, Taylor and Francis*, **74** (366a), 405–410.

HORVITZ, D. G. AND THOMPSON, D. J. (1952). A generalisation of sampling without replacement from a finite universe. *Journal of American Statistical Association*, **47** (260), 663–685.

HULME, D. AND SHEPHERD, A. (2003). Conceptualising chronic poverty. *World Development*, **31** (3), 403–423.

JIANG, J. AND LAHIRI, P. (2006). Mix model prediction and small area estimation. *Sociedad de Estadistica e Investigacion Operativa*, **15** (1), 1–96.

KALTON, G. (1983). Models in the practice of survey sampling. *International Statistical Review, Revue Internationale de Statistique, JSTOR*, 175–188.

KENWARD, M. G. AND ROGER, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997.

KERR, C. (1998). Community Water Development, water transport and storage by val curtis, doing the donkey work. *Waterlines*, **4** (2), 194–198.

KIPRUTO, H., MUNG, J., OGILA, K., ADEM, A., MWALILI, S., KIBUCHI, E., AND SANG, G. (2015). Spatial temporal modelling of Tuberculosis in Kenya using small area estimation spatial temporal modelling of tuberculosis in kenya using small area estimation. *International Journal of Science and Research*.

LE, B. D. (2015). Towards integration at last? The sustainable development goals as a network of targets. *Sustainable Development*, **23** (3), 176–187.

LE ROUX, N. (2009). The reason behind service delivery Protests in South Africa. *Institute for Security Studies, University of Pretoria, South Africa, 29 July*.

LEHTONEN, R. AND PAHKINEN, E. (2004). *Practical Methods for design and Analysis of Complex surveys*. John Wiley and Sons Ltd. Chichester.

LEVINE, S. AND ROBERTS, B. (2013). Robust estimates of changes in poverty and inequality in post-independence Namibia. *South African Journal of Economics*, **81** (2), 167–191.

LITTLE, R. J. (2006). Calibrated Bayes, a Bayes or frequentist roadmap. *The American Statistician*, **60** (3), 213–223.

LODGE, T. (1983). *Black Politics in South Africa since 1945*. Longman Publishing Group, London.

LODGE, T. (2003). *Politics in South Africa: From Mandela to Mbeki, The Reconstruction and Development Programme Delivery and Performance*. Indiana University Press.

MAGASELA, W. (2006a). Towards a constitution-based definition of poverty in post-apartheid South Africa, State of the Nation. *HSRC Press*.

MAGASELA, W. (2006b). Towards a constitution-based definition of poverty in post-apartheid South Africa,state of the nation. *HSRC Press*.

MAGUBANE, B. M. (1996). The making of the racist state: British imperialism and the union of South Africa. *Africa World Press*, 1875–1910.

MAJAVU, M. (2011). Making sense of municipal revolt. *Development in Focus*, **1** (2), 4–6.

MALA, K., SCHLÄPFER, A., AND PRYOR, T. (2009). Better or worse? the role of solar photovoltaic PV systems in sustainable development: Case studies of remote atoll communities in kiribati. *Renewable Energy, Elsevier*, **34** (2), 358–361.

MALZBENDER, D., GOLDIN, J., TURTON, A., AND EARLE, A. (2005). Traditional water governance and South Africa's National Water Act' tension or cooperation. *African water laws: plural legislative frameworks for rural water management in Africa*, 26–28.

MARCHETTI, S., GIUSTI, C., PRATESI, M., SALVATI, N., GIANNOTTI, F., PEDRESCHI, D., RINZIVILLO, S., PAPPALARDO, L., AND GABRIELLI, L. (2015). Small area model-based estimators using big data sources. *Journal of Official Statistics*, **31** (2), 263–281.

MARTINOT, E., CABRAAL, A., AND MATHUR, S. (2001). World Bank/GEF solar home system projects: experiences and lessons learned 1993-2000. *Renewable and Sustainable Energy Reviews, Elsevier*, **5** (1), 39–57.

MATTES, R., BRATTON, M., AND DAVIDS, Y. D. (2003). Poverty survival, and democracy in Southern Africa. *Afrobarometer Working Paper No. 23. Cape Town: IDASA*.

MCCULLOCH, C. E. AND NEUHAUS, J. M. (2001). *Generalized linear mixed models*. Wiley and Sons Online Library.

MERCER, L. D., WAKEFIELD, J., PANTAZIS, A., LUTAMBI, A. M., MASANJA, H., AND CLARK, S. (2015). Space-time smoothing of complex survey data:

small area estimation for child mortality. *The Annals of Applied Statistics, NIH Public Access*, **9** (4), 1889.

MINOT, N. AND BAULCH, B. (2005). Spatial patterns of poverty in Vietnam and their implications for policy. *Food Policy*, **30** (5-6), 461–475.

MOKOKELI, S. (2012). Jobless Youths a 'ticking' time bomb for South Africa. *Archives, 6 August, Gauteng*.

MUKHOPADHYAY, P. K. AND MCDOWELL, A. (2011). *Small Area Estimation for Survey Data Analysis using SAS Software*. SAS Institute Inc., USA.

NATTRASS, N. AND SEEKINGS, J. (2001). 'Two Nations'? race and Economic Inequality in South Africa today, why South Africa matters. *American Academy of Arts and Sciences*, **130** (1), 45–70.

NEWS24 (2014). English to blame for residence 'booing' Zuma. *City Press Newspaper, South Africa*.

NOBLE, M., BARNES, H., WRIGHT, G., AND ROBERTS, B. (2009). Small area indices of multiple deprivations in South Africa. *Social Indicators Research*.

NOBLE, M. AND WRIGHT, G. (2012). Using indicators of multiple deprivation to demonstrate the Spatial legacy of Apartheid in South Africa. *Springer and Science Media*.

NOBLE, M. AND WRIGHT, G. (2013). Ring of Fire: Socially perceived necessities in informal urban settlement in South Africa. *Policy and Politics*, **41** (2), 259–277.

NOBLE, M., WRIGHT, G., SMITH, G., AND DIBBEN, C. (2006). Measuring multiple deprivations at the small-area level. *Environment and Planning*, **38** (1), 169–185.

NTSALA, M. AND MAHLATJI, M. (2016). Service delivery protests resulting in the burning of libraries: a study of selected public libraries in South Africa.

PETRUCCI, A., PRATESI, M., AND SALVATI, N. (2005). Geographic information in small area estimation: small area models and spatially correlated random areas effects. *Statistics in Transition*, **7** (3).

PETRUCCI, A. AND SALVATI, N. (2004). *Small Area Estimation considering spatially correlated errors: the unit level random effects models*. University of Florence.

PFEFFERMANN, D. (2002). Small area estimation-new developments and directions. *International Statistical Review, Wiley Online Library*, **70** (1), 125–143.

PFEFFERMANN, D. (2013). New important developments in small area estimation. *Statistical Science*, **28** (1), 40–68.

PFEFFERMANN, D. AND BARNARD, C. H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, **9** (1), 73–84.

PICKETT, W., KOUSHIK, A., FAELKER, T., AND BROWN, K. S. (2000). Estimation of youth smoking behaviours in Canada. *Chronic Diseases and injuries in Canada*, **21** (3), 119.

PIERSON., C. AND FRANCIS, G. (2006). *The Welfare State Reader*. Polity Press, US.

PILLAY, U., TOMLINSON, R., AND TOIT, J. D. (2006). Democracy and delivery: Urban policy in South Africa. *HSRC press*.

PLAUT, M. AND SEEKINGS, J. (2002). Reflection on the South African revolution: A discussion between Martin Plaut and Jeremy Seekings. *In History Workshop Journal: Oxford University Press*, **54**, 209–220.

PRATESI, M. AND SALVATI, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical Methods and Applications*, **17** (1), 113–141.

PRATESI, M. AND SALVATI, N. (2009). Small area estimation in the presence of correlated random area effects. *Journal of Official Statistics*, **25** (1), 37.

RAGHUNATHAN, T. E., XIE, D., SCHENKER, N., PARSONS, V., DAVIS, W. W., DODD, K., AND FEUER, E. J. (2007). Combining information from multiple surveys for small area estimation: a Bayesian approach. *Journal of the American Statistical Association*, **102**, 474–486.

RAHMAN, A. (2008). A review of small area estimation problems and methodological developments. *University of Canberra, Australia*.

RAO, J. N. K. (2003). *Small Area Estimation*. New Jersey, John Wiley and Sons, Inc.

RAO, J. N. K. AND MOLINA, I. (2015). *Small area estimation*. New Jersey, John Wiley and Sons, Inc.

ROMA, E., BUCKLEY, C., JEFFERSON, B., AND JEFFREY, P. (2010). Assessing users' experience of shared sanitation facilities: A case study of community ablution blocks in Durban. *South Africa, Water*, **36** (5), 589–594.

ROUX, A., ABEDIAN, I., KHAN, I., BRIAN, A. D., SMIT, B., SMAL, D., HIRSCH, A., VAN DER MERWE, S., GOLDING, I., GELBA, S., VAN SEVENTER, D., VAN DER BERG, S., DA SILVA, L. P., AND KETLEY, R. (1996). Growth, employment and redistribution (GEAR): A macroeconomic strategy. *Department of Finance, Pretoria*.

RUNCIMAN, C. (2014). South African protests. *Social Change Research Unit, University of Johannesburg, South Africa*.

SAEI, A. AND CHAMBERS, R. (2003). Small area estimation: a review of methods based on the application of Mixed models. *Southampton Statistical Sciences Research Institute*.

SARNDAL, C. AND HIDIROGLOU, M. A. (1989). Small domain estimation: a conditional analysis. *Journal of the American Statistical Association, Taylor and Francis Group*, **84** (405), 266–275.

SARNDAL, C. E., SWENSSON, B., AND WRETMAN, J. (1992). *Model assisted survey sampling*. Springer-Verlag Inc, New York.

SCHOUTEN, M. AND MATHENGE, R. W. (2010). Communal sanitation alternatives for slums: A case study of Kibera, Kenya. *Physics and Chemistry of the Earth, Parts A, B and C*, **35** (13-14), 815–822.

SEEKINGS, J. (2007a). Poverty and Inequality after apartheid. *Paper prepared for the second 'After Apartheid Conference' 'Yale' in April*, 27–28.

SMITH, K. S., NOGLE, J., AND CODY, S. (2002). A Regression approach to estimating the average number of persons per household. *Demography*, **39** (4), 697–712.

SURENDER, R. (2014). The drivers of universal health care in South Africa: The role of ideas, institutions and actors. *UNRISD Working Paper*.

VISSER, W. (2004). Shifting RDP to GEAR: The ANC government dilemma in providing equitable system of social security for the new South Africa. *Paper Presented at the 40th Linzer Konferenz*.

WHITWORTH, A. (????). Evaluations and improvements in small area estimation methodologies, discussion paper. *NCRM*.

YOU, Y., RAO, J. N. K., AND GAMBINO, J. A. C. K. (2003). Model-based unemployment rate estimation for the Canadian Labour Force Survey: a hierarchical Bayes approach. *Survey Methodology*, **29** (1), 25–32.

ZUMA, J. (2014). State of the Nation Address of the President of South Africa. *13 February, Cape Town, McIntyre*.

Appendices

Appendix A3.1: Water question for Census 2001 and the choice options (H-26 PIPED WATER) Stats SA: Administered for all households including the institutions

In which way does this household obtain PIPED WATER for domestic use?

- No access to piped (tap) water
- Piped (tap) water on community stand:distance greater than 200m from dwelling
- Piped (tap) water on community stand:distance less than 200m from dwelling
- Piped (tap) water inside the yard
- Piped (tap) water inside the dwelling
- ☐ Not applicable (homeless)

Appendix A3.2: Water question for Census 2011 and the choice options (H-26 Piped Water)

Stats SA: Administered for all households including the institutions

In which way does this household mainly get piped water for household use?

- 1= Piped (tap) water inside dwelling/institution
- 2=Piped (tap) water inside yard
- 3 = Piped (tap) water on community stand:distance less than 200m from dwelling/institution
- 4 = Piped (tap) water on community stand:distance between 200m and 500m from dwelling/institution
- 5 = Piped (tap) water on community stand:distance between 500m and 1000m (1km) from dwelling/ institution
- 6 = Piped (tap) water on community stand:distance greater than 1000m (1km) from dwelling/institution
- 7 = No access to piped (tap) water
- 9 = Unspecified

APPENDIX A3.3: WATER QUESTION FOR GHS 2011 AND THE CHOICE OPTIONS

STATS SA: ADMINISTERED FOR ALL HOUSEHOLDS INCLUDING THE INSTITUTIONS

What is the household's main source of drinking water

- 01 = Piped (tap) water in dwelling
- 02 = Piped (tap) water on site or in yard
- 03 = Borehole on site
- 04 = Rainwater tank on site
- 05 = Neighbour's tap
- 06 = Public tap
- 07 = Water-carrier/tanker
- 08 = Borehole off site/communal
- 09 = Flowing water/stream/river
- 10 = Dam/pool/stagnant water 11 = Well
- 12 = Spring
- 13 = Other
- 99 = Unspecified

Appendix A3.4: Sanitation question for Census 2001 and the choice options

(H-27 TOILET FACILITY)

Stats SA: Administered for all households including the institutions

What is the MAIN type of TOILET facility used by this household?

- 1 = Flush toilet (connected to sewerage system)
- 2 = Flush toilet (with septic tank)
- 3 = Chemical toilet
- 4 = Pit toilet with ventilation (VIP)
- 5 = Pit toilet without ventilation
- 6 = Bucket laterine
- 7 = Other
- 0 = None
- ☐ Not applicable (homeless)

APPENDIX A3.5: SANITATION QUESTION FOR CENSUS 2011 AND THE CHOICE OPTIONS

(H-27 TOILET FACILITY)

STATS SA: ADMINISTERED FOR ALL HOUSEHOLDS INCLUDING THE INSTITUTIONS

What type of TOILET used is by this household?

- 1= Flush toilet connected to a public sewerage system
- 2 = Flush toilet connected to a septic tank
- 3 = Chemical toilet
- 4 = Pit latrine/toilet with ventilation pipe
- 5 = Pit latrine/toilet without ventilation pipe
- 6 = Bucket toilet
- 7 = None
- 8 = Other (specify)
- 9 = Unspecified

Appendix A3.6: Sanitation question for GHS 2011 and the choice options (H-27 TOILET FACILITY)

Stats SA: Administered for all households including the institutions

What type of TOILET used is by this household?

- 1= Flush toilet connected to a public sewerage system
- 2 = Flush toilet connected to a septic tank
- 3 = Chemical toilet
- 4 = Pit latrine/toilet with ventilation pipe
- 5 = Pit latrine/toilet without ventilation pipe
- 6 = Bucket toilet
- 7 = None
- 8 = Other (specify)
- 9 = Unspecified

Appendix A3.7: Energy/fuel question for Census 2001 and the choice options

(H-28 ENERGY/FUEL(COOKING)

Stats SA: Administered for all households including the institutions

What type of energy/fuel does this household MAINLY use for cooking/ heating /lighting?

- 1 = Electricity
- 2 = Gas
- 3 = Paraffin
- 4 = Wood
- 5 = Coal
- 7 = Animal dung
- 8 = Solar
- 9 = Other
- ☐ Not applicable (homeless)
- 99 = Unspecified

Appendix A3.8: Energy/fuel question for Census 2011 and the choice options

(H-28 ENERGY/FUEL(COOKING)

Stats SA: Administered for all households including the institutions

What type of energy/fuel does this household MAINLY use for cooking/ heating /lighting?

- 1 = Electricity
- 2 = Gas
- 3 = Paraffin
- 4 = Wood
- 5 = Coal
- 7 = Animal dung
- 8 = Solar
- 9 = Other
- ☐ Not applicable (homeless)
- 99 = Unspecified

Appendix A3.9: Energy/fuel question for GHS 2011 and the choice options

(H-28 ENERGY/FUEL(COOKING)

Stats SA: Administered for all households including the institutions

What type of energy/fuel does this household MAINLY use for cooking/ heating /lighting?

- 01 = Electricity from mains
- 02 = Electricity from generator
- 03 = Gas
- 04 = Paraffin
- 05 = Wood
- 06 = Coal
- 07 = Candles
- 08 = Animal dung
- 09 = Solar energy
- 10 = Other (specify)
- 11 = None

Appendix A3.10: Master sample design and estimation, 2007

Provincial Code Numbers and Names

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

Note: Stats SA: Master sample design 2007: Pr-code and Pr.Name denotes province code and province name .

APPENDIX A3.11: MASTER SAMPLE DESIGN AND ESTIMATION, 2007

STATS SA: DESIGN VARIABLES

Description of Stratum Number

- A 6-digit number representing stratum formed during Master Sample 2006 where during Master Sample 2006 where (digit 1 = Province based 2005 provincial boundaries, digit 2 and 3= metro/non-metro, digit 4=Geography type)

APPENDIX A3.12: MASTER SAMPLE DESIGN AND ESTIMATION, 2007

STATS SA: DESIGN VARIABLES

Description of Primary Sampling Unit Number itemised (PSUno)

- An 8-digit number derived from the EAs used to form PSU (digit 1 = Province, digit 2 and 3 = municipality code, digit 4 and 8= unique PSU with the municipality). This is the PSU number that was used for selecting the sample of PSUs. During sample selection, some PSUs are segmented where the segment number is a 3-digit number; concatenated to the PSUno resulting in a variable called ‘PSUno-Seg’ which is 11-digit string. PSUs that are not segmented, will have the last three digits of the PSUno-Seg equal to ‘000’; where those with segmentation will have a 3-digit number greater that ‘001’

Appendix A3.13: Master sample design and estimation, 2007

Stats SA: Design variables

Description of Dwelling number

- A unique 5-digit number assigned to dwelling units during listing

Appendix A3.14: Master sample design and estimation, 2007

Stats SA: Design variables

Description of Rotation

- The rotation number (1, 2, 3 or 4) was assigned at the PSU level, and indicates the quarter the sampled dwelling units will be rotated out of the sample and replaced by new sample of dwelling units from the same PSU (or the next PSU on the list when the originally sampled PSU has been exhausted)

Appendix A3.15: Master sample design and estimation, 2007

Stats SA: Design variables

Description of Unique number (UQNR)

- Unique household identifier that is defined by concatenation of PSU number Segment, Dwelling number and household number

Appendix A3.16: Master sample design and estimation, 2007

Stats SA: Design variables

Description of Sample Weights

- Calibrated adjusted base weight for unique households Adjustments are made to the base weight to account for Primary Sampling Units (PSUs) that were sub-sampled due to growth or those that were segmented (informal PSUs), non-coverage of very small Census Enumeration Areas (EAs) that were excluded at the design phase, and unit non-response

Appendix A4.1: Access to tap drinking water by province, 2009-2015

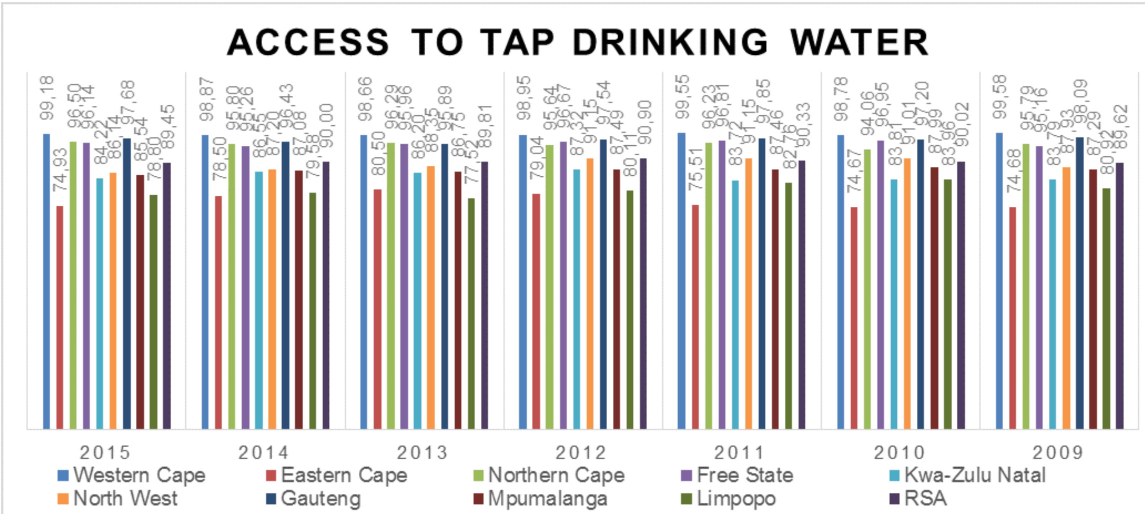

Figure 6.1: Proportion of HHs accessing tap drinking water by province, GHS

Appendix A4.2: No access to HH services by province, GHS 2009-2015

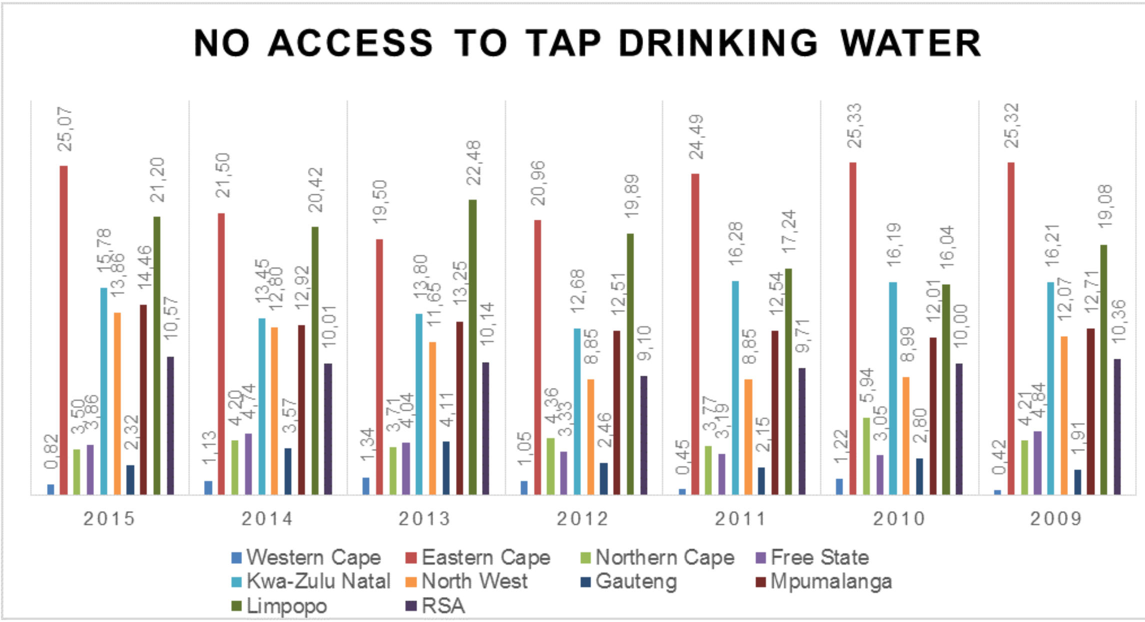

Figure 6.2: Proportion of HHs with No tap drinking water by province, GHS

Appendix A4.3: Access to Services by municipality in the CD, Census 2011

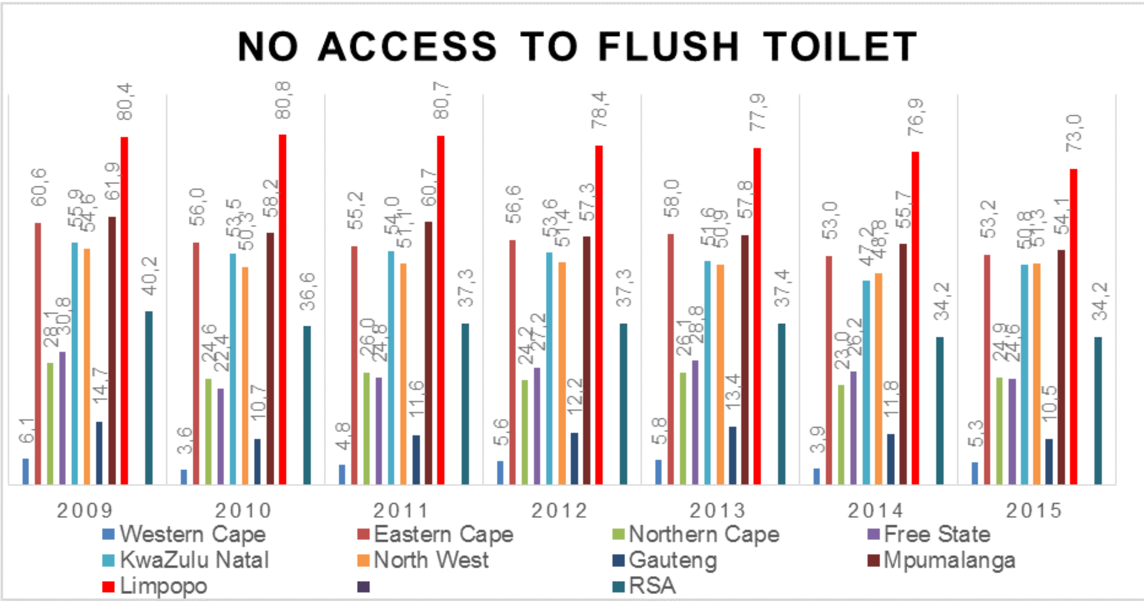

Figure 6.3: Proportion of HHs with No flushing Toilet by province, GHS

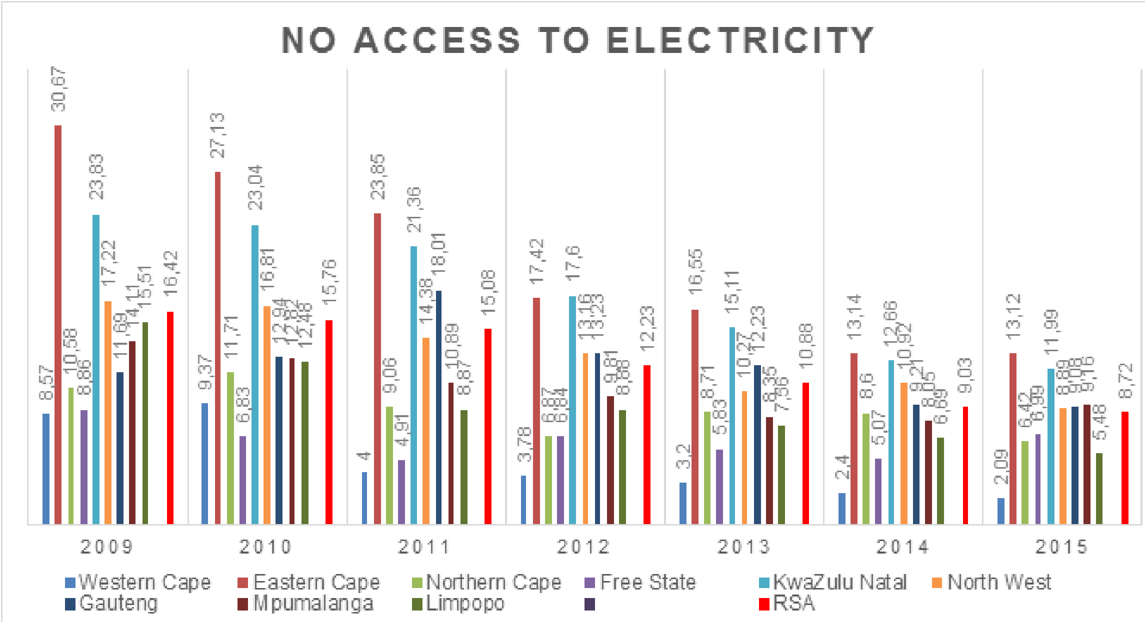

Figure 6.4: Proportion of HHs with No electricity for lighting by province, GHS

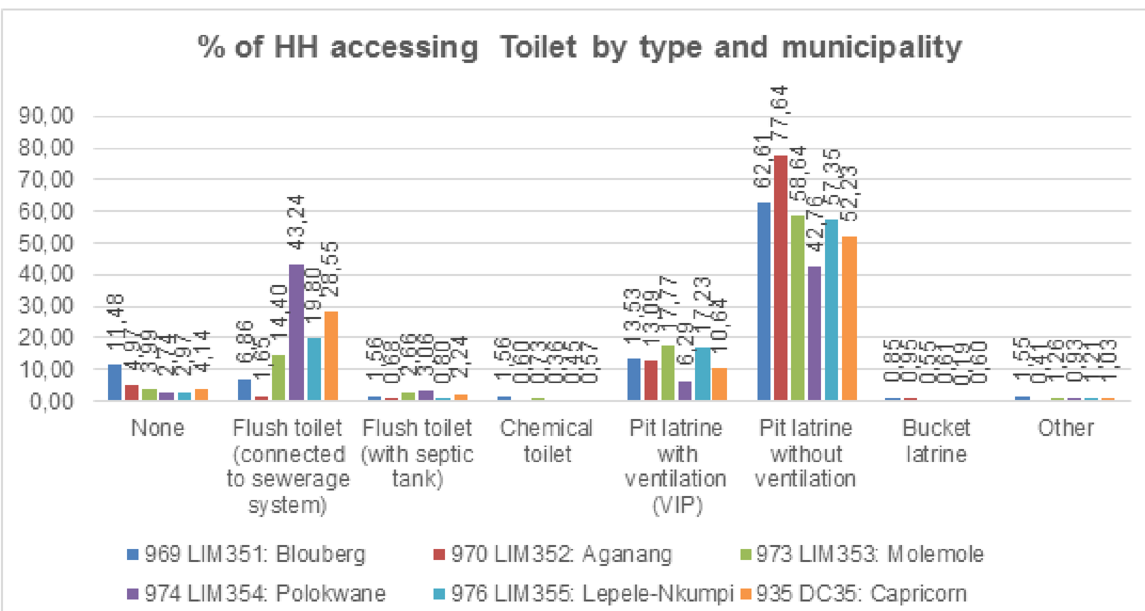

Figure 6.5: Proportion of HHs accessing toilet by type and municipality in the CD, Census 2011

Appendix A4.4: Access to type of water source in the CD, Census 2011

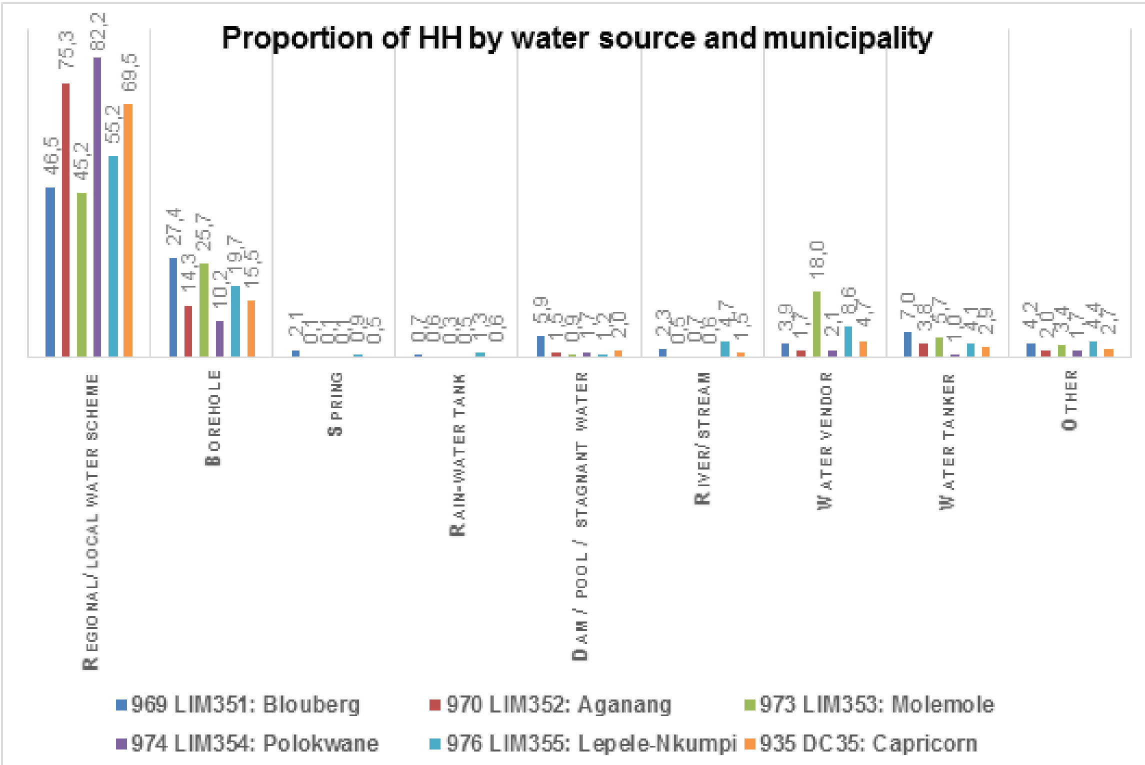

Figure 6.6: Proportion of HHs by type of water source and municipality, Census 2011

Table 6.1: Source of water by small area in the CD, Census 2011

Villages	Water	Borehole	Spring	Rain	Stagnant	R/S
935 DC35: Capricorn	240989	53864	1612	2245	6956	5101
969 LIM351: Blouberg	19164	11273	868	301	2417	940
969003 GaRamaswikana	13	10	3	-	222	50
969011 GaMamadi	777	107	3	3	211	-
969041 Buffelshoek	488	185	8	10	130	178
969046 Borwalathoto	74	44	-	1	185	1
969053 GaRamutla	-	11	-	-	264	1
969069 Ga-Kobe	112	188	40	-	132	-
969071 Mophamamona	63	110	-	4	114	-
970 LIM352: Aganang	24443	4649	33	209	501	158
970038 Ga-Modikana	143	69	1	2	131	-
973 LIM353: Molemole	13490	7663	30	79	257	203
973014 Tshitale	10	3	-	3	-	139
973030 Ramokgopa	1703	1009	2	15	128	29
974 LIM354: Polokwane	151493	18730	165	898	3072	1014
974008 Ditenteng	4	-	-	-	246	-
974024 Koloti	1054	477	1	6	270	-
974025 Mabokelele	644	150	1	-	123	98
974028 Sebayeng	2485	176	2	5	528	6
974032 Mehlakong	167	4	2	3	111	23
974042 Bloodriver	1644	527	1	9	141	2
974048 GaMamadila	171	-	-	1	129	-
974053 Tshebeng	748	67	-	2	121	-
974086 Makanye	2068	10	2	9	194	-
974092 Tholongwe	872	5	-	19	135	129
974093 Megoring	2039	5	3	7	149	103
974115 Mankgaile	305	2	2	4	124	3
974120 GaRamphere	344	1	1	22	126	63

Note: Small areas in the CD; R/S denotes river/stream.

Table 6.2: Source of water by small area in the CD, Census 2011 (cont.)

Villages	Water	Borehole	Spring	Rain	Stagnant	R/S
976 LIM355: Lepele-Nkum	32399	11550	516	759	708	2786
976009 Ga-Mafefe	5	92	3	4	2	208
976010 Ramonwana	2	-	-	-	-	123
976014 Mmashadi	-	29	-	-	-	250
976015 Madikeleng	14	9	1	2	-	216
976037 Mphahlele	1058	1266	7	64	125	24
976038 Maejane	245	148	1	-	3	350
976039 Mashite	697	44	1	5	3	234
976065 Madisha-Ditoro	2	675	2	40	2	1
976067 Nkotokwane	2	-	-	-	1	123
976069 Tjiane	1	24	-	3	93	-
976073 Lenting	136	145	1	-	-	-
976075 Magatle	37	900	3	146	5	48
976076 Marulaneng	137	91	4	2	1	-
976078 Mokgophong	236	135	2	7	4	269
976080 GaMolapo	68	170	2	116	21	3
976081 Byldrift	166	16	2	14	6	181
976083 Malatane	138	11	105	17	1	112
976084 Khureng	962	27	-	5	4	4
976085 Seruleng	-	6	-	1	-	-

Note: Small areas in the CD, R/S denotes river/stream.

Table 6.3: Source of water by small area in the CD, Census 2011 (cont.)

Villages	Vendor	Tanker	Other	Unspecified	Not appl.
935 DC35: Capricorn	16431	10124	9208	-	-
969 LIM351: Blouberg	1595	2879	1750	-	-
969003 GaRamaswikana	-	-	1	-	-
969011 GaMamadi	4	4	97	-	-
969041 Buffelshoek	15	15	32	-	-
969046 Borwalathoto	-	1	2	-	-
969053 GaRamutla	-	4	3	-	-
969069 Ga-Kobe	1	2	6	-	-
969071 Mophamamona	5	-	8	-	-
970 LIM352: Aganang	560	1244	655	-	-
970038 Ga-Modikana	18	9	3	-	-
973 LIM353: Molemole	5377	1699	1025	-	-
973014 Tshitale	9	1	9	-	-
973030 Ramokgopa	572	192	166	-	-
974 LIM354: Polokwane	3869	1906	3176	-	-
974008 Ditenteng	3	1	4	-	-
974024 Koloti	506	24	101	-	-
974025 Mabokelele	301	17	96	-	-
974028 Sebayeng	11	84	138	-	-
974032 Mehlakong	51	-	1	-	-
974042 Bloodriver	142	118	216	-	-
974048 GaMamadila	2	-	4	-	-
974053 Tshebeng	42	2	15	-	-
974086 Makanye	4	10	18	-	-
974092 Tholongwe	57	8	17	-	-
974093 Megoring	12	18	38	-	-
974115 Mankgaile	-	2	1	-	-
974120 GaRamphere	8	31	9	-	-

Note: Small areas in the CD.

Table 6.4: Source of water by small village area in the CD, Census 2011 (cont.)

Villages	Vendor	Tanker	Other	Unspecified	Not appl.
976 LIM355: Lepele-Nkum	5031	2396	2601	-	-
976009 Ga-Mafefe	30	452	9	-	-
976010 Ramonwana	-	2	-	-	-
976014 Mmashadi	-	2	-	-	-
976015 Madikeleng	17	-	-	-	-
976037 Mphahlele	78	90	249	-	-
976038 Maejane	1	-	10	-	-
976039 Mashite	56	117	41	-	-
976065 Madisha-Ditoro	183	45	10	-	-
976067 Nkotokwane	1	2	1	-	-
976069 Tjiane	196	-	1	-	-
976073 Lenting	250	17	43	-	-
976075 Magatle	273	65	248	-	-
976076 Marulaneng	274	3	17	-	-
976078 Mokgophong	132	100	8	-	-
976080 GaMolapo	627	7	127	-	-
976081 Byldrift	72	1	2	-	-
976083 Malatane	14	1	2	-	-
976084 Khureng	21	3	1	-	-
976085 Seruleng	304	-	1	-	-

Note: Small areas in the CD.

APPENDIX A5.1: PROCEDURE ON MARKOV CHAIN MONTE CARLO METHODS

Table 6.5: Procedure on Markov Chain Monte Carlo methods

Block	Parameter	Sampling	value	Prior
1	β_0	N-Metropolis	1.0000	general(0)
-	β_1	-	1.0000	general(0)
2	s2	Conjugate	0.0476	igamma(shape=0.05, scale=0.05)

Note: Sampling methods, number of subjects,initial value and Prior Distributions.

APPENDIX A5.2: PROCEDURE ON MARKOV CHAIN MONTE CARLO METHODS CONT

textbfPar	**Sampling**	**Sub**	**No.Sub**	**Sub.V**	**PriorD**
gamma	N-Metropolis	Munic	5	Aganang	-
-	-	-	-	Blouberg	-
-	-	-	-	Lepele-Nkumpi	-
-	-	-	-	Molemole	-
-	-	-	-	Polokwane	$N(\beta_0 + \beta_1 ** \log(cen), s2)$

Note: Sampling methods, subjects (sub), number of subjects (No.Sub), subject values (Sub.V) and Prior Distributions (PriorD).

Appendix A5.3: Aganang estimated Household - size by services

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
22	1	1	0	19.44	0.879	0.937	0.048
19	1	1	0	16.86	0.879	0.937	0.055
18	1	1	0	16.00	0.879	0.937	0.058
18	1	1	0	16.00	0.879	0.937	0.058
18	1	1	0	16.00	0.879	0.937	0.058
18	1	1	0	16.00	0.879	0.937	0.058
17	1	1	0	15.15	0.879	0.937	0.061
17	1	1	0	15.15	0.879	0.937	0.061
17	1	1	0	15.15	0.879	0.937	0.061
17	1	1	0	15.15	0.879	0.937	0.061
17	1	1	0	15.15	0.879	0.937	0.061
16	1	1	0	14.29	0.879	0.937	0.065
16	1	1	0	14.29	0.879	0.937	0.065
16	1	1	0	14.29	0.879	0.937	0.065
16	1	1	0	14.29	0.879	0.937	0.065
16	0	1	0	14.28	0.879	0.937	0.065
16	0	0	0	16	0	0	0
16	1	1	0	14.29	0.879	0.937	0.065
16	1	1	0	14.29	0.879	0.937	0.065
16	1	1	0	14.29	0.879	0.937	0.065

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

Appendix A5.4: Aganang estimated Household - size by services Cont.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
15	1	1	0	13.43	0.879	0.937	0.069
14	0	1	0	12.56	0.879	0.937	0.074
14	1	1	0	12.57	0.879	0.937	0.074
14	1	1	0	12.57	0.879	0.937	0.074
14	1	1	0	12.57	0.879	0.937	0.074
14	1	1	0	12.57	0.879	0.937	0.074

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

APPENDIX A5.5: BLOUBERG ESTIMATED HOUSEHOLD - SIZE BY SERVICES.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
2	1	1	0	2.261	0.879	0.937	0.414
1	1	1	0	1.402	0.879	0.937	0.668
5	1	1	0	4.839	0.879	0.937	0.193
3	1	1	0	3.120	0.879	0.937	0.300
1	1	1	0	1.402	0.879	0.937	0.668
5	1	1	0	4.839	0.879	0.937	0.193
3	1	1	0	3.120	0.879	0.937	0.300
7	1	1	0	6.557	0.879	0.937	0.142
1	1	1	0	1.402	0.879	0.937	0.668
3	1	1	0	3.120	0.879	0.937	0.300
3	1	1	0	3.120	0.879	0.937	0.300
1	1	1	0	1.402	0.879	0.937	0.668
1	1	1	0	1.402	0.879	0.937	0.668
1	1	1	0	1.402	0.879	0.937	0.668
5	1	1	0	4.839	0.879	0.937	0.193
1	1	1	0	1.402	0.879	0.937	0.668
2	1	1	0	2.261	0.879	0.937	0.414
7	1	1	0	6.557	0.879	0.937	0.142
8	1	1	0	7.417	0.879	0.937	0.126
5	1	1	0	4.839	0.879	0.937	0.193

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

APPENDIX A5.6: BLOUBERG ESTIMATED HOUSEHOLD - SIZE BY SERVICES CONT.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
6	1	1	0	5.698	0.879	0.937	0.164
1	1	1	0	1.402	0.879	0.937	0.668
5	1	1	0	4.839	0.879	0.937	0.193
4	1	1	0	3.980	0.879	0.937	0.235
2	1	1	0	2.261	0.879	0.937	0.414
4	1	1	0	3.980	0.879	0.937	0.235
4	1	1	0	3.980	0.879	0.937	0.235
2	1	1	0	2.261	0.879	0.937	0.414
8	1	1	0	7.417	0.879	0.937	0.126
2	1	1	0	2.261	0.879	0.937	0.414
6	1	1	0	5.698	0.879	0.937	0.164
1	1	1	0	1.402	0.879	0.937	0.668
5	1	1	0	4.839	0.879	0.937	0.193
4	1	1	0	3.980	0.879	0.937	0.235
6	1	1	0	5.698	0.879	0.937	0.164
6	1	1	0	5.698	0.879	0.937	0.164
2	1	1	0	2.261	0.879	0.937	0.414
2	1	1	0	2.261	0.879	0.937	0.414
2	1	1	0	2.261	0.879	0.937	0.414

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

Appendix A5.7: Lepelle-Nkumpi estimated Household - size by services

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
1	1	1	0	1.402	0.879	0.937	0.668
1	1	1	0	1.402	0.879	0.937	0.668
13	1	1	0	11.71	0.879	0.937	0.080
5	1	1	0	4.839	0.879	0.937	0.193
6	1	1	0	5.698	0.879	0.937	0.164
2	1	0	0	2	0	0	0
4	1	1	0	3.980	0.879	0.937	0.235
1	1	1	0	1.402	0.879	0.937	0.668
6	1	1	0	5.698	0.879	0.937	0.164
7	1	1	0	6.557	0.879	0.937	0.142
10	1	1	0	9.135	0.879	0.937	0.102
3	1	1	0	3.120	0.879	0.937	0.300
6	1	0	0	6	0	0	0
1	1	0	0	1	0	0	0
2	1	1	0	2.261	0.879	0.937	0.414
4	1	1	0	3.980	0.879	0.937	0.235
3	1	1	0	3.120	0.879	0.937	0.300
2	1	1	0	2.261	0.879	0.937	0.414
6	1	1	0	5.698	0.879	0.937	0.164

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

Appendix A5.8: Lepelle-Nkumpi estimated Household - size by services Cont.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
2	1	1	0	2.261	0.879	0.937	0.414
5	1	1	0	4.839	0.879	0.937	0.193
5	1	1	0	4.839	0.879	0.937	0.193
9	1	1	0	8.276	0.879	0.937	0.113
2	1	1	0	2.261	0.879	0.937	0.414
6	1	1	0	5.698	0.879	0.937	0.164
5	1	1	0	4.839	0.879	0.937	0.193
3	1	1	0	3.120	0.879	0.937	0.300
4	1	1	0	3.980	0.879	0.937	0.235
8	1	1	0	7.417	0.879	0.937	0.126
1	1	1	0	1.402	0.879	0.937	0.668
2	1	0	0	2	0	0	0
5	1	1	0	4.839	0.879	0.937	0.193
1	1	1	0	1.402	0.879	0.937	0.668
3	1	1	0	3.120	0.879	0.937	0.300
1	1	1	0	1.402	0.879	0.937	0.668
8	1	1	0	7.417	0.879	0.937	0.126
1	1	1	0	1.402	0.879	0.937	0.668
4	1	1	0	3.980	0.879	0.937	0.235
3	1	1	0	3.120	0.879	0.937	0.300

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

APPENDIX A5.9: MOLEMOLE ESTIMATED HOUSEHOLD - SIZE BY SERVICES

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
4	1	0	0	4	0	0	0
2	1	1	1	2.110	0.879	0.937	0.444
1	1	1	0	1.402	0.879	0.937	0.668
4	1	1	1	3.828	0.879	0.937	0.244
2	1	1	1	2.110	0.879	0.937	0.444
2	1	1	1	2.110	0.879	0.937	0.444
1	1	0	0	1	0	0	0
1	1	0	0	1	0	0	0
2	1	0	0	2	0	0	0
5	1	1	1	4.687	0.879	0.937	0.199
1	1	1	1	1.250	0.879	0.937	0.749
3	1	1	1	2.969	0.879	0.937	0.315
1	1	1	0	1.402	0.879	0.937	0.668
2	1	1	1	2.110	0.879	0.937	0.444
1	1	1	1	1.250	0.879	0.937	0.749
1	1	1	1	1.250	0.879	0.937	0.749
2	1	1	1	2.110	0.879	0.937	0.444
2	1	1	1	2.110	0.879	0.937	0.444
1	1	1	1	1.250	0.879	0.937	0.749

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1.

Appendix A5.10: Molemole estimated Household - size by services Cont.

Table 6.6: Molemole estimated Household - size by services Cont.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
2	1	1	1	2.110	0.879	0.937	0.444
1	1	1	1	1.250	0.879	0.937	0.749
1	1	1	0	1.402	0.879	0.937	0.668
1	1	1	1	1.250	0.879	0.937	0.749
1	1	1	1	1.250	0.879	0.937	0.749
1	1	1	1	1.250	0.879	0.937	0.749
2	1	1	0	2.261	0.879	0.937	0.414
2	1	1	1	2.110	0.879	0.937	0.444
2	1	1	1	2.110	0.879	0.937	0.444
1	1	1	1	1.250	0.879	0.937	0.749
1	1	1	1	1.250	0.879	0.937	0.749
2	1	0	0	2	0	0	0
3	1	1	1	2.969	0.879	0.937	0.315
1	1	1	1	1.250	0.879	0.937	0.749
2	1	1	1	2.110	0.879	0.937	0.444
7	1	1	1	6.406	0.879	0.937	0.146
1	1	0	0	1	0	0	0
3	1	1	0	3.120	0.879	0.937	0.300
1	1	0	0	1	0	0	0
2	1	1	1	2.110 0	.879	0.937	0.444

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1 .

APPENDIX A5.11: POLOKWANE ESTIMATED HOUSEHOLD - SIZE BY SERVICES

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
2	1	1	0	2.261	0.879	0.9375	0.414
12	1	0	0	12	0	0	0
3	1	0	0	3	0	0	0
1	1	1	0	1.402	0.879	0.937	0.668
6	1	1	0	5.698	0.879	0.937	0.164
6	1	1	0	5.698	0.879	0.937	0.164
6	1	1	0	5.698	0.879	0.937	0.164
4	1	1	0	3.980	0.879	0.937	0.235
2	1	1	0	2.261	0.879	0.937	0.414
5	1	1	0	4.839	0.879	0.937	0.193
4	1	1	0	3.980	0.879	0.937	0.235
4	1	1	0	3.980	0.879	0.937	0.235
1	1	1	0	1.4027	0.879	0.937	0.668
2	1	0	0	2	0	0	0
3	1	1	0	3.120	0.879	0.937	0.300
9	1	1	0	8.276	0.879	0.9379	0.113

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1 .

APPENDIX A5.12: POLOKWANE ESTIMATED HOUSEHOLD - SIZE BY SERVICES CONT.

HSiZE	W	L	T	EBLUP	MSE	SE-EBLUP	CV-EBLUP
5	1	1	0	4.839	0.879	0.937	0.193
7	1	1	0	6.557	0.879	0.937	0.142
3	1	1	0	3.120	0.879	0.937	0.300
3	1	1	0	3.120	0.879	0.937	0.300
9	1	1	0	8.276	0.879	0.937	0.113
3	1	1	0	3.120	0.879	0.937	0.300
1	1	1	0	1.402	0.879	0.937	0.668
1	1	1	0	1.402	0.879	0.937	0.668
3	1	1	0	3.120	0.879	0.937	0.300
1	1	1	0	1.402	0.879	0.937	0.668
6	1	1	0	5.698	0.879	0.937	0.164
3	1	1	0	3.120	0.879	0.937	0.300
3	1	1	0	3.120	0.879	0.937	0.300
3	1	1	0	3.120	0.879	0.937	0.300
1	1	0	0	1	0	0	0
5	1	1	0	4.839	0.879	0.937	0.193
8	1	1	0	7.417	0.879	0.937	0.126
6	1	1	0	5.698	0.879	0.937	0.164
2	1	1	0	2.261	0.879	0.937	0.414
3	1	1	0	3.120	0.879	0.937	0.300
8	1	0	0	8	0	0	0
1	1	1	0	1.402	0.879	0.937	0.668
10	1	1	0	9.135	0.879	0.937	0.102

Note: HSiZE: Household size, W denotes variable water1, L: electricity for lighting1, T: toilet1 .