A BAYESIAN HIERARCHICAL MODELLING OF SMALL AREA VARIATION IN YOUTH UNEMPLOYMENT IN NAMIBIA

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN APPLIED STATISTICS AND DEMOGRAPHY
OF
THE UNIVERSITY OF NAMIBIA

BY

LINDA VUTE SHITENGA

200308122

OCTOBER 2023

SUPERVISOR: DR. PETRUS IIYAMBO (UNIVERSITY OF NAMIBIA)

ABSTRACT

Youth unemployment has been one of Namibia's socio-economic problems, which has the potential to have significant and serious social repercussions on economic growth and development and could cause social exclusion and unrest in the affected country. Youth unemployment rates estimates are only available at the national (46.1 percent in 2018) and regional levels in Namibia; however, the Namibia Labour Force survey (NLFS) does not provide such statistics at small area e.g., at constituency level. The census data could be used to provide estimates of youth unemployment at constituency level; however, the data is only obtained every 10 years which in most cases the time frame is too long given the developmental changes that may take place during the 10-year periods. In view of these challenges, it is paramount to estimate the variation in unemployment rates at constituency level for possible targeted interventions within regions. In comparison to conventional small area estimation (SAE) models, the hierarchical Bayesian approach to SAE problems has several benefits, one of which is the ability to properly account for the kind of surveyed variable. For this reason, the main objective of this study was to estimate the risk of youth unemployment at constituency level using the 2018 NLFS data. The likelihood was estimated using a hierarchical Bayesian model and results from the study showed that the chance of youths being unemployed was very high among male youths than female youths in urban areas with OR=1.35 (1.10, 1.66) and OR=0.79 (0.65, 0.96) respectively. Several models were fitted, and the best model was used to estimate the probability of being unemployed amongst the male and female youths (with the DIC values of 4900.90 for the males and 5719.48 for females). The best model considered the fixed effects together with the unstructured spatial effects at constituency and regional levels. Even though employment is the result of aggregate demographic and socio-economic factors, the study recommends that employment opportunities specifically targeting youths should be created by either government or private sectors in constituencies especially those in the rural constituencies. Furthermore, special attention should be paid to integrating the youth into the labour market by improving their educational levels.

Keywords: Unemployment rate, youth**,** small area estimation**,** Hierarchical Bayesian

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CAR	Conditional Autoregressive
DIC	Deviance Information Criterion
EB	Empirical Bayes
FB	Fully Bayesian
GMCAR	Generalized Multivariate Conditional Autoregressive
GMRF	Gaussian Markov Random Field
HB	Hierarchical Bayesian
IAR	Intrinsic Autoregressive
INLA	Integrated Nested Laplace Approximation
MCAR	Multivariate Conditional Autoregressive
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
STAR	Structured Additive Regression Model
OR	Odds Ratio

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Lord Almighty for giving me the strength and ability to successfully complete my master’s studies at the University of Namibia. It was indeed a struggle and without Him, it would not have been possible. To my supervisor Dr. Petrus Iiyambo, and my other lecturers for the master’s programme Prof. Lawrence Kazembe, Dr. Opeoluwa Oyedele and Prof. Lillian Pazvakawambwa, thank you for your guidance, advice, and inspiration throughout the duration of my research work. I wish also to express my sincere gratitude to my family, friends for constantly encouraging me to push through each time I got overwhelmed and overburdened with my schoolwork.

Finally, I would also like to thank appreciate my colleagues especially Mr. Titus Johannes and fellow colleagues from Namibia Statistics Agency (NSA) for sharing their knowledge and technical expertise with regards to data analysis of this study and report writing.

DEDICATIONS

This research/thesis is dedicated to my mother that has been my pillar of strength given the difficulties I have been going through during the past years. In the same vein this thesis is also dedicated to my former supervisor Mr. Daniel Oherein (Former Manager: Social Statistics at the Namibia Statistics Agency) who mentored and nourished me in the field of Labour Statistics. I am forever grateful to him.

DECLARATIONS

I, Linda Vute Shitenga, declare hereby that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institution of higher education.

No part of this thesis may be reproduced, stored in any retrieval system, or transmitted in any form, or by means of electronic, mechanical, photocopying, recording or otherwise without the prior permission of the author, or the University of Namibia in that behalf.

I, Linda Vute Shitenga, grant the university the right to reproduce this thesis in whole or in part, in any manner or format, which the University of Namibia may deem fit, for any person or institution requiring it for study and research; providing that the University of Namibia shall waive this right if the whole thesis has been or is being published in a manner satisfactory to the university.

Signature Date

Shitenga…………………………… OCTOBER 2023……………………………..

CHAPTER 1: INTRODUCTION

1.1 Background of the study

Youth unemployment in Namibia has been one of the social-economic problems, with the national youth unemployment rate recorded at 46.1 percent as of 2018. Youth unemployment is defined as the number of unemployed 15–34-year-olds expressed as a percentage of the youth labour force (OECD, 2021). As per the Namibian Labour Force survey (NLFS) which used the broader definition of unemployment, "applies when a person was available for work in the preceding seven days but does not require the person actively sought work. This definition of unemployment is appropriate for developing countries where there are limited formal avenues through which people can look for work. Youth participation in education, training, or employment has a significant impact on the nation's future economic development, growth, and stability. Although it has the potential to have major and serious social implications that could result in social marginalization and unrest, youth unemployment is frequently disregarded (Dietrich, 2012). In most European countries youth unemployment has risen significantly because of the economic crisis (Dietrich, 2012). Compared to youth unemployment, according to Dietrich (2012) adult unemployment experienced the effects of the crisis with some delay but in the long run, the recession has affected all age groups as it was evident that in the 2000s there were significant changes in the pattern of youth unemployment due to gender, citizenship, or educational level.

Youth unemployment is considered a barrier to the Southern African region's development (Devlin, 2013). On the African continent some of the highest rates of unemployment are in Southern Africa, where 51 percent of young women and 43 percent of young men are unemployed (Delvin, 2013). In sub-Saharan Africa it is reported that “young people under the age of 25 represent 3/5 of sub-Saharan Africa’s unemployed population (ILO, 2014). Difficulties finding and sustaining employment detract from a young person’s lifetime productivity and earnings, making it more challenging to escape poverty (Delvin, 2013).

The World Bank reported in 2014 that the youth accounted for 60 percent of all unemployed Africans. However, youth unemployment in North African countries remains the highest in the world reaching more than 29percent in 2013 (ILO, 2014c). Youth unemployment is expected to continue increasing in the future, because currently unemployment was estimated at 74 million in 2013, which showed there was a huge increase by 3.4 million from 2007 to 2012 (ILO, 2013b).

One of the main causes of youth unemployment in the SADC region, particularly in Zimbabwe, is the incompatibility of the curriculum with the needs of twenty-first century industry (Machadu and Jena, 2015). As a result of this disjuncture, school-leavers’ qualifications are largely irrelevant to national economic needs.

According to the 2018 Namibia Labour Force Survey (LFS), conducted by the Namibia Statistics Agency, youth unemployment rate in Namibia increased from 43.4 percent in 2016 to 46.1 percent in 2018. This represents a 2.7 percent slight increase. In addition, with the youth making up more than half (52 percent) of the labour force population in Namibia, the increasing levels of youth unemployment rate is still one of the major socio-economic problems in the country. Due to a lack of frequent data at the small area level, particularly on youth unemployment there has been a significant increase in the need for small area labour force statistics in recent years given the high rate of youth unemployment in the country. Hence data at lower level is necessary to identify the specific constituencies in each region that contribute significantly to high youth unemployment rates in Namibia to make it easier for necessary interventions to be devised by policy makers. For this reason, this study aimed at estimating the risk of youth unemployment at constituency level using the Bayesian Hierarchal model.

1.2 Statement of the problem

The Namibia Statistical Agency (NSA) only generates official statistics on the indicators of the labor force (LF) at the national and regional levels. In Namibia, the census that yields estimates of youth unemployment rates at lower levels, such as constituencies, is only carried out every ten years. As a result, after a few years of collection, the information from the census data may not necessarily be accurate reflections of the actual situation on the ground. This study is employing the use of the small area estimates (SAE) approaches for the Fay-Herriot (FH) model.

The Fay-Herriot (FH) model, an area-level model, was used in this study to identify which constituencies have high probability of youth unemployment levels for potential targeted interventions because the current labor force surveys do not produce reliable indicators at constituency levels due to the smaller sample sizes of youth unemployment statistics at these levels. There is however a growing demand for up-to-date statistics on the labor force countrywide.

This will be particularly valuable in terms of resource allocation within regions. When the unemployment rate is low, for example, in Khomas region, that is only an average figure; however, people in those peri urban areas are nearly 90 percent unemployed. However, because it is not targeted per constituency, this type of reporting overshadows planning, resulting in misdirected response programs or interventions.

This kind of reporting ill-informed the current programmes as they do not target constituencies but rather regions, which leaves the issues partly resolved. The statistics at the constituency level will increase the effectiveness of the targeted interventions.

1.3 Research objectives

1.3.1 Main objective of the study

The main objective of this study was to model small area variation in youth unemployment rate in Namibia at constituency level in all the 14 regions of Namibia using hierarchical Bayesian model.

1.3.2 Specific objectives of the study:

The specific objectives of the study were:

- To estimate the risk of youth unemployment at each constituency level in Namibia.
- To evaluate different spatial models applicable for the analysis of constituency-level LF indicators for youth unemployment.

1.4 Significance of the study

Countries face a range of challenges in their labour force across the development spectrum, such as insufficient employment generation and that timely, relevant, and accurate labour force statistics which are essential for planning, implementing, monitoring, and evaluating policies (Pietschmann et. al, 2016) such as employment creation policy that is spearheaded by the Ministry of Labour, Industrial Relation, and Employment Creation. These polices are linked to the SDGs. For example, SDG 1 which is targeted at addressing poverty, SDG 2which is targeted at addressing hunger and malnutrition, SDG 3 which is targeted at addressing health problems, SDG 4 which is aimed at addressing the non-completion of education for school for children, and SDG 8 which is targeted at addressing the issues related to employment. Currently these statistics are only produced at macro level such as regional levels which is often difficult for the regional planners to use these statistics as they require it at lower levels such as, constituency level. Hence, the findings from this study will be useful in the monitoring and evaluating of youth unemployment regional programs and targeted interventions especially at constituency levels.

1.5 Limitation of the study

The study used secondary data from the 2018 LFS. The LFS coverage was limited to persons in private households excluding those in institutions at the time of the survey, such as school hostels, army/police barracks, hospitals wards, etc. Household members residing in these institutions were only included in the survey if they live in their own private accommodation.

1.6 Organization of chapters

This thesis is organized as follows: Chapter 1 presents the background information, problem statement, research objectives, limitation, and significance of the study, while Chapter 2 presents the literature review. Chapter 3 presents the methodology of this research study while Chapter 4 shows the analysis and interpretations of results. Chapter 5 presents conclusions and recommendations made based on the results of the study.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter reviews previous research and scholar work of youth unemployment on the international level, SADC regions as well as for Namibia. This chapter provides key definitions of the study covered. Three key variables of the study were outlined in the literature which are: youth unemployment, small area variation, and Bayesian Hierarchal modelling. Relevant statistical models used in modelling youth unemployment in small areas include Bayesian approach, Monte Carlo method and their short comings are discussed, parameter estimation as well as model selection criteria is also discussed.

2.2 Definitions

Section 2.2 presents an overview of the key definitions used in this study. This section is centred on youth, unemployment, youth unemployment, small area variation, and Bayesian Hierarchal models which are the key terms used throughout the study.

2.2.1 Youth

United Nations (United Nations, n.d.) (O'Higgins, 2007) defined youth as young persons between the age of 15 and 24 years. However, according to the African Youth Charter by the African Union, defines youth as persons between the ages of 15 to 34 years. The definition of youth is very broad, and it varies from country to country, and it is also dependent on political, institutional, and cultural factors. The broadness and notion of the term ‘Youth’ should be put into consideration. This study of youth unemployment is centred on the Namibian perspective, hence for the purpose of this study, youth will mean any persons between the ages of 15 and 34 years including students, who while studying, were at the same time in paid or self-employment or available for work following the labour force priority rule which is; precedence is given to employment over unemployment and to unemployment over economic inactivity.

2.2.2 Unemployment

Broad unemployment (used by this study) applies to when a person was available for work seven days prior to interview but does not require the person actively sought work. This definition of unemployment is appropriate for developing countries where there are limited formal avenues through which people can look for work. Unemployment is regarded as a key measure of the economy which is often regarded as an unemployment rate. Unemployment rate refers to the number of unemployed people in a geographically specified area divided by the number of people in the labor force (Somer & Adam, 2021).

2.2.3 Youth Unemployment

Youth unemployment is a condition whereby young people are actively looking for a job, but cannot find a job, within the age limit defined by the United Nations as 15-24 years (O'Higgins, 2007). According to (Namupala, 2016) unemployment represents the number of unemployed young people aged 15 – 34 years, as a percentage of the total labour force, where the total labour force comprises the unemployed and the employed.

In this study, youth unemployment refers to a condition of young persons between the age of 15 and 34 who are without work and are available for work as per the broad defined by ILO broadly. Youth unemployment is one of the many socio-economic problems affecting all countries in all over the world. Socio-economic problems have a serious effect on socio-economic status of the country as well as the living standards of the people. These problems have impact on the nation as it can increase the rate of nepotism, poverty, crime, depression and suicide, parochialism, bribery etc., according to the economic report on Africa (ILO, 2010), youth unemployment reflect the failure to make use of the important factors of production, labour for fostering economic growth.

Teshita (2018) found that youths are among the most important resources countries need to have to acquire prosperity as they can make changes to the social economic development if they are utilised in a good way. It is very crucial to understand the structure, patterns, and causes of youth unemployment as it can help in policy interventions which leads to well fused data to totally direct the interest of decision makers (NSA, 2015).

2.2.3.1 Factors Influencing Youth Unemployment

This section reviews the factors influencing youth unemployment. Examples of factors that were found to influence the youth unemployment were age, ethnicity, region, level of education, and gender. These factors are reviewed based on the three perspectives namely: on the global, African, and Namibian perspective.

2.2.3.1.1 Youth Unemployment rate on Global Perspective

Approximately 88 million young people were unemployed in 2003 throughout the world, accounting for 47 percent of the 186 million people unemployed globally and representing an increase of almost 4 percent between 1993 (11.7 percent) and 2003 (14.4percent) (Economic Report on Africa, 2005). In 2003, globally, young people (persons of 15 - 30 years) were 3.5 times more likely to be unemployed than adults (Economic Report on Africa, 2005). In terms of geographic distribution, young people in developing countries were 3.8 times more likely to be unemployed compared with older workers, whereas in the industrialized economies young people were 2.3 times more likely to be unemployed (Economic Report on Africa, 2005). Global youth unemployment figures are particularly disturbing because young people only make up 25 percent of the working-age population (Economic Report on Africa, 2005).

According to Verd, Barranco, & Bolíbar (2019), Spain, Greece, Italy, and Croatia, are some of the countries in the European Union where unemployment has hit the youth the hardest during the Great Recession. Also, in its aftermath, unemployment rate among Spanish workers under 25 years of age reached peaks above 50 percent in 2012, 2013 and 2014. Since these peaks, the youth unemployment rate has been slowly declining, although in 2017 it continued to be above 35 percent (Spanish Labour Force Survey, 2017).

The global concern with youth unemployment has been manifested by an increased commitment to addressing the problem by national governments, non-governmental organizations, and regional and international organizations. For example, Goal 8 of the United Nations Millennium Development Goals, which call for a global partnership for development, specifically refers to the creation of employment opportunities for young people. The United Nations, ILO and World Bank have also come together to form the Youth Employment Network (YEN), which brings together leaders in industry, youth and civil society and policymakers to respond to the challenge of youth unemployment (ILO, 2004). YEN operates on the basis of four global priorities employability, entrepreneurship, employment creation and equal opportunity which also function as a framework for analysis and action. Since Namibia is one of the countries which has volunteered to be a ‘champion country’ for the purpose of showcasing successful youth employment practices, this study also examined Namibia's record in view of the standards set and recommendations made by YEN and other relevant bodies and actors.

2.2.3.1.2 Youth Unemployment rate on African Perspective

According to Nwuke (2002), Africa's population is very young. That is, young people make up more than 50 percent of the population of the countries in Africa, the region that the United Nations Economic Commission for Africa serves. Young people have potential resource for growth and social development if they are productively engaged. They could also be a source of devastating social tension and conflict if they are not well taken care of.

In many countries, the degree to which youth can contribute to the possibilities of their countries and the continent in general is constrained by circumscribed life chances, job opportunities being one of the major circumscriptions. The countless number of jobless young people on the streets of major African cities bears ample witness to the limited jobs and other opportunities that the youth of the continent have. Nwuke (2002) further stated that the phenomenon resulted in the fact that youth unemployment in Africa ranked the second highest in the world following the Middle East and North Africa where the youth unemployment rate was recorded at 25.6 percent during the same period according to the economic report on Africa for 2005 (Africa, United Nations Economic Commision for Africa, 2005).

There is a high proportion of young people in countries such as Benin, Niger and Zambia. Young people under the age of 18 make up to 55 percent of the population, this has led to the problem of youth unemployment to be even more pronounced in Africa. In addition, according to the Kaapanda (2007), Mauritius' youth unemployment rate was recorded at 23.8 percent in 1995 with a share of youth unemployed compared to the population at 12 percent.

In 2000 in South Africa, 55.8 percent of the youth were unemployed of whom 57.9 percent were between 15 and 24 years old (Kaapanda, 2007). It is also argued that each year there are about 8.7 million new entrants into the labour market in Africa, for whom jobs should be found according to the Economic Report on Africa of 2005 has realized that youth unemployment in Africa is characterized by interrelated factors which may not necessarily be unique to Africa, but which exhibit themselves to a different degree.

According to the to the Economic Report on Africa of 2005, the reasons for a high youth unemployment rate in Africa has been related to the following factors: a) lack of employment opportunities due to the under-development of economies; b) high rates of population growth; c) sluggish or stagnant economies; d) low literacy and numeracy rates; e) poor quality education and an education sector that increasingly equips young people with limited industrial skills and which fails to align the curricula with the needs of the labour market. Other factors that ILO has observed include globalization, armed conflict and population growth (Economic Report on Africa, 2005).

2.2.3.1.3 Youth Unemployment rate on Namibia Perspective

Several researchers have investigated factors that are associated with the likelihood of youth unemployment in Namibia. Young people generally have poor employment prospects. According to (Gary & Nicole, 1998) low level of educational qualifications and poor performance in literacy and numeracy are the major factors contributing to youth unemployment in Australia. In Namibia, young people tend to have a higher probability of being unemployed due to lack of working experience and ability to get absorbed into the job market.

Namibia is one of the developing countries and the Germany colonialism has a bigger impact on the economy and the unemployment rate to date. The colonial history and a harsh desert physical environment do not provide many opportunities for mass industrialization and employment according to Ministry of Labor (1997).

The socio-economic structure which was inherited at independence in 1990 was characterized by (a) an underdeveloped economy with a small modern sector, dominated by mining, a small manufacturing sector, a very underdeveloped agricultural and informal sector, and a low gross domestic product (GDP) growth rate; (b) low levels of education and skills among most of the black population; and (c) racial inequity in the distribution of employment opportunity, income, wealth, and social services.

According to the Namibia Labour Force Survey of 2018, Namibia has an overall unemployment rate of 33.4 percent, and a youth unemployment rate of 46.1 percent. This implies that almost half of Namibia’s economically active population is unemployed. According to Kaapanda (2007), the causes of youth unemployment in Namibia includes the education system that has failed to provide young people with the necessary skills that will allow them to be employed. The effect of youth unemployment include crime, alcoholism, feelings of personal humiliation and family disintegration.

The Namibia Statistics Agency reported that the country's youth unemployment rate rose from 37.8 percent in 2012 to 46.1 percent in 2018, as seen in Figure 1 (Namibia Labour Force Survey [NLFS], 2012, 2018).

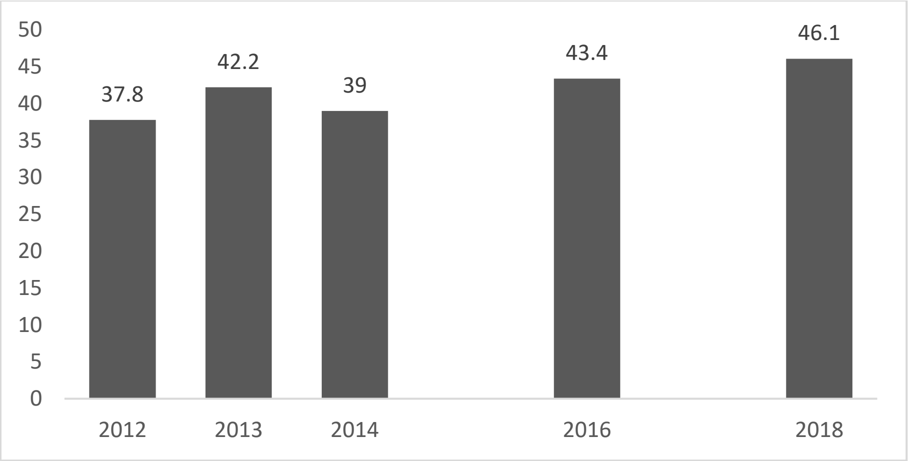


Figure 1: Youth unemployment rate (%) in Namibia from 2012 to 2018
**The youth unemployment rate (%) for 2015 and 2017 is not recorded as there was no labor force survey conducted in Namibia during the two years.*

According to the New Era newspaper dated 26th July 2021, youth unemployment is expected to reach 50 percent by the year 2021, this was a figure revealed by Sheila Roseau who is United Nations Population Fund (UNFPA) country representative during the launch of the third National Youth Policy. On contrary, the most recent statistics from the NSA report of 2019 indicate that the country's overall unemployment rate dropped slightly from 34 percent in 2016 to 33.4 percent in 2018. Therefore, it is very crucial for the government and its development partners to invest in young people for the potential of socio-economic benefits as youth unemployment poses a potential threat to the social stability and economy of the country at large.

2.2.4 Small Area Variation

A variety of statistical methods can be used in small-area analysis to test whether there is more variation than would be expected by chance as indicated by (Diehr, 1992). Small area analysis (SAA) describes statistical methods or techniques used to produce adequate estimates when there is insufficient information or small sample size as further described by (Diehr, 1992). This can be achieved by the use of small area methods that make use of auxiliary data available at the small area level, such as administrative data or data from the last census to construct predictor variables for use in a statistical model that can be used to predict the estimate of interest for all small areas. The SAA focuses on specific areas or populations to identify significant differences and needs among well-defined small areas within a larger geographic boundary. It may refer to regions and subregions like constituencies and enumeration areas/primary sampling units as minor geographic areas hence researchers use SAA conclusions to describe relationships between observed variations, identify community needs, and influence concentrated community efforts to high need areas. López-Vizcaíno et al., (2015) applied the methodology of small area estimation on labour force indicators to the e Spanish Labour Force Survey of Galicia. The labour force indicatory that the study focus on were the totals of employed and unemployed people and unemployment rates where small area estimators of these quantities were derived from four multinomial logit mixed models, including a model with correlated time and area random effects. In the same study, mean squared errors were used to measure the accuracy of the proposed estimators and they were estimated by analytic and bootstrap methods. They introduced area-level multinomial mixed models, without temporal components, for estimating the total number of employed and unemployed people.

Berg & Fuller (2014) developed a small area procedure to the Canadian Labour Force Survey for a two-way table of proportions where the estimated proportions were from a complex survey and estimation was difficult because of the observed proportions that do not have multinomial distributions, the observed proportions were correlated with estimated variances, benchmarking was required, and mean models were nonlinear. They then specified a predictor that was based on a nonlinear mixed model were a bootstrap estimator of the mean squared error of a benchmarked predictor was suggested and performed in simulations. The procedure was then applied to the proportions in the two-way table defined by occupations crossed with Canadian provinces. Through the application of the prediction procedure to the Canadian LFS data, this led to gains in the estimated mean squared errors relative to the direct estimators between approximately 30 percent and 80 percent and this procedure was further supported by the comparison of the predictors to the Census 2006 proportions.

Molina et al., (2007) developed a new methodology for the estimation of unemployment or employment characteristics in small areas based on the assumption that the sample totals of unemployed and employed individuals follow a multinomial logit model with random area effects. This method was illustrated with the United Kingdom labour force data aggregated by se-age groups. Furthermore, two different estimators of the mean-squared error were given; an analytical approximation obtain Taylor linearization and an estimator based on bootstrapping in which the simulation study for comparison of the two estimators showed good performance of the bootstrap estimator.

2.2.5 Bayesian Models

Sample surveys have long been a well-liked and effective way to gather information on socio-demographic, labour, and health factors, as well as unemployment, income, and poverty levels. These data are tabulated starting at the unit level, up to an aggregate level like county or state. While sample surveys are reliable sources for direct estimates of means or totals of variables of interest for large areas or domains, for smaller domain sample size plays a crucial role in prediction bias and in determining the uncertainty associated with these estimates (Nandy et al. (2022). Given this background, Bayesian models can help in providing good quality small area estimates.

Bayesian statistics involves evaluation of a conditional probability of an event A given event B. According to Walsh (2002), as opposed to the point estimators (means, variances) used by classical statistics, Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior density for these parameters. As such, he further stated that Bayesian statistics provides a much more complete picture of the uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed.

The Bayesian probability model's formula can generally be written as follows:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

where $P(A|B)$ denotes the probability of event A occurring, given event B has occurred, $P(B|A)$denotes the probability of event B occurring, given event A has occurred, $P(A)$ denotes the probability of event A occurring, and $P(B)$ denotes the probability of event B occurring.

2.3 Statistical models

Section 2.3 reviews literature of the statistical models applied and referred to in this study.

Anjoy et al. (2018) applied the Hierarchical Bayes approach of SAE to the Household Consumer Expenditure Survey using 2011/2012 Population Census. Esteban et al. (2020) derived the employed, unemployed, and inactive persons and unemployment rate estimates using four multinomial logit mixed models. One of the models was based on the correlated time and area random effects, which was applied on the Spanish Labour Force Survey of Galicia. In their effort to estimate asthma prevalence across districts considering the sampling weights, Vandendijck et al. (2016) developed a predictive model-based approach to estimate the prevalence of a binary outcome for both the sampled and non-sampled individuals, using hierarchical Bayesian models that consider the sampling weights. In addition, they also carried out a simulation study to compare the performance of the proposed method with other established methods of SAE of which the results indicated that their proposed method achieved great reductions in mean squared error when compared with standard approaches. In the same vein, this method also performed equally well when compared with methods that are more elaborate when there is a relationship between the responses and the sampling weights.

2.3.1 Estimation of Model parameters

Approaches to parameter estimation of the models in disease mapping or SAE are discussed in the following sections. Empirical Bayes (EB) and a Fully Bayesian (FB) approaches are usually used for smoothing purposes (Ugarte et al., 2008). For maximum likelihood estimation of Generalized Linear Mixed Models with counts, numerical integration of values is usually required. The penalized quasi-likelihood technique, a Laplace approximation to the quasi-likelihood can reduce the problem to a series of weighted least squares regressions (Breslow and Clayton, 1993).

2.3.1.1 Maximum Likelihood Estimation

When the data is observed, the likelihood function is constructed. The likelihood function is the product of the joint probability function of the data, but viewed as a function of the parameters, treating the observed data as fixed quantities.

Assuming that the data values, $y = (y_1, \dots, y_n)'$ are obtained independently, the general form of a likelihood function of the parameter θ given the data y is given by

$$L(\theta; y) = p(y_1, \dots, y_n|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$$

The log-likelihood function denoted by $l(\theta; y)$ is obtained by taking the natural logarithm of the likelihood function. In the case when $l(\theta; y)$ is explicit (that is, it is differentiable), the maximum likelihood estimates that maximize $l(\theta; y)$, we evaluate the first order partial derivative of $l(\theta; y)$ with respect to the parameter of interest and then equate the derivative to zero. The parameter of interest is then solved to obtain its maximum likelihood estimator. However, when $l(\theta; y)$ is not differentiable, the maximum likelihood estimates are obtained using numerical methods, for example Newton Raphson.

2.3.1.2 The likelihood Function

Let $\{y_i\}$ $i = 1, \dots., m$ be a random sample with probability density function $f(y_i|\theta)$, where $\theta = \{\theta_1, \dots\dots, \theta_p\}$ is a p length vector of relative risk parameters. Then, the likelihood function of $\{y_i\}$ is defined as,

$$L(\theta|y) = \prod_{i=1}^{m} f(y_i|\theta)$$

The assumption is made here that the sample values of $y = \{y_1, \dots, y_m\}$ given θ are independent, making it possible to take the product of individual contribution of $f(y_i|\theta)$ in (2.2) (Lawson, 2008). In Subsections 2.5.1.1 and 2.5.1.2, we present the likelihood functions for the Binomial and Poisson distributions respectively.

2.3.1.3 Binomial likelihood for count data

In the case where we examine arbitrary small areas, usually a count of disease is observed within each spatial unit. We define this count as y_i and assume that there are m small areas. We also assume that there is a finite population within each small area out of which the count (n_i) of a disease has arisen.

In this situation, we use a binomial distribution to model the count data (the number of diseases) conditional on the observed population in each area. Hence, we can assume that given the probability of a case is p_i, then y_i is distributed independently as $y_i \sim Bin(p_i, n_i)$ and that the likelihood function of the parameter p_i given the data y_i is given by

$$L(p_i|y_i) = \prod_{i=1}^{m} \binom{n_i}{y_i} {p_i}^{y_i} \ (1 - p_i)^{n_i - y_i}$$

The commonest link function for the probability p_i to a linear predictor is a logit link so that $p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$. The model specification is envisaged within η_i to include spatial and non-spatial components.

2.3.1.4 Poisson likelihood for count data

The most encountered model for small area count data is the Poisson model. This model assumes that the unemployed count $y_i's$ have a mean and are independently distributed as $y_i \sim Poiss(\mu_i)$. The likelihood function of μ given the data y is given by:

$$L(y|\mu) = \prod_{i=1}^{m} \mu_i \exp(-\mu_i) / y_i!$$

The expectation of Y_i is $E(Y_i) = \mu_i = \varepsilon_i \theta_i$ where ε_i is the expected rate for the ith area and θ_i is the relative risk for the ith area (Lawson, 2008).

2.3.2 Integrated Nested Laplace approximation

Integrated Nested Laplace Approximation (INLA) is mostly used in implementing Approximate Bayesian Inference (Martino and Rue, 2010). Martino and Rue (2010) explained INLA as a new approach that is used for statistical inferences for Latent Gaussian Markov Random Field (GMRF) models and was introduced by Rue, Martino, and Chopin (2009). According to Martino and Rue (2010), it provides fast deterministic alternatives to MCMC that is now the standard tool for inference in GMRF models.

A Latent GMRF model is a hierarchical model where, at the first stage a distributional for the observables y usually assumed to be conditionally independent given some latent parameters η and, possibly, some additional parameters θ_1 is found. A Latent GMRF model is presented as ((Martino & Rue, 2010)).

$$\pi(y|\eta, \theta_i) = \prod_j \pi(y_j|\eta_j, \theta_1)$$

The latent parameters η are part of a larger latent random field x, which constitutes the second stage of our hierarchical model. The latent field $x$is modelled as a GMRF with precision matrix Q depending on some hyperparameters θ_2.

$$\pi(x|\theta_2) \propto \exp\{-\frac{1}{2}(x-u)^T Q(x-u)$$

The third, and last, stage of the model consists of the prior distribution for the hyper parameters $\theta = (\theta_1, \theta_2)$.

The INLA approach provides a recipe for fast Bayesian inference using accurate approximations to $(\boldsymbol{\theta}|\boldsymbol{y})$ and $(x_i|y)$, i = 0,..., n - 1, i.e. meaning that the marginal posterior density for the hyperparameters and the posterior marginal densities for the latent variables.

Different types of approximations are available, and the approximate posterior marginal can be used to compute summary statistics of interest, such as posterior means, variances or quantiles Harris (2017).

2.3.3 Empirical Bayes approaches

Most of the statistical techniques for smoothing the estimates use Empirical Bayes (EB) method. The basic idea from Efron and Morris (1975) is pooling the information across the regions through a suitable model using James-Stein (1961) estimators (Gómez-Rubio & López-Quılez, 2006).

In the EB approach, the posterior distributions of the parameters of interest given the data are first obtained assuming that the model parameters are known. The model parameters are then estimated by suitable methods and inferences for those parameters are made from the estimated posterior distributions (Ugarte, Goicoa & Militino, 2009).

The application of EB approaches can be found in studies conducted by Clayton and Kaldor (1987), Cressie and Chan (1989) and Cressie (1992) using Poisson likelihood, log-normal prior and the spatial modeling of log-estimates.

2.3.3.1 Fully Bayesian Approach

The Fully Bayesian (FB) approach in SAE exploits the posterior distribution of the Relative Risks to obtain reliable and smooth estimates (Ugarte et al., 2008). This approach became popular because of its flexibility and the availability of free software for implementing Marcov Chain Monte Carlo (MCMC) methods allowing the fit of complex models (Ugarte et al., 2008). One of the main advantages of the FB approach is that it provides, for each area, samples of the whole posterior distribution of the RRs, supplying more information than just a single point estimate.

All the parameters are assigned prior distribution whose parameters are also assigned hyperprior distributions to cope with their possible variability in a FB setting. The estimation procedures were done using MCMC simulation techniques. Best, Richardson and Thomson (2005) summarised the most recent Hierarchical Bayesian models that are used for disease mapping using Full Bayes estimation.

In the Bayesian framework, all information about the parameter θ directly from the data is contained in the likelihood function. The values of the parameters that correspond with the largest values of the likelihood function are the parameters that are most supported by the data. To obtain the posterior distribution $p(y_i|\theta)$, the probability distribution of the parameters once the data have been observed, the Bayes' theorem is applied such that:

$$p(\boldsymbol{\theta}|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(\boldsymbol{y}|\boldsymbol{\theta})d\theta} = \frac{p(\theta)L(\theta|y)}{p(y)} \propto p(\theta)L(\theta|y)$$

where ∝ means "is proportional to" (Meaning that the expressions are equal when the right-most term is multiplied by a normalizing constant that does not depend on θ). To obtain the posterior distribution, the prior distribution is multiplied by the likelihood function and the constant, which does not depend on the parameter θ, is then determined.

2.3.4 Prior Distributions

Lawson (2008) defined a prior distribution of the parameter θ as a distribution assigned to the parameter θ before observing the data $\{y_i\}$. He further stated that prior distributions provide additional "data" for a problem, and they can be used to improve estimation or identification of parameters. Given a single parameter, the prior distribution is denoted by $\mathrm{g}(\theta)$, while for a parameter vector ϑ, the joint prior distribution is denoted as g(ϑ). Lawson (2008) considered two types of priors, namely, propriety and non-informative priors. In Subsections 2.5.2.1 and 2.5.2.2, we present some properties of these two types of priors.

2.3.4.1 Propriety

A prior distribution is said to be impropriety if its integration of the random variable θ over its range (Ω) is not finite ((Mwahi, 2014). That is,

$$\int_{\Omega} \mathrm{g}(\theta)d\theta = \infty$$

A prior distribution is improper if its normalizing constant is infinite. While impropriety is a limitation of any prior distribution, it is not necessarily the case that an improper prior will lead to impropriety in the posterior distribution.

2. 3.4.2 Informative Priors

An informative prior is a prior that is not dominated by the likelihood function and that has an impact on the posterior distribution. That is, if a prior distribution dominates the likelihood, it is clearly an informative prior.

2. 3.4.3 Non-informative Priors

These priors do not make strong preferences over values of the variables. They have a relatively flat form yielding close-to-uniform preference for different values of the variables. This tends to mean that in any posterior analysis the prior distributions will have little impact on the posterior distribution of the parameters compared to the likelihood of the data (Lawson, 2008).

2. 3.4.4 Conjugate Priors

A prior is conjugate for a family of distributions if the prior and the posterior belong to the same family. Specifically, for a Binomial likelihood with parameter θ, and the beta prior distribution with parameter θ, the posterior distribution of θ is also a Beta distribution (Lawson, 2008). For example,

$$y \sim Bin(n, \theta)$$

$$\theta \sim Beta(\alpha, \beta)$$

$$\theta | y \sim Beta(y + \alpha, n - y + \beta)$$

Hence, a beta prior for θ is a conjugate prior. The conjugate priors are useful as building blocks in more complicated models, even though the full model will not be conjugate (Lawson, 2008).

2. 3.4.5 Jeffery's Priors

Jeffrey's prior (Jeffrey, 1961) is a prior that does not change much over the region in which the likelihood is significant and does not assume large values outside that range. This is based on the Fisher information matrix. Jeffrey's prior for a single parameter θ is defined as

$$P(\theta) \propto \sqrt{I(\theta)}$$

where $I(\theta) = -E[\frac{\partial^2 \ln P(y|\theta)}{\partial\theta^2}]$ is the Fisher information matrix based on the likelihood function $P(y|\theta)$. Based on equation (2.4), Jeffrey's prior distribution for a binomial likelihood with a single parameter θ and sample size *n* is given by

$$P(\theta) = \sqrt{\frac{\theta}{\theta(1-\theta)}} = n^{1/2}\theta^{-1/2}(1-\theta)^{-1/2}$$

2. 3.4.6 Priors for Spatial data

As discussed in Section 2.1, spatial data contain information about the attribute of interest as well as its location in space. The broad principle of regression and modelling apply to spatial datasets. Markov Random Fields (MRFs) are the dominant approach to analysing aerially aggregated spatial data, such as disease counts in administrative units (Paciorek, 2013).

Paciorek (2012) argued that the most common form of MRF represents the spatial dependence structure such that areas that share boundaries are considered neighbours, with an area conditionally independent of any non-neighbour areas, a so-called first order neighbourhood structure. A common alternative to this standard nearest neighbour is to extend the neighbourhood structure beyond bordering areas (Song et al., 2008).

2.3.5 Bayesian Hierarchical Modelling

As mentioned earlier, sample surveys have long been popular and efficient means for collecting data on economic, labour, and health variables, measuring unemployment, determining income and poverty status, and other socio-demographic variables. These data are tabulated starting at the unit level, up to an aggregate level like county or state. While sample surveys are reliable sources for direct estimates of means or totals of variables of interest for large areas or domains, for smaller domain sample size plays a crucial role in prediction bias and in determining the uncertainty associated with these estimates. Domains may be defined by geographic areas, socio-demographic groups, or other sub-populations (Rao and Molina, 2015). A domain (area) is regarded as large if the domain-specific sample is large enough to yield “direct estimates” with adequate precision whereas a small area would denote any domain for which direct estimates of adequate precision cannot be produced (Rao and Molina, 2015).

Since design-based direct estimators fail to produce reliable estimates for small areas, model-based “indirect” estimators have become popular. These estimators borrow information from related areas, variables, and/or time periods thereby increasing the effective sample size for small area estimation problems. Datta (2009), Pfeffermann (2013), and Rao and Molina (2015) provide comprehensive review of popular model-based estimators.

In Bayesian modelling, the parameters have prior distributions. These distributions control the form of the parameters and are specified by the investigator based on their prior belief concerning their behaviour (Lawson, 2008).

2. 3.5.1 Spatial Model Structure for Point and Areal data

Let $\mu_i = E(y_i|X_i, \mathrm{g})$ be related via a link function, h(.) to a linear predictor:

$$h(\mu_i) = X_i^T\beta + K_i\mathrm{g}$$

where, K_i is the ith row of mapping matrix. Let $\mathrm{g}(.)$ be the unknown latent spatial process represented as a piecewise constant surface on a fine rectangular grid.

$$\mathrm{g} \sim N(0, kQ)^-$$

where, Q is an MRF precision matrix and k is a precision parameter, recognizing the potential singularity of Q by using the generalized inverse notation (Paciorek, 2008).

Parciorek (2008) indicated that for point data, K_i would be a sparse vector with a single one (1) that matches the location of the observations to the grid cell in which it falls. We discuss the potential MRF models in Sections (2. 3.5.2 to 2. 3.5.3).

2. 3.5.2 Improper Conditional Autoregressive model

This model was developed from the lattice models of Kunsch (1987), and it uses the definition of spatial distribution in terms of differences between distributions and allows the use of a singular normal joint distribution. The prior for this model is defined as

$$p(\mu|r) = \frac{1}{r^{m/2}} \exp\left(-\frac{1}{2r}\sum_i \sum_{j\epsilon\delta_i} (\mu_i - \mu_j)^2\right)$$

where δ_i is a neighbourhood of the *i*th tract. This neighbourhood is assumed to be defined for the first neighbour only (Rue et al., 2005). Neighbourhoods could consist of first and second neighbours defined by common boundary or by a distance cut-off.

The uncorrelated heterogeneity (v_i) is defined to have a conventional zero-mean Gaussian prior distribution:

$$p(v) = \sigma^{-m/2} \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^{m} {v_i}^2\right)$$

where both *r* and σ are assumed to have improper inverse exponential hyper priors;

$prior(r,\sigma) = e^{-\varepsilon/2r}$ where $e^{-\varepsilon/2\sigma}, \sigma, r > 0$ and *t*=0.001 (Besag et al., 1991).

The full posterior distribution where Poisson likelihood is assumed for the tract counts is given by (Lawson, 2008).

$$p(u,v,r,\sigma|y_{i)} = \prod_{i=1}^{m}\{\exp(-e_i\theta_i)\,(e_i\theta_i)^{y_i}|y_i!\} \times \frac{1}{r^{\frac{m}{2}}}\exp\left(-\frac{1}{2r}\sum_i\sum_{j\epsilon\delta_i}\left(u_i - u_j\right)^2\right) \times \sigma^{-\frac{m}{2}}\exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{m} v_i^2\right\} \times prior(r,\sigma)$$

(16)

This posterior distribution can be sampled using MCMC algorithms such as the Gibbs or Metropolis-Hastings samplers. The conditional moments of the intrinsic Gaussian formulation are defined as simple functions of the neighbouring values and number of neighbours $(n\delta_i\ )$:

$$E(u_i|\ldots) = \bar{u}_\iota \text{ and var } (u_i|\ldots) = r/n\delta_{i.}$$

The conditional distribution is defined as $[u_i|\ldots]\sim N(\overline{u_\iota}, r\,ln\,\delta_i)$ where $\overline{u_\iota} = \sum_{j\in\delta i} u_j \ln \delta_i$ is the average over the neighbourhood of the *i*th region (Lawson, 2008).

2.3.5.3 Proper Conditional Autoregressive model

Spatially-referenced vector of interest is defined as $\{u_i\}$. One specification of the proper CAR formulation yields (Lawson, 2008).

$$[u_i| \dots]\sim N(\mu_i, r\, ln\, \delta_i)\,; \;\; \mu_i = t_i + \emptyset \textstyle\sum_{j\epsilon\delta i}(u_j - t_j) \ln \delta_i$$

where $t_i = x_i'\beta$ is the trend, r is the variance of the distribution and $\emptyset$ must lie on a predefined range which is a function of eigenvalues of a matrix. In detail, the range is the maximum and minimum $(\emptyset_{min} < \emptyset < \emptyset_{max})$ of $diag\{n\delta_i^{1/2}\}.C_{ij}\{n\delta_i^{-1/2}\}$

where $C_{ij} = c_{ij}$ and $c_{ij} = \frac{1}{n\delta_i}\; if\; i{\sim}j$ or 0 otherwise.

2.3.6 Posterior Distributions

The prior distributions and likelihood provide two sources of information about the parameters of any distribution. A prior distribution provides information about the parameter through prior beliefs or assumptions whereas a likelihood provides information via the data. The product of the likelihood and the prior distribution is known as the posterior distribution. This distribution describes the behaviour of the parameters after the data are observed and prior assumptions are made (Lawson, 2008).

Let ϑ be a parameter vector and $\{y_i\}\; i = 1, \dots, m$ denotes a random variable such that $y = \{y_i, \dots, y_m\}$. Then the posterior distribution of the parameter ϑ given the data y is defined as

$$p(\vartheta|y) = L(y|\vartheta)\mathrm{g}(\vartheta)/\mathrm{C}$$

where $C = \int_p L(y|\vartheta)\,\mathrm{g}(\vartheta)d\vartheta$, $\mathrm{g}(\vartheta)$ is a prior distribution of ϑ, $L(y|\vartheta)$ is a likelihood function of y given the parameter ϑ. Here, the quantity C is called the normalizing constant.

2.3.6.1 The Binomial examples

Let y_i denote the number of unemployed youths. We assume that $y_i{\sim}Bin(p_i, n_i)$. Thus, the general form of a likelihood function is given by equation (2.1). Assume that $logit(p_i) = x_i'\alpha + z_i'y$. In this case, $z_i'y$ is a vector random effect, x_i' is a vector of individual cases and y is a unit vector. Suppose a logit link is appropriate for the probability and that a random effect (v_i) at the individual level is to be included. Thus,

$$p_i = \frac{\exp\,(\alpha_0+v_i)}{l+\exp\,(\alpha_0+v_i)}$$

represents a basic model with intercept to capture the overall rate and prior distribution for the intercept. The random effects could be assumed to be $\alpha_0{\sim}N(0, \tau_{\alpha_0})$, and $v_i{\sim}N(0, \tau_v)$ (Lawson, 2008). The hyper prior distribution for the variance parameters may follow either a Gamma, Inverse Gamma, or a Uniform distribution. For instance, a Gamma distribution is defined as follows:

$y_i{\sim}Bin(p_i, n_i)$ such that $\text{logit}(p_i) = \alpha_0 + v_i$ and the appropriate priors summarized as

$$\alpha_0{\sim}N(0, \tau_{\alpha_0})$$

$$v_i{\sim}N(0, \tau_v)$$

$$\tau_{\alpha_0}{\sim}G(\psi_1, \psi_2)$$

$$\tau_v{\sim}G(\phi_1, \phi_2)$$

2.3.6.2 Bayesian Markov Chain Monte Carlo method

Simulation techniques based on Markov chains are very general and flexible. They are widely used to carry out posterior inference in the case where the product of the likelihood and the prior are analytically intractable. Ntzoufras (2011) stated that MCMC methods enable quantitative researchers to use highly complicated models to estimate the corresponding posterior distribution with a given degree of accuracy. An alternative to MCMC methods is the approximate Bayesian inference which is based on integrated nested laplace approximation (Rue et al., 2009; Palcious et al., 2012 and Holand et al., 2013). However, in the following section we only discuss MCMC methods.

2.3.6.2.1 The MCMC algorithm

Ntzoufras (2011) defined a Markov chain as stochastic process $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{t}\}$ such that

$$f\left(\theta^{(t+1)} \middle| \theta^{(t)}, \dots, \theta^{(1)}\right) = f(\theta^{(t+1)} | \theta^{(t)})$$

That is, the distribution of θ at sequence $t + 1$ given all the preceding θ values depend only on the vector $\theta^{(t)}$ of the previous sequence t. When the Markov chain is irreducible, aperiodic, and positive-recurrent, the distribution $\theta^{(t)}$ converges to its equilibrium distribution, which is independent of the initial values of the chain $\theta^{(0)}$ (Nummelin, 2004). In order to generate a sample from $f(\theta|y)$, we must construct a Markov chain with two properties namely, $f\left(\theta^{(t+1)} \middle| \theta^{(t)}\right)$ should be “easy to generate from”; and the equilibrium distribution of the selected Markov chain must be the posterior distribution of interest $f(\theta|y)$ (Ntzoufras, 2011).

From this Markov chain, we can then follow a standard approach to Bayesian inference using the MCMC as follows:

1. Select an initial value $\theta^{(0)}$.
2. Generate T values until the equilibrium distribution is reached.
3. Monitor the convergence of the algorithm using convergence diagnostics. If convergence diagnostics fail, we then generate more observations.
4. Cut off the first $b$observations.
5. Consider $\{\theta^{(1)}, \theta^{(2)}, \ldots.., \theta^{T}\}$ as the sample for the posterior analysis.
6. Plot the posterior distribution.
7. Finally, obtain summaries of the posterior distribution.

2.3.6.2.2 Describing the target distribution using MCMC output

Let $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}, \ldots, \theta^{(T)}$ be a random sample from the MCMC output. Ntzoufras (2011) indicated that for any function $G(\theta)$ of the parameter of interest, we could:

1. Obtain a sample of the desired parameter $G(\theta)$ by simply considering

   $$G\big(\theta^{(1)}\big), G\big(\theta^{(2)}\big), \ldots, G\big(\theta^{(t)}\big), \ldots, G(\theta^{(T)})$$

2. Obtain any posterior summary of $G(\theta)$ from the sample using traditional sample estimates. For example, we can estimate the posterior mean by

   $E(G(\theta)|y) = \overline{G(\theta)} = \frac{1}{T'}\sum_{t=1}^{T'} G(\theta^{(t)})$ and the posterior standard deviation by

   $$SD(G(\theta)|y) = \frac{1}{T^1 - 1}\sum_{t=1}^{T'}[G\big(\theta^{(t)}\big) - E(G(\theta)|y]^2$$

3. Calculate and monitor correlations between parameters.

4. Produce plots of the marginal posterior distributions (histogram, density plots, error bars, boxplots, etc.).

2.3.6.3 Monte Carlo error

The Monte Carlo error measures variability of each estimate due to simulation. It is required that the Monte Carlo error must be low to calculate the parameter of interest with increased precision. There are two most common ways to estimate MC error namely, the batch mean method and window estimator method. These two methods are discussed in full details by Ntzoufras (2011, pp. 39-40).

In the next subsection, we discuss some of the most common basic MCMC algorithms.

2.3.6.3 The Metropolis-Hastings algorithm

Metropolis et al. (1953) originally formulated the Metropolis algorithm by introducing the Markov-chain-based simulation methods widely used in the field of science. This algorithm was further generalized by Hastings (1970) to what is now known as Metropolis-Hastings algorithm.

2.3.6.3 The Metropolis-Hastings algorithm is summarized below:

Let *p(x)* be a target distribution from which we wish to generate a sample of size $N$and $x^{(t)}$ a vector of generated values in $t$iteration of the algorithm. Then:

1. Set initial value $x^{(0)}$.
2. For $t = 1, \dots, N$ repeat the following steps:
    a. Set $x = x^{t-1}$.
    b. Generate new candidate value $x'$from a proposal distribution $q(x \rightarrow x') = q(x'|x)$.

c. Calculate $\alpha = \min\,(1, \frac{p(x)q(x|x')}{p(x)q(x'|x)})$.

d. Update $x^{(t)} = x'$ with probability α and $x^{(t)} = x = x^{(t-1)}$ with probability $\alpha - 1$.

The Metropolis-Hastings algorithm will converge to its equilibrium distribution regardless of whatever proposal distribution q is selected (Ntzoufras, 2011).

2.3.6.4 The Gibbs Sampler

This is another MCMC algorithm which according to Ntzoufras (2001) was introduced by Geman and Geman (1984) as an algorithm for simulating samples from the posterior distribution. The algorithm can be summarized by the following steps:

1. Set initial value $\theta^{(0)}$.
2. For $t = 1, \dots, T$, repeat the following steps:
   a. Set $\theta = \theta^{(t-1)}$
   b. For $j = 1, \dots, d$ update θ_j from $\theta_j \sim f\left(\theta_j \middle| \theta_j, y\right)$.
   c. Set $\theta^{(t)} = 0$ and save it as the generated set of values at $t + 1$ iteration of the algorithm.

Thus, given a particular state of chain $\theta^{(t)}$, we generate the new parameters by

$$\theta_1^{(t)} \text{ from } f(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}, y)$$

$$\theta_2^{(t)} \text{ from } f(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}, y)$$

$$\theta_3^{(t)} \text{ from } f(\theta_3|\theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)} \dots, \theta_p^{(t-1)}, y)$$

. . . . . .

. . . . . .

$\theta_j^{(t)}$ from $f(\theta_j|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_p^{(t-1)}, y)$

. . . . . .

. . . . . .

$\theta_p^{(t)}$ from $f(\theta_p|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{p-1}^{(t)}, y)$.

Generating values from $f\left(\theta_j\middle|\theta_{\setminus j}, y\right) = f(\theta_j|\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_{p-1}^{(t-1)}, y)$ is relatively easy since it is a univariate distribution and can be written as $f(\theta_j|\theta_j, y) \propto f(\theta|y)$, where all variables except θ_j are held constant at their given values (Ntzoufras, 2011).

2.4 Model Selection and Assessments

The foremost goal of model comparison is achieving good balance between two opposing pressures, goodness of fit and complexity. There are different representative measures that help to achieve this balance. In addition, the need to select the best model is of great importance in statistical modelling to ensure goodness-of-fit and adjust or penalize for model complexity. The observed data is usually from an unknown probability distribution and as a result, several models are fitted to the data in order to find the model that best fit the data. Some of the model selection measures are: Deviance Information Criterion (DIC), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The details of these measures are discussed in Subsections 2.8.1, 2.8.2 and 2.8.3, respectively as follows:

2.4.1 The Deviance Information Criterion

The Deviance Information Criterion was introduced as an easily computable and rather universally applicable Bayesian criterion for posterior predictive model comparison (Spiegelhalter et al., 2002). Like many other model selection criteria, it compromises between data and model complexity, and it generalizes AIC which appears as a special case under a vague prior. For example, experiences indicating that DIC works well if the sampling distribution belongs to an exponential family but less so if it is a mixture are hard to explain, and hence modifications are hard to justify. This criterion assumes that a better fitted model has $\Delta DIC > 10$. The value of $4 \leq \Delta DIC \leq 10$ implies a moderate difference and $\Delta DIC < 4$ implies non-distinguishable difference.

2.4.2 The Akaike Information Criterion

For maximum likelihood or empirical Bayesian, one can use the Akaike Information Criterion (Cui et al., 2008). The Akaike Information Criterion was developed by Akaike (1973, 1974) as estimators of the expected Kullback-Lieber discrepancy between the model generating the data and a fitted candidate model (Cui et al., 2008). The Akaike information is defined as $AIC = -2\ln f(y|\theta_k) + 2k$ where k is the number of parameters in the model, f denotes the likelihood function while ln denotes the natural logarithm. The best model that fits the data is determined by examining their relative distance to the real value. The best model is the one with a minimum AIC value. AIC have several advantages e.g., valid for both nested and non-nested models, compare models with different error distribution and avoid multiple testing issues. However, AIC cannot be used to compare models of different data sets.

2.4.3 The Bayesian Information Criterion

Schwarz (1978) as a competitor to the AIC information criterion (Cavanaugh, 1999) introduced Bayesian Information Criterion. BIC is defined as $BIC = -2\ln f(y|\theta_k) + k\ln n$. BIC was derived to serve as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. In large-sample settings, the fitted model favoured by BIC ideally corresponds to the candidate model, which is a posterior most probable, i.e., the model which is rendered most plausible by the data at hand. The computation of BIC is based on the empirical log-likelihood and does not require the specification of priors.

The penalty term of BIC is more stringent than the penalty term of AIC (for $n > 8$, k $ln\ n$ exceeds $2k$). Consequently, BIC tends to favour smaller models than AIC. BIC can be used to compare non-nested models and to compare models based on different probability distributions. However, when the criterion values are computed, no constants should be discarded from the goodness-of-fit term. In a model selection application, the optimal fitted model is identified by the minimum value of BIC.

CHAPTER 3: METHODOLOGY

This chapter provides more details on the methodology that was applied in this study.

3.1 Introduction

In terms of Hierarchical Bayesian modelling unemployment risk amongst Namibian youth, the availability of limited literature for the study suggests that this is the first study that attempted to estimate the geographical variation in youth unemployment at constituency level in Namibia.

3.2 Research Design

The study used secondary data obtained from the 2018 Namibia LFS. The 2018 LFS follows a quantitative cross-section design and is conducted annually on a de facto method. The 2018 LFS sample is based on the 2017 updated National Sampling Frame based on information collected during the 2011 Census conducted by NSA. The labour force surveys in Namibia are conducted by NSA which is the sole entity mandated to produce and disseminate official statistics on labour force.

For economic and social planning reasons, a clear knowledge and understanding of the size, composition and other characteristics of the labour force is indispensable to national, as well as regional planning. The LFS also presents a major step towards the systematic production of labour force information for meeting the realization of one of Namibia’s national development goal Vision 2030. In addition, timely collection and release of labour force statistics are a priority for Namibia to monitor and assess the impact of all policies government has implemented that affect the labour market.

The survey collected data on the labour market activities of individuals aged 15 years and above who lived in Namibia during the reference period (seven days prior to the interview). The LFS is normally conducted during the months of September and October. The LFS, for the sample design employed a stratified two-stage cluster sample, where the first stage units were the PSUs, and the second stage units were the households. The sample sizes were determined to give reliable estimates of the population characteristics at the regional level (i.e., lowest domain of estimation). The PSU were drawn from the master sample frame design for the 2011 Population and Housing Census as mentioned earlier. At the second stage, a fixed number of households in the chosen PSUs were systematically selected. Subsequently all adults of working age (15 years and older) were asked to participate in the study. A key question which was asked was “Have you worked for pay, or profit or family gain for one hour or more in the last week” (Namibia Statistics Agency, 2013).

3.3 Population of the study

The study population was all 13304 youths from ages 15 - 34 years from the 2018 NLFS.

3.4 Sample of the study

This study used 8,714 unemployed youth between the age of 15 and 34 years as per the African youth charter definition of youth unemployment as described below. It should also be noted that even school going youth are part of the analysis as per the ILO specifications.

Employed: The employed comprise all persons of working age who during a specified brief period, such as one week or one day, were in the following categories: a) paid employment (whether at work or with a job but not at work); or b) self-employment (whether at work or with an enterprise but not at work). Temporary absence from work includes reasons such as illness, maternity and parental leave, holiday, training, and industrial disputes (ILO, 2015).

Unemployed in the broad sense: The unemployed comprise all persons of working age who were: a) without work during the reference period, i.e., they were not in paid employment or self-employment; and b) currently available for work, i.e., they were available for paid employment or self-employment during the reference period (ILO, 2015).

3.5 Procedures

The target population of the LFS includes all persons living in private households in Namibia. The survey follows a stratified two-stage random sampling design. The first stage involves the random selection of Primary Sampling Units (PSUs) and second stage involves the random selection of private households within the selected PSUs were all household members are interviewed (NSA, 2018).

In this study, all 121 constituencies of Namibia were covered (I =1,…,121) by the LFS. PSUs were selected depending on population size of the constituency, and thus certain constituencies have more PSUs covered than others (NSA, 2018).

Data sorting and cleaning was carried out using Stata program (Version 14) to remove missing data. For the constituencies and regions, R statistical software (Version 3.0) was used to create boundaries. The boundaries file created by R were used to compute the neighbourhood information of the map of all constituencies and regions in Namibia using BayesX. The boundary maps of all regions and constituencies in Namibia were also used in the different models fitted.

BayesX is the software for Bayesian Inference in Structured Additives Regression Models that is used for spatial analysis for this thesis. BayesX is a free software that can be downloaded from the site (http://www.stat.uni-muenchen.de/~bayesx/). Version 3.0.2 of BayesX developed in (17.07.2015) was used for this thesis. Version 3.0.2 of BayesX permits Bayesian Inference on Markov Chain Monte Carlo simulation techniques. For all the models fitted in BayesX version 3.0.2, 52000 iterations were run with a burn of 2000 for each model.

3.6 Data Analysis methods

Hierarchical Bayesian (HB) approach has recently been proposed for SAE because of the following advantages as stipulated by (Trevisani & Torelli, 2007).

- Their specification is straight forward and allows to take into account the different sources of variation.
- Inferences are clear-cut and computationally feasible in most cases by using standard MCMC techniques.
- One can use this approach when the variable of interest is count or a proportion, alternative model specifications can be considered.

The evaluation of the uncertainty in the unobserved random effects that are also contributing to the variation in the average probability of regional and constituency unemployment rate through the full posterior inference approach was achieved by incorporating the unobservable random effects of interest into the full hierarchical Bayesian model.

3.7 Description of key variables

The outcome variable for this study is the employed and unemployed youth counts data aggregated to the region and constituency level to carry out HB analysis. In addition, socio economic variables were also considered as covariates for modelling.

Table 3.1 shows the description of the outcome variables used in the study and the covariates used to fit models in BayesX. The covariates considered for the study were the characteristics of head of household namely the highest level of education completed, marital status and citizenship. Table 1 below gives the description of these variables.

Table 1 : Description of variables

Covariates	**Description**
Outcome variable	
Unemployed youth (unemployed)	0 = No, 1 = Yes
Spatial effects	
Region (reg)	Administrative boundaries, there were 14 regions in Namibia. See Appendix B
Constituency (const)	Administrative boundaries, there were 121 constituency in Namibia. See Appendix C, part A and part B.
Covariates	
Head of household (head)	0 = No, 1 = Yes
Attending school (Attend)	0 = No, 1 = Yes
No education (noedu)	0 = No, 1 = Yes
Tertiary education (tedu)	0 = No, 1 = Yes
Marital status (mstatus)	0 = Not married, 1= Married
Type of residence (Urban)	0= Rural, 1 = Urban

3.8 Model Building

In order to estimate unemployment rate amongst youths at constituency levels, several models were developed. The outcome variables for this study are binary; hence the binary logistic regression was used which is of the general form:

$$\text{logit}(\pi) = \log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Taking into consideration that classical Generalized Linear Models (GLMs) make no, or limited use of spatial structure of the data, neither do they consider possible nonlinear effects of the risk factors (Mwahi, 2014), a Structured Additive Regression Model (STAR) was fitted, followed by jointed spatial models which was done through the shared component latent variables approach.

According to Mwahi (2014), a STAR model replaces the strictly linear predictor in classical models by a more flexible semi-parametric predictor that incorporates non-linear effects and spatial effects (structured and unstructured). The STAR model that was used in this study is of the form:

$$\eta = \gamma_0 + \gamma_1 noedu + \gamma_2 tedu + \ldots + \gamma_k urban + f^{str}{}_{(const)} + f^{unstr}{}_{(const)}$$

where η is the predictor and γ_i are the fixed effects parameters. Spatial effects of the constituency are split into spatially and correlated part $f^{str}{}_{(const)}$ and an uncorrelated part $f^{unstr}{}_{(const)}$ where the correlated part is modelled by a Markov random field prior (Intrinsic Gaussian Markov random fields (IGMRFs) where the neighbourhood matrix and possible weights associated with the neighbours are obtained from the constituency map. This study assumes diffuse priors for fixed effects, exchangeable normal priors for unstructured spatial random effects and conditional autoregressive priors for the structured spatial random effects.

The study modelled unemployed youth with fixed effects provided in Table 3.1 and the analysis was carried out using version 3.0.2 of the BayesX (Belitz et al, 2009) of which simulation techniques through the MCMC were used for model parameters estimation. For all these models, convergence was reached at 12000 iterations and were run with a burn of 2000 for each model.

The following candidate models were fitted in BayesX (the description of the models fitted in BayesX is given in Table 3.2).

$M_0 = Fixed\ effects\ only\ for\ unemployed$

$M_1 = fspatial\ (const)$

$M_2 = fspatial\ (reg)$

$M_3 = fspatial\ (const) + frandom\ (const)$

$M_4 = fspatial\ (reg) + frandom\ (reg)$

$M_5 = fspatial\ (const) + fspatial\ (reg)$

$M_6 = fspatial\ (const) + frandom\ (const) + fspatial\ (reg) + frandom\ (reg)$

$M_7 = covaraites + fspatial\ (const)$

$M_8 = covaraites + fspatial\ (reg)$

$M_9 = covaraites + fspatial\ (const) + fspatial\ (reg)$

$M_{10} = covaraites + fspatial\ (const) + frandom\ (const)$

$M_{11} = covaraites + fspatial\ (reg) + frandom\ (reg)$

$M_{12} = covaraites + fspatial\ (const) + frandom\ (const) + fspatial\ (reg) + frandom\ (reg)$

Table 2 : Description of the models for the study.

Model	Variable description
M_0	Fixed effects
M_1	Structured spatial effects at constituency level
M_2	Structured spatial effects at regional level
M_3	Spatial effects (structured + unstructured) at constituency level
M_4	Spatial effects (structured + unstructured) at regional level
M_5	Structured spatial effects at constituency level + regional level
M_6	Spatial effects (structured + unstructured) at constituency + regional level
M_7	Fixed effects + structured spatial effects at constituency level
M_8	Fixed effects + structured spatial effects at regional level
M_9	Fixed effects + structured spatial effects at constituency + regional level
M_{10}	Fixed effects + spatial effects (structured + unstructured) at constituency level
M_{11}	Fixed effects + spatial effects (structured + unstructured) at regional level
M_{12}	Fixed effects + spatial effects (structured + unstructured) at constituency + regional level

Model M_0 is a log-transformed linear logistic regression whereas Models M_1 *to* M_{12} were joint spatial models, and they are all described in detail as follows:

- Model M_0 in this study represents a log-transformed linear logistic regression models which also does not account for any spatial effects, and the hierarchical structure of the data. Geo-additive models that simultaneously account for the spatial dependence in the variables were fitted.

- Models M_1 and M_2 examines only the spatial correlated random effects in the employed and unemployed youths at constituency and regional levels respectively.
- Models M_3 and M_4 were fitted by introducing the uncorrelated spatial random effects to models M_1 and M_2.
- Model M_5 was fitted by taking into consideration the region and constituency correlated spatial effects simultaneously.
- Model M_6 is an improvement to Model M_5. This was fitted by introducing uncorrelated spatial effects.
- Models M_7 and M_8 accounts for the effects of spatially structure covariates at constituency and regional levels respectively.
- Model M_9 was fitted considering the structural spatially structured covariates at constituency and regional structured spatial effects simultaneously.
- Models M_{10} and M_{11} considered the effects of covariates of employed and unemployed youth by incorporating both unstructured and structured random effects at constituency level respectively.
- Model M_{12} considered both the effects of a large area level (regional) together with the effects of small area (constituency) level.

From the models above, DIC was used to select the best models fitted.

3.9 Research ethics

The 2018 LFS was conducted by the NSA under the Statistics Act, 2011 (Act No.9 of 2011), which mandates the agency among others, to constitute the central statistical authority of the country and to collect, produce, analyse, and disseminate official and other statistics in Namibia. Thus, by virtue of this Act, all information collected that could be linked to identified individuals or households were removed from the published microdata sets.

CHAPTER 4: RESULTS

4.1 Introduction

This chapter presents the results of the models fitted and the model selection criterion used. In addition, the parameter estimates for the best model fit was carried out. The interpretation of the results and the subsequent presentation of the odds ratio (OR) of unemployment youth at regional and constituency levels into the maps were undertaken. Section 4.2 presents the descriptive statistics of the analyses; Section 4.3 presents the results of multivariate fitted models. Section 4.4 presents the covariates results (fixed effects). Lastly, Section 4.5 presents the results for the spatial effects of youth unemployment mapped at region and constituency levels.

4.2 Descriptive statistics

Table 3 below shows the estimated unemployment rate at national and regional level by sex as published in the NLFS 2018 by the NSA. The results showed that 46.1 percent of the Namibian youth were unemployed with more than half of the regions recording an unemployment rate that was higher than the national figure. The table also illustrates that unemployment was much higher among youth in rural areas as compared to those in urban areas with 49.1 and 44 percent respectively. Kavango East recorded the highest (62.5 percent) youth unemployment rate with Erongo recording the lowest (36.8 percent). The results further shows that unemployment rate was high amongst female youth as compared to male youth with 48.8 percent and 43.7 percent respectively.

Table 3: The distribution of youth unemployment rate at national and regional levels and at urban/rural by sex

	Overall		Male		Female	
	Count	**%**	**Count**	**%**	**Count**	**%**
Namibia	**265,770**	**46.1**	**125,206**	**43.7**	**140,564**	**48.5**
Place of residence						
Urban	150,506	44.0	70,513	42.6	79,993	45.4
Rural	115,264	49.1	54,693	45.1	60,571	53.3
Region						
!Karas	10,117	44.7	4,769	41.0	5,348	48.6
Erongo	20,800	36.8	11,428	37.0	9,372	36.5
Hardap	8,820	41.9	4,392	37.4	4,428	47.6
Kavango East	21,362	62.5	9,118	62.0	12,244	62.9
Kavango West	8,035	46.8	3,289	44.6	4,746	48.4
Khomas	57,524	43.0	26,435	40.0	31,089	46.0
Kunene	12,777	53.0	6,176	50.8	6,601	55.2
Ohangwena	24,561	51.4	10,240	48.8	14,321	53.4
Omaheke	8,012	46.6	3,527	35.8	4,485	61,0
Omusati	19,187	39.7	8,928	38.1	10,259	41.3
Oshana	22,294	47.2	10,783	48.5	11,511	46.0
Oshikoto	22,439	50.2	12,013	50.4	10,426	49.9
Otjozondjupa	19,007	47.4	9,076	42.5	9,931	52.9
Zambezi	10,835	49.7	5,032	47.8	5,803	51.4

4.3 Results of multivariate models

The following section presents the results of fitted models to the data.

4.3.1 Results of the model's comparison

In both Tables 3 and 4, models M_0 and M_{12} were fitted to assess the difference in the values of Deviance Information Criteria (DIC) in relation to the best model fitted among the geo-additive models (M_0 to M_{12}) that takes into consideration the spatial dependence in the data. For Tables 4.1 and 4.2, sample deviance (Deviance), the effective of parameters (pD), Deviance Information Criteria (DIC) were estimated as proposed by Spiegelharter et al. (2002) and ΔDIC was calculated.

4.3.1.1 Model in respect of male youth unemployment

Table 4: Sample Deviance, pD, DIC and ΔDIC for models in respect of male youth unemployment

Model	Deviance	pD	DIC	ΔDIC
M_0	4945.85	6.77	4959.39	
M_1	5441.23	61.42	5564.08	663.18
M_2	5597.53	10.08	5617.70	716.79
M_3	5402.96	71.16	5545.29	644.38
M_4	5596.73	10.86	5618.45	717.55
M_5	5447.30	58.47	5564.24	663.34
M_6	5404.79	70.66	5546.11	645.20
M_7	4829.92	44.84	4919.61	18.71
M_8	4890.67	16.71	4924.10	23.19
M_9	4837.70	36.43	4910.56	9.65
M_{10}	4783.67	61.58	4906.83	5.93
M_{11}	4889.50	17.39	4924.29	23.39
M_{12}	4789.16	55.87	4900.90	0.00

In Table 4, model M_9 (considers the fixed effects with combined spatial effects of both regions and constituencies) explains the chance of male youths being unemployed better than considering the effects of regions and constituencies separately (with ΔDIC value of 9.65). Model M_9 shows the most improved when unstructured effects are considered in model M_{10} (DIC value of 4906.83).

Based on the results illustrated in Table 4, the best model that explains the chance of being unemployed amongst the male youths was model M_{12} (with the DIC value of 4900.90) which considered the fixed effects together with the unstructured spatial effects at constituency and regional levels. Looking at the difference in DIC values of the other models relative to model M_{12}, it can be concluded that model M_{10} was weakly differentiated.

4.3.1.2 Model in respect of female youth unemployment

Table 5: Deviance, pD, DIC and Δ DIC for models of female youth unemployment.

Model	Deviance	pD	DIC	ΔDIC
M_0	5788.08	7.02	5802.12	
M_1	6155.55	54.19	6263.94	544.46
M_2	6270.55	12.04	6294.63	575.14
M_3	6134.13	63.98	6262.09	542.61
M_4	6269.34	12.30	6293.95	574.47
M_5	6161.48	50.01	6261.49	542.01
M_6	6142.69	57.94	6258.57	539.09
M_7	5607.90	59.31	5726.52	7.04
M_8	5708.63	17.95	5744.53	25.04
M_9	5624.21	50.22	5724.64	5.16
M_{10}	5590.23	67.31	5724.85	5.37
M_{11}	5707.62	18.36	5744.34	24.85
M_{12}	5603.16	58.16	**5719.48**	0.00

With regards to female youth's unemployment, the best model fitted remained the same as for the male youth_which is Model 12 (M_{12} with DIC value of 5719.48) as shown in Table 5. The table show that model M_9 and M_{10} were weakly differentiated as indicated by the DIC differences from the best model. It should also be noted that as much as $M_{7,}$ $M_{9,}$ and M_{10} were equally useful, M_{12} was more proved to be more useful as compared to the earlier stated because of the DIC that is the lowest.

4.4 Fixed effects results

From the best model fitted in Tables 4 and 5, the summaries of the socio-demographic factors considered in this study to explain their influences on the youth unemployment amongst males and female were produced and are presented in Table 6 below.

It is important to point out that Odd Ratios (OR) calculated in case-control studies as the incidence of outcome is not known also if OR>1 indicates increased occurrence and if OR<1 indicates decreased occurrence of an invent. The Confidence Intervals for the OR and the associated p-values are used to determine their significance.

Table 6: Odds ratios and their 95% confidence intervals (CIs) for fixed effects summaries of the best STAR models.

	Male		**Female**	
Covariates	**OR**	**(95% C.I.)**	**OR**	**(95% C.I)**
Is youth Head of Household?				
No	1.00		1.00	
Yes	0.19	(0.16, 0.22)	0.23	(0.20,0.27)
Is youth attending school?				
No	1.00		1.00	
Yes	0.61	(0.47, 0.76)	0.47	(0.37,0.61)
Youth has no formal education				
No	1.00		1.00	
Yes	0.48	(0.38, 0.61)	1.11	(0.85,1.46)
Youth has tertiary education				
No	1.00		1.00	
Yes	0.51	(0.39, 0.66)	0.48	(0.38,0.60)
Youth's marital status				
Never married	1.00		1.00	
Married	0.47	(0.34, 0.67)	0.48	(0.39,0.59)
Type of residence				
Rural	1.00		1.00	
Urban	1.35	(1.10, 1.66)	0.79	(0.65,0.96)

Table 6 above provides OR’s and their respective 95 % CIs. It can be observed that female youths with no formal education are more likely to be unemployed with OR=1.11 (0.85, 1.46). When the same model is generated for male youths, the OR=0.48 (0.38, 0.61) which shows that male youths are with no formal education are less likely to be unemployed. However, the probability of being unemployed is very high among male youths in urban areas with OR=1.35 (1.10, 1.66). The same trend can also be observed amongst female youths in urban areas with OR=0.79 (0.65, 0.96). The table also illustrates that the risk of being unemployed is low for head of household youths.

4.5 Spatial effects

The spatially structured effects of youth unemployment were mapped at regional and constituency levels to give results of smoothed risk estimates from three Bayesian Models (M_1, M_2 and M_5) models fitted. A Conditional Autoregressive (CAR) distribution prior to account for over-dispersion and spatial autocorrelation for both male and female was used in the modelling of the Relative Risk (RR) of youth unemployment.

In all the spatial geographical effects maps presented in this section, the darker colour indicates constituencies with high odds of youth unemployment whereas the lighter colour indicates constituencies with low odds. For the probability maps (Figures 5 through 7), the white coloured areas denote areas with strictly lower uptake of youth unemployment; while the black coloured areas denote areas with strictly high uptake and grey coloured areas shows areas that are of no significance difference.

4.5.1 Spatially structured effects of youth unemployment at regional level and constituency levels by sex

Figures 2 (a) and (b) below shows the spatially structured for youth unemployment at regional level by sex of the respondents. It can be observed that the OR for youth unemployment ranges from 0.8898 to 1.2234 for males and from 0.8652 to 1.2645 for females as generated from M_2. The OR for youth unemployment was found to be highest in the !Karas region when M_2 is applied to both males and female. The OR was lowest in the Hardap, Khomas, Erongo, Omaheke and Zambezi regions amongst males while and the same trend is observed in the Omusati and Kavango West, regions amongst females.

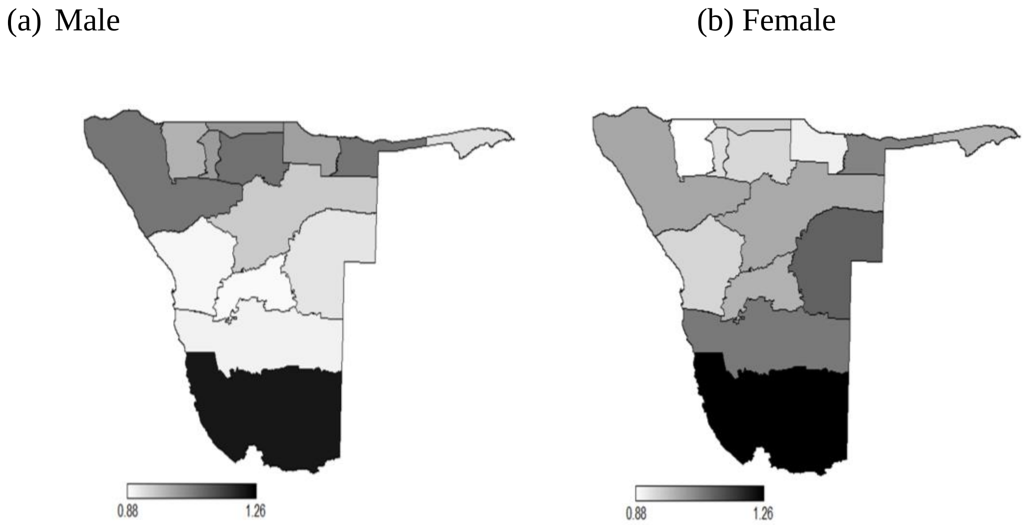


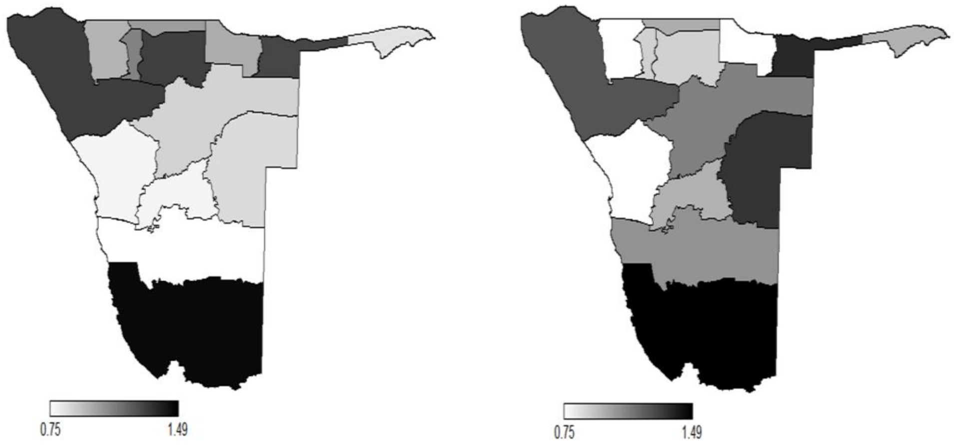
Figure 2 (a) Male and (b) Female: Spatially structured effects of youth unemployment at regional level

As indicated in Figures 3 (a) and (b) below, it is interesting to note that the ORs of most constituencies in the Zambezi region are high at constituency level as compared to the regional level for both male and female youths. This implies that both male and female youth unemployment was likely to be high in all constituencies in the Zambezi region and low in the rest of the constituencies of the rest regions.

(a) Male (b) Female

Figure 3 (a) Male and (b) Female: Spatially structured effects of youth unemployment at constituency level.

Figures 4 (a) and (b) indicates that there was a significant difference in the youth unemployment for males and females. Female youth were more likely to be unemployed when compared to male youth in Zambezi and Omaheke regions followed by the Hardap and Otjozondjupa regions. This trend is however on the reverse side for Oshikoto region where males were more likely to be unemployed than females. Moreover, the Erongo, Kunene, Kavango East and !Karas regions had similar spatial pattern among males and females when it comes to the likelihood of being unemployed. The OR ranged from 0.7559 to 1.4575 for males as compared to that that of females that ranges from 0.6811 to 1.4932.



(a) Male (b) Female

Figure 4 (a) Male and (b) Female: Total (structured and unstructured) spatial geographical effects of youth unemployment at regional level.

Posterior probability maps helps to identify whether or not the observed spatial effects are significant. Figures 5 (a) and (b) below show clear evidence of significant spatial effects of youth unemployment, with higher uptake occurring in !Karas region in both males and females. However, there was a difference in high probabilities amongst males and females in Kunene and Oshikoto regions for the males and Omaheke and Kavango East regions for the females. A lower uptake of youth unemployment occurred in most northern and northern-east regions amongst males (Omusati, Ohangwena, Kavango West, Kavango East, and Zambezi) and more central regions (Hardap, Khomas and Otjozondjupa regions) amongst the females.

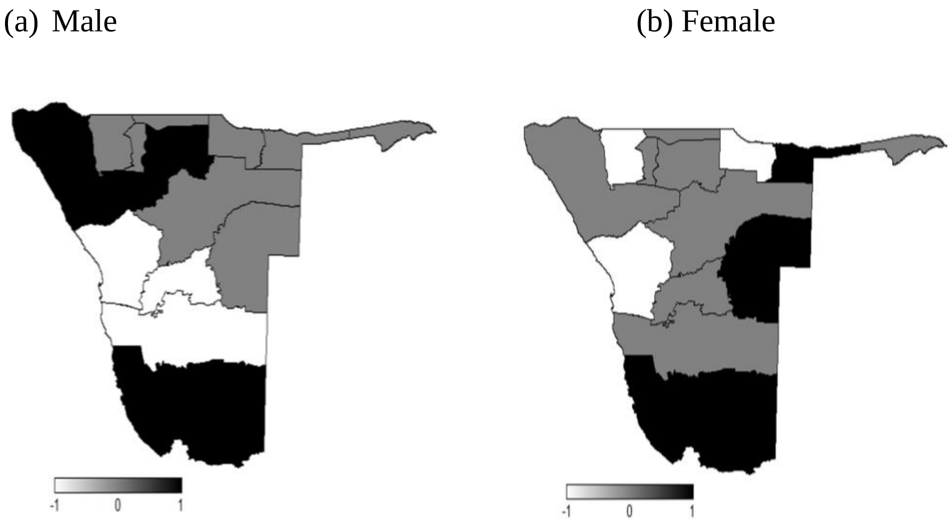


Figure 5 (a) Male and (b) Female: Probability maps of total (structured and unstructured) spatial geographical effects of youth unemployment at regional level at nominal level of 95%.

In Figures 6 (a) and (b) below show that a high rate of youth unemployment was observed in the Daweb, Kabbe South, Keetmanshoop rural and Katima Mulilo rural constituencies amongst males while in females it was observed in the Epembe and Epukiro constituencies. The youth unemployment was however very high in both males and females in Linyanti constituency in the Zambezi region.

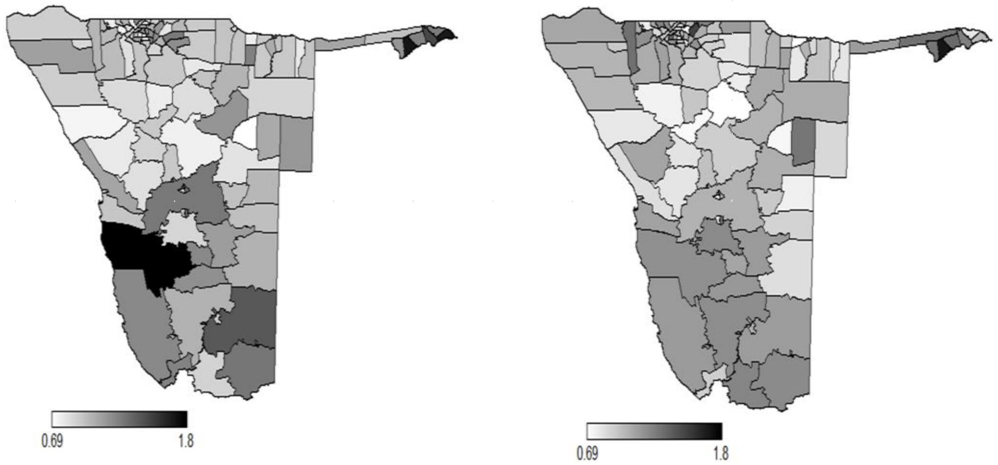

(a) Male (b) Female

Figure 6 (a) Male and (b) Female: Total (structured and unstructured) spatial geographical effects of youth unemployment at constituency level.

Figures 7 (a) and (b) show that there was no significance different in the youth unemployment amongst male and female youths with in almost all the constituencies except in the Daweb constituency in the !Karas region and in the Kabbe South constituency in the Zambezi region inmales.


(a) **Male** **(b) Female**

Figures 7 (a) Male and (b) Female: Probability maps of total (structured and unstructured) spatial geographical effects of youth unemployment at constituency level at nominal level of 95%.

CHAPTER 5: DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

5.1 DISCUSSIONS

Despite Namibia having been labelled with upper-middle income status with a GDP of about N$68.80 per capita in 2020, the World Bank has ranked Namibia as one of the most unequal societies, with a very high GINI coefficient of 0.59 as estimated in the 2015/16 Namibia Household Income and Expenditure Survey. Unemployment is recognised as one of the greatest socio-economic challenges and that is higher among young women is especially high in South Africa, Namibia, and Lesotho" (SADC, 2015). According to the NSA the national unemployment rate was 33.4 percent and for youth aged 15 – 34 years (unit of this study), this rate further increased to about 46.1 percent in 2018. These statistics are known to be higher amongst female as compared to male youths with 48.5 percent and 34.3 percent respectively (NSA, 2018). As a more complete measure, the share of 15 – 34-year-olds not in employment, education or training was 34.9 percent at national level with considerable gender disparities with female accounting for 37.1 percent and males at 32.7 percent.

The youth not in education and not in employment or training (NEET) concept was introduced alongside the unemployment rate in consideration of the fact that where youth do not reflect as employed or part of the labour force is because they are in education or training, this is positive rather than negative. The NEET rate is therefore intended to reflect those youths who are not part of the labour force for reasons other than education and training. And as per the Namibia Labour Force Survey of 2018, NEET was recorded at 34.9 and the rate was particularly higher at the age group of 20-24 year (45.2%) which is worrisome as one would expect such youths to be either in schools or just starting off with their careers. Little is known, nor has there been an attempt to explicitly analyse various aspects of youth unemployment including duration analysis, skills mismatch, and determinants of youth unemployment to inform better policy formulation and decision-making in Namibia. Although the government have noted with concern the increasing number of youths that are jobless in the country, the seriousness of the youth unemployment predicament in Namibia cannot be over emphasized enough as less efforts have been made in understanding the youth unemployment, specifically at lower geographical levels such as constituencies for targeted interventions.

Since no studies in Namibia have investigated youth unemployment at lower geographical level to find the most affected constituencies that are contributing to the high national youth unemployment, the study adopted an application of the Fully Bayesian approach. The hierarchical Bayesian models in BayesX were used to model small area variation in youth unemployment at constituency levels in Namibia using the Namibia Labour Force Survey data of 2018. This study also compares the experiences of male and female youth in the same region regarding unemployment, calculating both shared and distinctive variation in uptake. The constituencies with a large disparity between male and female youths' unemployment represent worrying statistics.

It is also important to note that the advantages of employing HB models for SAE issues have mostly been acknowledged because they consist of having a larger selection of instruments at hand to deal with complex and more realistic models as well as to obtain accurate assessments of variability. Moreover, the Bayesian framework makes it reasonably simple to extend to stochastic sampling variances, and that the non-standard specifications can only be practically implemented using HB models.

To construct a multivariate explanatory model for youth unemployment as a tool for programmatic methods to reduce youth unemployment, the study showed how to use the STAR and joint modelling approaches in looking at the youth unemployment variation at lower levels in this case constituencies.

Logistic regression models were used to help understand the factors associated with the probability of youth being unemployed, building on the existing methodological contributions by Gemperli et al. (2004), Kazembe et al. (2010) and Kneib (2005) although the later were in public health which makes this study the first to do small area analysis specifically on youth unemployment High rates of youth unemployment have been a prominent economic and social issue in the Namibian landscape hence understanding the patterns, structure and causes of youth employment is essential for designing policy interventions. Moreover, evidence-based planning needs well-synthesized data to diametrically direct the interest of decision-makers.

Additionally, a geographical model was used in this study to emphasize both shared and gender-specific distribution patterns for the analysis of youth unemployment among male and female youths. The shared terms can be interpreted as surrogate or observable covariate of a latent variable (Kazembe et al., 2009). In joint analysis, the latent components have a direct interpretation in terms of related risk factors which are either shared by either gender or specific to one of the genders (Mwahi, 2014). Joint modelling not only achieves a considerable improvement in terms of DIC (Tables 4 and 5) but also a gain in precision of the relative risk estimates; this provides strong evidence that a joint modelling approach is a valuable extension of individual analyses (Held et al., 2005). Held et al. (2005) stated that the components in joint analysis are assumed to be independent, and we cannot allow for the possibility of any interaction between the true, but unobserved covariates. It's possible that many risk variables and other outcomes have regional characteristics in common. If so, like it would be the case of gender disparity studies, joint modelling may lead to improved inference by reducing the number of alternative explanations for the observed variability (Best et al., 2005).

Several models were fitted of which the selection procedure was conducted using the succeeding reduction in DIC values and the change in DIC. The best model that was considered in this study was the one with the smallest DIC because like many other model selection criteria it compromises between data and model complexity, and it generalizes AIC which appears as a special case under a vague prior. Based on the results best model that explains the chance of being unemployed amongst the youths was model M_{12} with DIC values of 4900.90 in males and 5719.48 in females which considered the fixed effects together with the unstructured spatial effects at both constituency and regional levels.

With regards to fixed effects, it is evident that education plays a very big role in the probability of youths being unemployed because those with no formal education are at higher risks of being unemployed especially female youths with OR = 1.11 as compared to male youths with OR= 0.48. The study also showed that male youths that live in urban areas are more at risk of being unemployment compared to female youths with OR = 1.35 and OR= 0.79 respectively, and this could be attributed to the issue of more males migrating to urban centres in hopes of finding jobs. Other than explaining spatial variation, this study also estimated the effect of individual level covariates such as head of household status, school attendance, tertiary education attendance and marital status and found that all did not affect the chances of youth being unemployed for both male and female.

With regards to the spatial effects, the youth unemployment was mapped at regional and constituency levels to give results of the smoothed risk estimates from the Bayesian models. As shown in Figure 5 Khomas, Erongo and Hardap regions amongst male youth were observed to have a very low probability of unemployment. When the same model is applied to female youths, low probability of unemployment was observed in Kavango West, Omusati and Erongo regions. However, Figure 6 surprisingly shows that most of the constituencies in some regions have either a low or moderate probability of unemployment except Daweb constituency in the Hardap region, Linyanti and Kabbe South in zambezi region for male youths. The study also shows that both males and females were at high risk of being unemployed mostly in constituencies such as the Kabbe South (in Zambezi region), Keetmanshoop rural (in !Karas region) and Katima Mulilo rural (in Zambezi) for males while for females it was observed in the Epembe (in Ohangwena region) and Epukiro constituencies (in Omaheke region).

This study had several limitations due to the data and the methods used. This study used the 2018 NLFS data. This is the latest national labour force survey conducted in Namibia to date; hence the youth unemployment situation may have changed over the past years given the negative impacts of Covid-19 pandemic that has taken place from 2020 to date. It should also be noted that the study included school going youths as per the ILO standards of computing labour force indicators where employment takes precedence of unemployment which takes precedence over economic inactivity. The unemployment might have been different if these youths were to be removed from the computations.

5.2 CONCLUSIONS AND RECOMMENDATIONS

As stated earlier in Chapter 3, the study adopted a fully Bayesian model by Trevisani & Torelli (2007). As expected, the spatial analysis section shows that there is high risk of youths being unemployed across the constituencies considered in the study specifically those that are situated in rural areas compared to those in urban areas. The evaluation of the uncertainty in the unobserved random effects that are also contributing to the variation in the average probability of regional and constituency unemployment rate through the full posterior inference approach was achieved by incorporating the unobservable random effects of interest into the full hierarchical Bayesian model. The study discovered significant differences between male and female youths because female youths were at a higher risk of being unemployed across constituencies, particularly females with no formal education and those living in cities.

Based on these findings, the regions identified as having an elevated or increased risk of youth unemployment because of having more constituencies with a high variation in youth unemployment may need to be prioritized in terms of policy interventions aimed at job creation. Moreover, regions that were found to have a reduced risk of youth unemployment because of having more constituencies with low unemployment such as Erongo, Kunene, Omusati, Omaheke and Kavango West might requires future control interventions related to monitoring and evaluation.

Model 12 with the smallest DIC for both males (4900) and females (5719.48) were found to be the best model taking into effect the regional and constituency effects. The results further showed that education plays a very big role in the possibility of youths being unemployed because those with no formal education are at higher risks of being unemployed especially female youths with OR = 1.11 compared to male youths with

OR= 0.48. The study also showed that male youths that live in urban areas are more at risks of unemployment compared to female youths with OR = 1.35 and OR= 0.79 respectively.

The study also shows that both males and females were at high risk of being unemployed, mostly in the constituencies of the Zambezi region such as the Kabbe South Keetmanshoop rural (in !Karas region) and Katima Mulilo rural constituencies (in Zambezi region) for males while for females it was observed in the Epembe (in Ohangwena region) and Epukiro constituencies (in Omaheke region).

Policy factors play an important role in determining unemployment rates (Scarpetta, 1996) and further address labour market policies with the aim of facilitating entry into the labour market. This study has revealed that policy recommendations will help address structural changes in the current labour market. This could be a way to lessen the imbalance between demand and supply of young workers if they possess skills that speak to the labour market demands. It is therefore hoped that the results of this study will assist policy makers particularly those involved in employment creation policy planning to develop comprehensive programmes or targeted interventions particularly in constituencies (especially those in Zambezi region and some in !Karas region) that were found to have elevated risk of youth unemployment.

In view of the COVID-19 crisis of 2020 to date, the high rates of informality and vulnerable employment are of concern, since it was mostly the youth who have been hardest hit by the crisis (ILO, 2021). Overall, an ILO study for SADC, which has been endorsed by SADC Ministers for Employment and Labour in April 2021, projected the possible employment loss for Namibia as a result of the crisis to be around 30 percent and that in particular females have been mostly negatively affected, as they are projected to account for 62 percent of all employment losses in the country as a result of the COVID-19 pandemic.

Based on the study findings, it may not be obvious which specification is better when estimating counts, but pertinent quantities should be accurately characterized. In conclusion, future research in the subject of SAE should concentrate on defining appropriate devices for model determination. Additionally, model structure must be complicated to consider more realistic situations such as adding auxiliary information, more complex sampling designs must be experienced, and various real population phenomena must be investigated to magnify differences between models at comparison in simulation studies “small areas at different level of territorial aggregation, small area population with different distributional characteristics, etc.” (Trevisani & Torelli, (2017)).

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd international Symposium on Information Theory*, 267-281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.

Akerkar, R., Martino, S., & Rue, H. (2010). Implementing approximate Bayesian inference for survival analysis using integrated nested Laplace approximations. *Preprint Statistics, Norwegian University of Science and Technology, 1*, 1-38.

Belitz, C., Brezger, A., Kneib, T., & Lang, S. (2009). *BayesX-software for Bayesian Inference in Structure Additive Regression Models*, version 2.01. Retrieved October 10, from http://www.dtst.uni-muechen.de/bayes2.01.

Best, N., Richardson, S., Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res, 14*, 35-59.

Cui, W., & George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138(4), 888-900.

Cavanaugh, J. E., & Neath, A. A. (1999). Generalizing the derivation of the Schwarz information criterion. *Communications in Statistics-Theory and Methods*, 28(1), 49-66.

Chikako, T. U. (2018). Multilevel Modelling of Determinants of Youth Unemployment in Urban Ethiopia: Bayesian Estimation Approach. *International Journal on Data Science and Technology, 4(2),* 67-78.

Economic Report on Africa 2005. (2006). Meeting the Challenges of Unemployment and Poverty in Africa. *INTERNATIONAL JOURNAL OF RURAL STUDIES IJRS, 13(1), 22*.

Emily J. Berg, Wayne A. Fuller, Small Area Prediction of Proportions with Applications to the Canadian Labour Force Survey, *Journal of Survey Statistics and Methodology, Volume 2, Issue 3,* September 2014, Pages 227–256, https://doi.org/10.1093/jssam/smu011

Gary, N., & Nicole, F. (1998). *Factors Influencing Youth Unemployment in Australia: 1980-1994.* Australia: Longitudinal Surveys of Australian Youth.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence, (6),* 721-741.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., & Smith, T. (2004). Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *American Journal of Epidemiology*, 159, 64-72.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, *57*, 97–109.

Hu, S. (2007). Akaike information criterion. *Center for Research in Scientific Computation*.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

International Labour Organization 2005. *Youth Unemployment: From national challenge to a global development plan*: Background paper contributed by the ILO at the G8 Labour and Employment Ministers' Conference, London 10-11 March 2005.

International Labour Organization (2010). *Growth- Employment-Poverty REduction linkages: A framework for Recovery and Accerated Progress towards the Millennium Development goals*. Economic Report on Africa.

International Labour Organization (2019). *Youth Unemployment Estimates*. Retrieved from ILOSTAT: http://www.data.worldbank.com

Kazembe, L.N., Muula, A.S., & Simoonga, C. (2009). Joint spatial modeling of common morbidities of childhood fever and diarrhea in Malawi. *Health and Places,* 15, 165-172.

Kazembe, L.N., & Mpeketula, P.M.G. (2010). Quantifying spatial disparities in neonatal mortality using a structured additive regression model. *PLOS ONE, 5(6),* 11180.

Kneib, T. (2005). *Mixed model based inference in structured additive regression*. PhD thesis, Ludwig-Maximilians-University Munchen.

López-Vizcaíno, E., Lombardía, M. J., & Morales, D. (2015). *Small area estimation of labour force indicators under a multinomial model with correlated time and area effects.* Journal of the Royal Statistical Society. Series A (Statistics in Society), 535-565.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21 (6),* 1087.

Mwahi, E. M. (2014). *Shared-component model with application to mapping gender specific pattern in HIV testing and condom use in Namibia (Doctoral dissertation).*

Molina, I., Saei, A., & M. José Lombardía. (2007). Small Area Estimates of Labour Force Participation under a Multinomial Logit Mixed Model. Journal of the Royal Statistical Society. Series A (Statistics in Society), 170(4), 975–1000. http://www.jstor.org/stable/4623223

Namibia Statistics Agency. (2015). *Analysis of Youth Employment and Unemployment in Namibia 2012-2013 report.* Namibia Statistics Agency. Windhoek

Namibia Statistics Agency. (2019). *The Namibia Labour Force Survey 2018 Report.* Namibia Statistics Agency. Windhoek

Namupala, N. (2016). One that has given you little has your soul consoled”-Unemployed. *Journal for Studies in Humanities and Social Sciences*, 34-48.

Ndjavera, M. (2021, July 26). Youth Unemployment rate expected to reach 50%. *New era newspaper*. https://neweralive.na/posts/youth-unemployment-expected-to-reach-50

Nwuke, K. (2002). *Youth and Employment in Africa: paper prepared for the youth employment summit.* Addis Ababa: ESPD Background Paper.

O'Higgins, N. (2007). Youth unemployment and employment policy: a global perspective. *MPRA Paper.* University Library of Munich, Germany.

Scarpetta, S. (1996). Assessing the role of labour market policies and institutional settings on unemployment: A cross-country study. *OECD Economic studies, 26(1),* 43-98.

Somer, A., & Adam, H. (2021, 10 21). *Investopedia.* Retrieved from Investopedia: https://www.investopedia.com/terms/u/unemployment.asp

Statistics, C. B. (2021, 09 10). *tradingeconomics.* Retrieved from Trading Economics: https://tradingeconomics.com/namibia/youth-unemployment-rate

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology),* 64(4), 583-639.

Teshita, U. C. (2018). Multilevel Modelling of Determinants of Yputh Unemployment in Urban Ethiopia: Bayesian Estimstion Approach. *Internation Journal on Data Science and Technology,* 67-78.

Trevisani, M., & Torelli, N. (2007). Hierarchical Bayesian models for small area estimation with count data. *Dipartimento di Scienze Economiche e Statistiche, Universita Degli studi di Trieste*.

Trevisani, M. and Torelli, N. (2017) A Comparison of Hierarchical Bayesian Models for Small Area Estimation of Counts. *Open Journal of Statistics*, 7, 521-550

ANNEXURE A: BayesX codes

BayesX is the software for Bayesian Inference in Structured Additives Regression Models and is the software used for spatial analysis for this thesis. BayesX is free software that can be downloaded from the site (http://www.stat.uni-muenchen.de/~bayesx/). Version 2.1 of BayesX developed in (07.05.2012) was used for this thesis and all the models fitted 52000 iterations were run with a burn of 2000 for each model. Version 2.1 of BayesX permits Bayesian Inference on Markov Chain Monte Carlo simulation techniques.

Below is the syntax used for fully Bayesian for the study:

```
****************Loading datasets************************************

>dataset Female

>Female.infile using C:\Users\Linda Vute\Desktop\200308122\Female.txt

>dataset Male

>Male.infile using C:\Users\Linda Vute\Desktop\200308122\Male.txt

*********************Loading Maps**********************************

>map regmap

>regmap.infile using C:\Users\Linda Vute\Desktop\200308122\namReg.csv

>map constimap

>constimap.infile using C:\Users\Linda Vute\Desktop\200308122\namConst.csv

***********************unemployment models*****************

>bayesreg m0
```

```
>m0.regress uempl=head+attend+noedu+tedu+mastat+urban, family=binomial predict using Female

>bayesreg m1

>m1.regress uempl=const(spatial, map=constimap), family=binomial predict using Female

>bayesreg m2

>m2.regress uempl=reg(spatial, map=regmap), family=binomial predict using Female

>bayesreg m3

>m3.regress uempl=const(spatial, map=constimap)+const(random), family=binomial predict using Female

>bayesreg m4

>m4.regress uempl=reg(spatial, map=regmap)+reg(random), family=binomial predict using Female

>bayesreg m5

>m5.regress uempl=const(spatial, map=constimap)+ reg(spatial,map=regmap), family=binomial predict using Female

>bayesreg m6

>m6.regress uempl=const(spatial, map=constimap)+const(random)+ reg(spatial,map=regmap)+reg(random), family=binomial predict using Female

>bayesreg m7
```

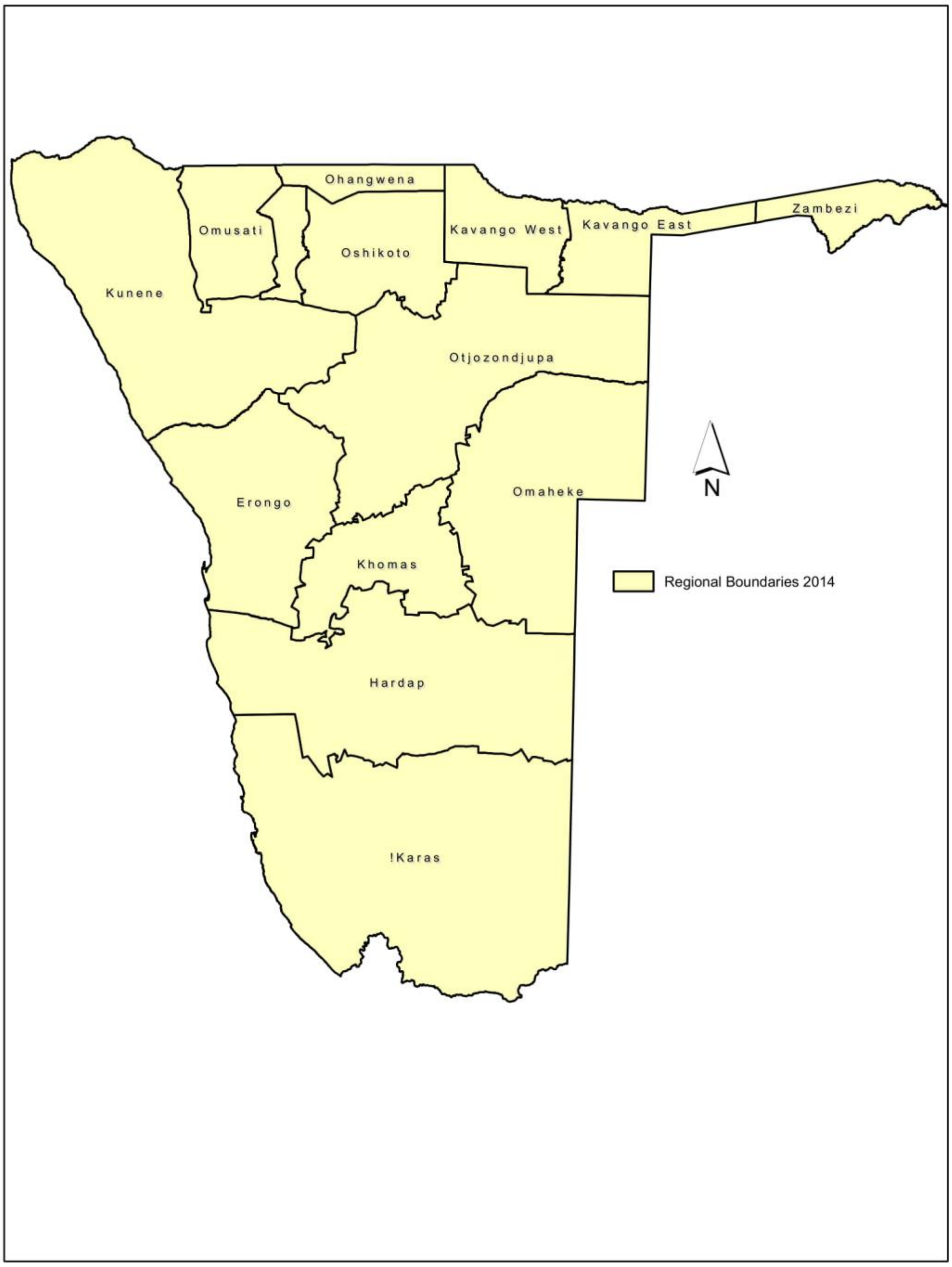
```
>m7.regress uempl=head+attend+noedu+tedu+mastat+urban + const(spatial, map=constimap), family=binomial predict using Female

>bayesreg m8

>m8.regress uempl=head+attend+noedu+tedu+mastat+urban + reg(spatial, map=regmap), family=binomial predict using Female

>bayesreg m9

>m9.regress uempl=head+attend+noedu+tedu+mastat+urban + const(spatial, map=constimap)+ reg(spatial, map=regmap), family=binomial predict using Female

>bayesreg m10

>m10.regress uempl = head + attend + noedu + tedu + mastat +urban + const(spatial, map=constimap)+ const(random), family=binomial predict using Female

>bayesreg m11

>m11.regress uempl=head+attend+noedu+tedu+mastat+urban + reg(spatial, map=regmap)+reg(random), family=binomial predict using Female

>bayesreg m12

>m12.regress uempl=head+attend+noedu+tedu+mastat+urban + const(spatial, map=constimap)+ const(random)+ reg(spatial, map=regmap)+reg(random), family=binomial predict using Female
```

ANNEXURE B: R codes

```
library("BayesX")
> m <- read.bnd("C:/Users/lshitenga/Documents/200308122/namConst.csv")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_f_const_spatial.txt", map = m, plotvar = "pmean", regionvar = "const",limits=c(0.91,1.31), swapcolors=TRUE, cols="grey") >drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_f_const_spatial.txt", map = m, plotvar = "pmean", regionvar = "const",limits=c(0.91,1.31), swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_const_spatialtotal.txt", map = m, plotvar = "pmean", regionvar = "const",limits=c(0.69,1.80), swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_const_spatialtotal.txt", map = m, plotvar = "pmean", regionvar = "const",limits=c(0.69,1.80), swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_const_spatialtotal_probability.txt", map = m, plotvar = "pcat95", regionvar = "const", legend = T, swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_const_spatialtotal_probability.txt", map = m, plotvar = "pcat95", regionvar = "const", limits=c(0,1), swapcolors=TRUE, cols="grey")
> m<-read.bnd("C:/Users/lshitenga/Documents/200308122/namReg.csv")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_f_reg_spatial.txt", map = m, plotvar = "pmean", regionvar = "reg", limits=c(0.88,1.26), swapcolors=TRUE, cols="grey")
```

```
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_f
_reg_spatial.txt", map = m, plotvar = "pmean", regionvar = "reg", limits=c(0.88,1.26
), swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_reg_s
patialtotal.txt", map = m, plotvar = "pmean", regionvar = "reg", limits=c(0.75,1.49),
swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_r
eg_spatialtotal.txt", map = m, plotvar = "pmean", regionvar = "reg", limits=c(0.75,1.
49), swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Male_Best/Male_reg_s
patialtotal_probability.txt", map = m, plotvar = "pcat95", regionvar = "reg", legend =
T, swapcolors=TRUE, cols="grey")
>drawmap(data="C:/Users/lshitenga/Documents/200308122/Female_Best/Female_r
eg_spatialtotal_probability.txt", map = m, plotvar = "pcat95", regionvar = "reg", lege
nd = T, swapcolors=TRUE, cols="grey")
```

APPENDIX A: Regional boundary map of Namibia



APPENDIX C: Constituency boundary map of Namibia