Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXVIII

IMPLEMENTING HIERARCHICAL BAYESIAN MODEL TO FERTILITY DATA: THE CASE OF ETHIOPIA

Direttore della Scuola: Ch.mo Prof.ssa Monica Chiogna

Supervisore: Ch.mo Prof.ssa Stefano Mazzuco

Dottorando: Haftu Gebrehiwot

February 1, 2016

Dedicated to my families, particularly to Mom!!

Abstract(English)

Background: Ethiopia is a country with 9 ethnically-based administrative regions and 2 city administrations, often cited, among other things, with high fertility rates and rapid population growth rate. Despite the country's effort in their reduction, they still remain high, especially at regional-level. To this end, the study of fertility in Ethiopia, particularly on its regions, where fertility variation and its repercussion are at boiling point, is paramount important. An easy way of finding different characteristics of a fertility distribution is to build a suitable model of fertility pattern through different mathematical curves. ASFR is worthwhile in this regard. In general, the age-specific fertility pattern is said to have a typical shape common to all human populations through years though many countries some from Africa has already started showing a deviation from this classical bell shaped curve. Some of existing models are therefore inadequate to describe patterns of many of the African countries including Ethiopia. In order to describe this shape (ASF curve), a number of parametric and non-parametric functions have been exploited in the developed world though fitting these models to curves of Africa in general and that of Ethiopian in particular data has not been undertaken yet. To accurately model fertility patterns in Ethiopia, a new mathematical model that is both easily used, and provides good fit for the data is required.

Objective: The principal goals of this thesis are therefore fourfold: *(1).* to examine the pattern of ASFRs at country and regional level,in Ethiopia; *(2).* to propose a model that best captures various shapes of ASFRs at both country and regional level, and then compare the performance of the model with some existing ones; *(3).* to fit the proposed model using Hierarchical Bayesian techniques and show that this method is flexible enough for local estimates vis-á-vis traditional formula, where the estimates might be very imprecise, due to low sample size; and *(4).* to compare the resulting estimates obtained with the non-hierarchical procedures, such as Bayesian and Maximum likelihood counterparts.

Methodology: In this study, we proposed a four parametric parametric model, Skew Normal model, to fit the fertility schedules, and showed that it is flexible enough in capturing fertility patterns shown at country level and most regions of Ethiopia. In order to determine the performance of this proposed model, we conducted a preliminary analysis along with ten other commonly used parametric and non-parametric models in demographic literature, namely: Quadratic Spline function, Cubic Splines, Coale-Trussell function, Beta, Gamma, Hadwiger distribution, Polynomial models, the Adjusted Error Model, Gompertz curve, Skew Normal, and Peristera & Kostaki Model. The criterion followed in fitting these models was *Nonlinear Regression with nonlinear least squares* (`nls`) estimation. We used **Akaike Information Criterion (AIC)** as model selecction criterion. For many demographers, however, estimating regional-specific ASFR model and the associated uncertainty introduced due those factors can be difficult, especially in a

situation where we have extremely varying sample size among different regions. Recently, it has been proposed that Hierarchical procedures might provide more reliable parameter estimates than Non-Hierarchical procedures, such as complete pooling and independence to make local/regional-level analyses. In this study, a Hierarchical Bayesian procedure was, therefore, formulated to explore the posterior distribution of model parameters (for generation of *region-specific ASFR point estimates* and *uncertainty bound*). Besides, other non-hierarchical approaches, namely Bayesian and the maximum likelihood methods, were also instrumented to estimate parameters and compare the result obtained using these approaches with Hierarchical Bayesian counterparts. Gibbs sampling along with Metropolis-Hastings argorithm in R (Development Core Team, 2005) was applied to draw the posterior samples for each parameter. Data augmentation method was also implemented to ease the sampling process. Sensitivity analysis, convergence diagnosis and model checking were also thoroughly conducted to ensure how robust our results are. In all cases, non-informative prior distributions for all regional vectors (parameters) were used in order to real the lack of knowledge about these random variables.

Result: The results obtained from this preliminary analysis testified that the values of the Akaike Information Criterion(AIC) for the proposed model, Skew Normal (SN), is lowest: *in the capital, Addis Ababa, Dire Dawa, Harari, Affar, Gambela, Benshangul-Gumuz*, and *country level data* as well. On the contrary, its value was also higher some of the models and lower the rest on the remain regions, namely: *Tigray, Oromiya, Amhara, Somali* and *SNNP*. This tells us that the proposed model was able to capturing the pattern of fertility at the empirical fertility data of Ethiopia and its regions better than the other existing models considered in 6 of the 11 regions. The result from the HBA indicates that most of the posterior means were much closer to the true fixed fertility values. They were also more precise and have lower uncertainty with narrower credible interval vis-á-vis the other approaches, ML and Bayesian estimate analogues.

Conclusion: From the preliminary analysis, it can be concluded that the proposed model was better to capture ASFR pattern at national level and its regions than the other existing common models considered. Following this result, we conducted inference and prediction on the model parameters using these three approaches: HBA, BA and ML methods. The overall result suggested several points. One such is that HBA was the best approach to implement for such a data as it gave more consistent, precise (the low uncertainty) than the other approaches. Generally, both ML method and Bayesian method can be used to analyze our model, but they can be applicable to different conditions. ML method can be applied when precise values of model parameters have been known, large sample size can be obtained in the test; and similarly, Bayesian method can be applied when uncertainties on the model parameters exist, prior knowledge on the model parameters are available, and few data is available in the study.

Sommario(Italian)

Background: L'Etiopia è una nazione divisa in 9 regioni amministrative (definite su base etnica) e due città. Si tratta di una nazione citata spesso come esempio di alta fecondità e rapida crescita demografica. Nonostante gli sforzi del governo, fecondità e cresita della popolazione rimangono elevati, specialmente a livello regionale. Pertanto, lo studio della fecondità in Etiopia e nelle sue regioni – caraterizzate da un'alta variabilità – è di vitale importanza. Un modo semplice di rilevare le diverse caratteristiche della distribuzione della fecondità è quello di costruire in modello adatto, specificando diverse funzioni matematiche. In questo senso, vale la pena concentrarsi sui tassi specifici di fecondità, i quali mostrano una precisa forma comune a tutte le popolazioni. Tuttavia, molti paesi mostrano una "simmetrizzazione" che molti modelli non riescono a cogliere adeguatamente. Pertanto, per cogliere questa la forma dei tassi specifici, sono stati utilizzati alcuni modelli parametrici ma l'uso di tali modelli è ancora molto limitato in Africa ed in Etiopia in particolare.

Obiettivo: In questo lavoro si utilizza un nuovo modello per modellare la fecondità in Etiopia con quattro obiettivi specifici: *(1).* esaminare la forma dei tassi specifici per età dell'Etiopia a livello nazionale e regionale; *(2).* proporre un modello che colga al meglio le varie forme dei tassi specifici sia a livello nazionale che regionale. La performance del modello proposto verrà confrontata con quella di altri modelli esistenti; *(3).* adattare la funzione di fecondità proposta attraverso un modello gerarchico Bayesiano e mostrare che tale modello è sufficientemente flessibile per stimare la fecondità delle singole regioni – dove le stime possono essere imprecise a causa di una bassa numerosità campionaria; *(4).* confrontare le stime ottenute con quelle fornite da metodi non gerarchici (massima verosimiglianza o Bayesiana semplice)

Metodologia: In questo studio, proponiamo un modello a 4 parametri, la Normale Asimmetrica, per modellare i tassi specifici di fecondità. Si mostra che questo modello è sufficientemente flessibile per cogliere adeguatamewnte le forme dei tassi specifici a livello sia nazionale che regionale. Per valutare la performance del modello, si è condotta un'analisi preliminare confrontandolo con altri dieci modelli parametrici e non parametrici usati nella letteratura demografica: la funzione splie quadratica, la Cubic-Spline, i modelli di Coale e Trussel, Beta, Gamma, Hadwiger, polinomiale, Gompertz, Peristera-Kostaki e l'Adjustment Error Model. I modelli sono stati stimati usando i minimi quadrati non lineari (nls) e il Criterio d'Informazione di Akaike viene usato per determinarne la performance. Tuttavia, la stima per le singole regioni pu'o risultare difficile in situazioni dove abbiamo un'alta variabilità della numerosità campionaria. Si propoone, quindi di usare procedure gerarchiche che permettono di ottenere stime più affidabili rispetto ai modelli non gerarchici ("pooling" completo o "unpooling") per l'analisi a livello regionale. In questo studia si formula un modello Bayesiano gerarchico ottenendo la distribuzione a posteriori dei parametri

per i tassi di fecnodità specifici a livello regionale e relativa stima dell'incertezza. Altri metodi non gerarchici (Bayesiano semplice e massima verosimiglianza) vengono anch'essi usati per confronto. Gli algoritmi Gibbs Sampling e Metropolis-Hastings vengono usati per campionare dalla distribuzione a posteriori di ogni parametro. Anche il metodo del "Data Augmentation" viene utilizzato per ottenere le stime. La robustezza dei risultati viene controllata attraverso un'analisi di sensibilità e l'opportuna diagnostica della convergenza degli algoritmi viene riportata nel testo. In tutti i casi, si sono usate distribuzioni a priori non-informative.
Risultati: I risutlati ottenuti dall'analisi preliminare mostrano che il modello Skew Normal ha il più basso AIC nelle regioni Addis Ababa, Dire Dawa, Harari, Affar, Gambela, Benshangul-Gumuz e anche per le stime nazionali. Nelle altre regioni (Tigray, Oromiya, Amhara, Somali e SNNP) il modello Skew Normal non risulta il milgiore, ma comunque mostra un buon adattamento ai dati. Dunque, il modello Skew Normal risulta il migliore in 6 regioni su 11 e sui tassi specifici di tutto il paese.
Conclusioni: Dunque, il modello Skew Normal risulta globalmente il migliore. Da wuesto risultato iniziale, si è partiti per costruire i modelli Gerachico Bayesiano, Bayesiano semplice e di massima verosimiglianza. Il risultato del confronto tra questi tre approcci è che il modello gerarchico fornisce stime più precise rispetto agli altri.

Acknowledgment

I have spent an exciting and enjoyable three years learning period at University of Padova, Italy, largely due to the amazing people that I had the good fortune to associate with. The research described in this thesis is the culmination of years of hard work, and would not have been possible without the help of a number of people. I want to take this opportunity to thank them for their help and support.

Above all, I am running out of words and phrases to explain my profound appreciation and heartfelt gratitude to Professor Stefano Mazzuco, my supervisor, for his closer follow-up, continuous encouragement, invaluable guidance, extraordinary patience and constructive comments and suggestions from which I have immeasurably benefited in sharpening my understanding, and broadening my interest predominantly, on the area I have studied. I have enjoyed and learnt a lot of lessons from him over those couple of years including patience, can-do attitude, sincerity, and many more. He was always keeping his door open 24 hours for all my challenges I encounter in the course of doing the thesis. I have no doubt that without his unwavering support and continuous encouragement this thesis would not have been finalized in its present structure. However, all the imperfections and shortcomings that may be inherent in this thesis are entirely mine.

Second, I extend my heartfelt gratitude to all my school professors, especially Professor Burno Scarp and Antone Canale who helped me with many materials and showed me directions in the course of doing the thesis. My deepest gratitude and heart-felt honor also goes towards Prof. Nicola Sartori and Prof. Monica Chaogna for your immeasurable support and unending care in making my study possible and ease. It is also a great pleasure to acknowledge my deepest thanks and sincere appreciation to Patrizia, the school secretary, for her kind endless help until this work came to existence.

I am also extremely indebted to my colleagues and friends: Giovanni, Giulio, Roberta, tDaniele, Ronaldo and Ismail for their unconditional love and support playing since day one of my arrival to Padova, Italy. Thank you all for the laughs and challenges we shared over the years!!

Last but not least, special gratitude to my families, specially to Mom, who inspired me with her moral and spiritual support to continue my work with determination.

Haftu Gebrehiwot

University of Padova
February 01, 2016

Contents

List of Figures

List of Tables

1 INTRODUCTION

Demography[1] is the scientific study of human population and its dynamics (*viz.*: their composition, distributions, densities, growth and other characteristics as well as the causes and consequences of changes in these factors). The five "*demographic processes*" that often determine the changes in population size (*growth* or *decline*), composition and distribution in space are: *Fertility, Mortality, Migration, Social Mobility* and *Marriage* (Sharma, 2004), as shown in Figure 1.1.

Figure 1.1: The five demographic processes:*Fertility, Mortality, Migration, Social Mobility* and *Marriage*

Fertility is, therefore, one of among these demographic processes, playing a key major role in changing population in size, composition and structure over time (Melake, 2005). It is defined as the reproductive performance of an individual, a couple, a group or a population; or the natural human capability of producing offspring. A fertility rate is a measure of the average number of children a woman will have during her reproductive ages/years (*aka*: childbearing ages)[2] There are various measures of fertility. The main ones are: *Crude Birth Rate (CBR)*, *General Fertility Rate (GFR)*, *Age Specific Fertility Rate (ASFR)* and *Total Fertility Rate (TFR)*. Analysis of fertility is not only required to understand the

[1] *Demography*: the term "demography" is derived from two *Greek Words*, namely **Demos** means ***population***; and **Graphics** means ***to draw***

[2] Reproductive or childbearing age/year/: *is the reproductive period or span of a woman, assumed for statistical purposes to be 15-49 years of age.*

demographic nature of a given population, but also is needed as it affects public policy, budgeting for education and health systems, and the likes. Government officials, policy makers, and academics are concerned about fertility because it can influence the overall development planning process.

Sub-Saharan Africa is still the most cited example of a continent in which those demographic elements as a whole and fertility in particular have hit a plateau. Compared to many developing nations in Asia and Latin America, fertility rate in those countries is higher, and its decline is later & slower yet, mainly due to social, economical, psychological and cultural factors, favoring high fertility rate, and discouraging policy makers from promoting strong family planning programmes (Caldwell et al., 1992). Ethiopia, where the present thesis focuses on, is not an exception to this rhetoric and phenomenon.

1.1 Overview:Fertility in Ethiopia

To highlight, Ethiopia, a country in East Africa, with 9 ethnically-based administrative regions (a.k.a: National Regional States) (viz.: Tigray, Afar, Amhara, Oromia, Gambella, SNNP, Somali, Benishangul-Gumaz, Harari) and 2 city administrations (viz.: Addis Ababa, Dire Dawa), as shown in Figure 1.2 below, is one of the developing nations, attributed, among other things, with rapid population growth and high fertility[3] rates. The country's population in the year 2014 was estimated at 96.6 million, increased by almost five-fold in just half a century, from 19.2 million in 1950; and its population growth rate was also 2.9 percent, which is a growth of nearly 2.8 million people the country adds every year to the total. This alarming figure in population size ranked the country as second most populous nation in Africa following Nigera and 14^{th} in the world (Teklu and Gebreselassie, 2013; The World Factbook, 2014). Similarly, the fertility rate in the stated year was 5.23 children per woman, which is among the highest in the planet. Rapid population growth and high fertility rates have a lot of important repercussion in the country. Study indicates that the rapid population growth and high fertility rates in Ethiopia are one of the major sticking points and challenges in achieving important national goals, such as :-food self sufficiency, universal primary education and accessibility of health services, increasing employment opportunities and reducing underemployment (Farina et al., 2001).

[3] High fertility is defined as a total fertility rate (TFR) of 5.0 or higher (World Population Perspective,2005). The TFR represents the average lifetime births per woman implied by the age-specific fertility rates prevailing in one historical period.

Figure 1.2: Ethiopia & its regions

Region	Women in the study	TFR
Tigray	1728	4.8
Affar	1291	5.5
Amhara	2087	4.1
Oromiya	2135	5.1
Somali	914	7.4
Benishangul-Gumuz	1259	5.1
SNNP	2034	5.1
Gambela	1130	3.1
Harari	1101	3.4
Addis Ababa	1741	1.6
Dire Dawa	1095	2.8

Table 1.1: Fertility across the regions, 2011 EDHS

Earlier studies point out that there are various reasons, which make fertility rate in Ethiopia remain high to-date, including but not limited to: early and universal marriage[4](Gezahegn, 2011), the high social and economic value attached to children, the low level of infertility [5](Gezahegn, 2011; Bertrand et al., 2000), cultural and traditional barriers to effectively utilize modern birth control methods[6](Bertrand et al., 2000), low socio-economic development, deeply-ingrained cultural values for large family (Kinfu, 2001; Machera, 1997), as well as poor health service, low economic status and autonomy of women, strong kinship networks, high economic & social values attached to children, the desire for more children and extremely low contraceptive/practice, high child mortality,and women's limited achievements in the sphere of educational & employment opportunities (Machera, 1997; Bertrand et al., 2000).

In Africa, families often prefer large number of children since they are considered as an economic asset rather than liabilities. In rural areas, parents want to have large number of children to get assistance in farming activities (Bairagi, 2001) and emotional as well as economic support during old ages (Fapohunda and Todaro, 1988). According to Caldwell (1982), the economic importance of children is over lifetime. African children do not only provide support during childhood and adolescent ages but also beyond these ages. More psycho-social and economic support is expected when parents are getting older. Old-age security is one of the major motivational forces for having as many children as possible in Africa. In traditional societies, children are also expected to strengthen the extent of kin

[4] The median age at first marriage is less than 17.1 years in 2011 and nearly all women are married before they reach age 35.

[5] Infertility is much lower in Ethiopia compared to other countries in Africa.

[6] Only 28.6% of currently married non-pregnant women aged 15-49 reported as users of contraceptives in 2011.

relations, which implies not only economic benefits but also physical protection. Getting larger in number is tantamount to strength in physical security. These scenarios are also true in Ethiopia. Like many countries in sub-Saharan Africa, traditional norms and values in Ethiopia are in favor of high fertility. Having many children is considered as a virtue and respect of God in a number of Ethiopian rural communities (Desta and Seyoum, 1998). However, in the last few years, fertility in Ethiopia, like many African countries, has shown a declining trend at the national level though this decline is insignificant and does not help the country escape from high fertility zone. According to the 1990 National Family and Fertility Survey, the Total Fertility Rate (TFR) of Ethiopia was 6.6 children per woman. However, this rate declined to 5.5 in 2000, to 5.4 in 2005 (Teller and Hailemariam, 2011) and further to 4.8 in 2011 (CSA and ICF International, 2012), which is a drop of almost a child per woman over two decades of time.

As seen in Table 1.1, there are also clear and remarkable disparities in fertility levels among the administrative regions, ranging from 1.6 children per woman (below the replacement level) in Addis Ababa, where annual growth rate was 2.1, to 7.4 children per woman in Somali region, which had had annual growth rate of 2.6. Fertility level was also observed to be higher than the national average in Oromiya, Benshangul-Gumuz, Affar and SNNP regions; and lower than the national average in the remaining six regions (CSA and ICF International, 2012). Given the fact that Ethiopia follows an ethnic based federal system, the variation in total fertility rate across different regions of the country could simply be a reflection of differences in cultural values and norms affecting fertility for most of them, with the exception of Addis Ababa and Dire Dawa City Administrations, where the population is of mixed type as they belong to different ethnic groups (ESPS, 2008).

It is palpable fact that high fertility rate affects the health status of mothers and the survival chance of their offsprings. Its effects even go to the extent of affecting the socio-economic development of a given country if proper care and action are not taken. For instance, it detracts from human capital investment, slows economic growth, exacerbates environmental threats, and many more (Mulugeta Eyasu, 2015; Casterline and Lazarus, 2010). Cognizant of this fact, the government of Ethiopia, like most in the developing countries, has been found perpetrating several concerted efforts to reduce the prevailing high fertility rate to a level the country sets as a target to achieve the Millennium Development Goals (MGDs), which is 4.0 children per woman by 2015.

So as to clarify the reason behind the prevailing high fertility , the recent decline in it, its variation & pattern observed at national and regional level, the important repercussion that this high fertility has at national level and in regions, and consequently, design effective family planning program to achieve the goal Ethiopia sets in 2015 and/or onward, local-level analysis of fertility, such as modeling fertility, estimation its parameters, prediction,and so on, is increasingly important so that one can effectively formulate flexible region-specific strategies; and make planning, monitoring, and policies, which could be in connection to education, reproductive health, child health services, population stabilization, human development and other related issues at regional levels, such as to decide where to build a new nursery-school, health center, health posts; where to increase midwifes or obstetrics departments in hospitals; or which kind of services can be offered to mothers and families; etc.

Until recently, many studies have been carried out particularly in developed countries and developing countries outside of Africa to model and look at fertility variations and explore further to understand factors and indicators influencing fertility level, but many of them were limited to a country level (Drèze and Murthi, 2000). Nevertheless, much less attention has been given to models for local or regional fertility curves, where we expect a wider variety of patterns than for country level. Given this variety of possible fertility patterns observed across region, in this thesis, we will propose a model that best captures the different age-specific fertility patterns of Ethiopia at country and regional levels. However, since there are regions in the study possessing lower number of observations like Somali, Benshangul-Gumuz, etc, thus, local estimates might be very imprecise, due to their low sample size. In this study, we, therefore, make use of a hierarchical Bayesian alternative to the fertility formulas so as to show that the hierarchical model outperforms and is flexible enough for most fertility pattern at different region of Ethiopia and in estimating the parameters in the proposed model.

1.2 Summary and Main contribution

Needless to say, local-level fertility analysis in Ethiopia is paramount important for it will help to effectively design flexible region-specific strategies or programs, which might be worthwhile in implementation of family planning programs, and other socio-economic policies down at regional levels. The modeling of fertility patterns is one of the essential methods/ approaches researchers often use to understand the demographic nature of a given population, and thereby, make budgeting, planning, and monitoring policy decisions at different levels, national and regional-levels.

Chapter 2 deals with model development for the fertility patterns shown at both country and regional levels. Putting it differently, in this chapter we first took advantage of real data set from 2011 Demographic and Health Survey (DHS) of Ethiopia, from which we extracted one-year age specific fertility rate (ASFR) and examined its pattern at both national and regional levels. Having thoroughly and attentively assessed all the patterns, we, then, proposed a four parametric Skew Normal model (Mazzuco and Scarpa, 2011) to fit the fertility schedules, and eventually showed that it is flexible enough in capturing these fertility patterns shown at both country and regional levels of Ethiopia. So as to determine the performance of the proposed model, a preliminary analysis of fitting the model along with ten other commonly used parametric and non-parametric models has been carried out.

Chapter 3 is primarily devoted to Bayesian inference to make local-level inference and prediction on the fertility parameters of *Skew Normal model* (Mazzuco and Scarpa, 2015) proposed this far, in Chapter 2, for fitting the Age-specific fertility rate data at country and regional levels in Ethiopian. As outlined in **Chapter 2**, this study uses data amassed from different regions, possessing varying sample sample sizes, some large and others small number of observations (childbearing mothers), as expounded in Table 1.1 or displayed in Figure 3.1. Moreover, study by Griffiths et al. (1987) suggests that inference based on the conventional classical statistics via MLE, often leads to significant bias and hence, is not accurate in a situation where the subject under scrutiny has small sample size. Consequently, in this chapter Bayesian inference is predominantly adopted to make inference and prediction on model parameters. Despite the issue of sample size in some regions, maximum likelihood method was also conducted to estimate the parameters of the model and compare the resulting values with Bayesian counterparts.
Besides, as another new contribution of this chapter, we first developed a fertility model based on whether or not each mother involved in the study gives rise a birth at a specific age, *i.e.*, taking mother's birth status (*Have birth*, *No birth*) during a specific age as major classification criterion. Given the result obtained in **Chapter 1**, we assume mother's birth status as a *Bernoulli random variable* having a probability distribution of *the proposed model*. Further, we modeled the number of births at this particular age, which was a binary data with non-negative valued random variable, as Binomial distribution having a probability, the value of proposed model at that age, and number of mother at that particular age as model parameters and eventually, estimated this model using a Bayesian approach using non-informative priors for the model parameters. Nevertheless, one stumbling block encounter in using this methodology was computational intractability. That is, the joint posterior distributions was non-linear, and too complex & intractable to easily drive the full conditional in standard/closed form. *Data Augmentation strategy* (*latent variable method*) had, hence, been instrumented as possible remedy in this respect. To wrap up, in this chapter and, of course, the next as well, a skew-normal (Mazzuco and Scarpa, 2015) latent variable methodology has been implemented in formulating our Bayesian model developed from Binomially distributed fertility (birth) data so as to overcome this computational plight. Hence, the fertility data from each of the 11 regions are modeled using a Bayesian model with non-informative prior distributions, *i.e.*, the priors are constructed by assuming

there is no information available about the process apart from the data. Furthermore, the results and techniques developed in this section was used directly in developing and analyzing the subsequent hierarchical Bayesian models. One of the major sticking points with

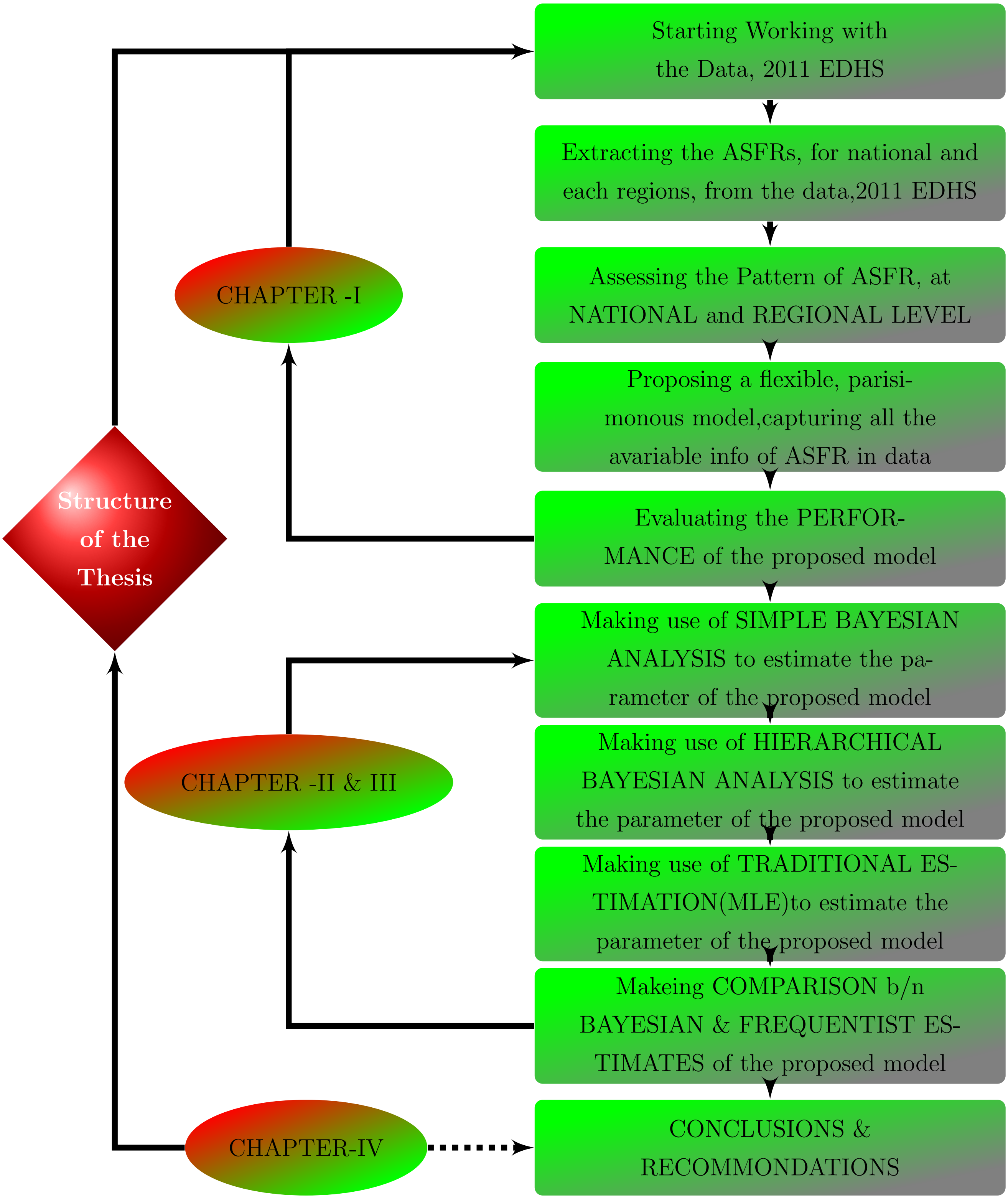


Figure 1.3: Thesis Structure

Bayesian Analysis/method utilized in **Chapter 3** is that those region-specific ASFR parameters had been imprecise estimates, particularly in those regions which contain smaller observations(mothers of age 15-49), namely Somali region, Affar region and SNNP. Albert and Chib (1993); Nilsson et al. (2011) have suggested that Hierarchical procedures might provide more reliable parameter estimates than Non-Hierarchical procedures, the mainstream Bayesian and maximum likelihood methods, especially with small sized regions. Therefore, in this chapter, **Chapter 4**, Hierarchical Bayesian method was formulated and briefly discussed to explore the posterior distribution of model parameters. Besides, other non-hierarchical approaches, namely Bayesian and the maximum likelihood methods, were also instrumented to estimate parameters and compare the result obtained using these approaches with Hierarchical Bayesian counterparts.

Last but not certainly the least, **Chapter 5** combines the thesis together by providing summary conclusions of previous chapters and puts forward some policy recommendations to be in place at both national and regional level in the country. The structural flow of the contribution is described diagrammatically in Figure 1.3

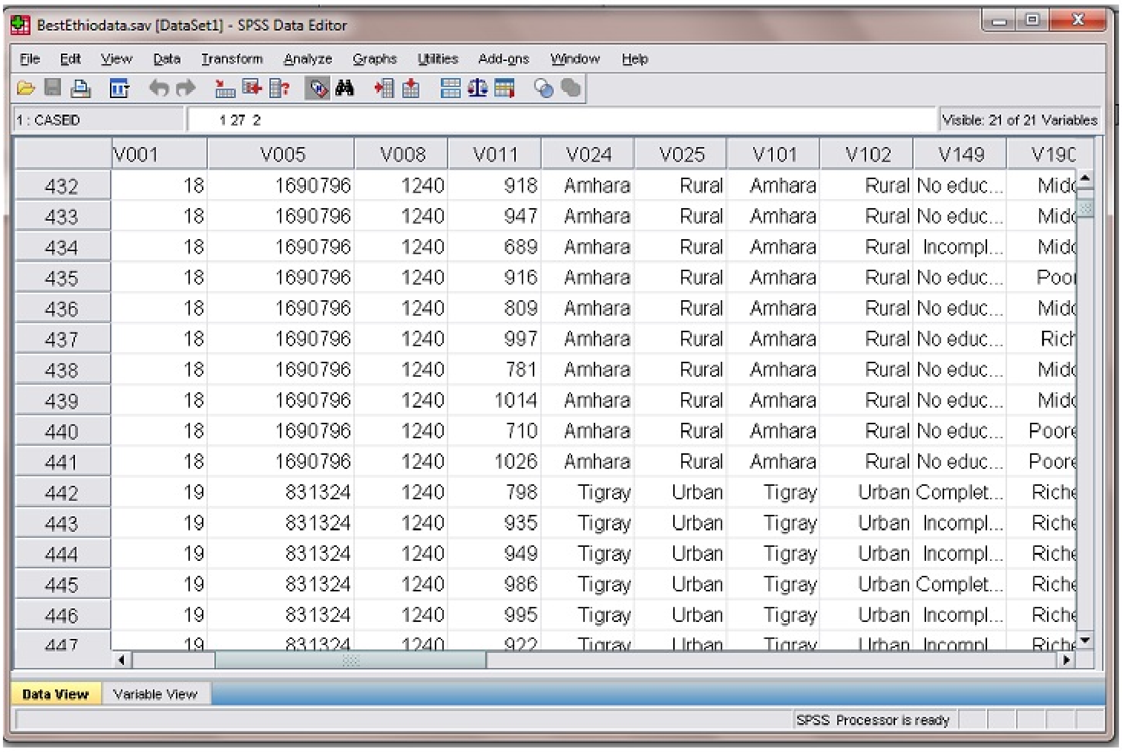
2 MODELING ASFR IN ETHIOPIA

2.1 Introduction

Ethiopia is characterized by distinct physiological and ethnic diversities which present challenges and opportunities for its development. Because of Ethiopian cultural, economical, and geographical diversity, the magnitude of regional variation in fertility levels is anticipated to be much larger. In this vein, regional level fertility perspectives or analyses are so important to understand existing disparities. This thesis in general focuses on regional fertility patterns and its implication. Therefore, the remain parts of the chapter are unfolded as in the following way: the next section is devoted to introduce about the data utilized in the entire work of this thesis. Then, the pattern of ASFRs at national and regional level is described in the upcoming section, in subsection 2.3. Following this, we provide a brief review of some existing models for fertility patterns and then propose our model in subsection 2.4 and subsection 2.5 respectively. Fitting fertility models to DHS data is done in subsection 2.6. The results of fitting our model as well as other models to the fertility data are presented in subsection 2.7. This paper concludes with a discussion of the implications of our findings in the last section of the chapter.

2.2 Data Source

In developing countries like Ethiopia, population information is usually derived from censuses and surveys collected occasionally as found necessary. Vital registration systems that yield continuous flow of information are incomplete or non-existent (Farina et al., 2001). In the absence of a complete and highly systematized vital registration system, census and survey data are used to estimate demographic parameters with all their defects and shortcomings. In this thesis, we make use of data obtained from the 2011 Ethiopia Demographic and Health Survey (EDHS), available online on ICF International's website `http://www.measuredhs.com`. This is the third Demographic and Health Survey (DHS) in Ethiopia, which is funded by USAID and conducted by Central Statistical Agency(CSA), Ethiopia, and ICF International (the owner of the raw data) in Calverton, Maryland as part of the worldwide MEASURE DHS project, a five-year project engaging in collecting and analyzing data needed to plan, monitor, and evaluate population, health, and nutrition programs in countries worldwide. The survey contains ample data on fertility, health, contraceptive use, breast-feeding practices, nutrition of women and children, and others though, on this work,we only focused on different key indicators relating to fertility information of 16515 eligible women in the reproductive age (a.k.a.: childbearing age) [7], taken from 11 regions located across the country. Figure 2.1 shows the partial



	V001	V005	V008	V011	V024	V025	V101	V102	V149	V190
432	18	1690796	1240	918	Amhara	Rural	Amhara	Rural	No educ...	Midc
433	18	1690796	1240	947	Amhara	Rural	Amhara	Rural	No educ...	Midc
434	18	1690796	1240	689	Amhara	Rural	Amhara	Rural	Incompl...	Midc
435	18	1690796	1240	916	Amhara	Rural	Amhara	Rural	No educ...	Poo
436	18	1690796	1240	809	Amhara	Rural	Amhara	Rural	No educ...	Midc
437	18	1690796	1240	997	Amhara	Rural	Amhara	Rural	No educ...	Rich
438	18	1690796	1240	781	Amhara	Rural	Amhara	Rural	No educ...	Midc
439	18	1690796	1240	1014	Amhara	Rural	Amhara	Rural	No educ...	Midc
440	18	1690796	1240	710	Amhara	Rural	Amhara	Rural	No educ...	Poore
441	18	1690796	1240	1026	Amhara	Rural	Amhara	Rural	No educ...	Poore
442	19	831324	1240	798	Tigray	Urban	Tigray	Urban	Complet...	Riche
443	19	831324	1240	935	Tigray	Urban	Tigray	Urban	Incompl...	Riche
444	19	831324	1240	949	Tigray	Urban	Tigray	Urban	Incompl...	Riche
445	19	831324	1240	986	Tigray	Urban	Tigray	Urban	Complet...	Riche
446	19	831324	1240	995	Tigray	Urban	Tigray	Urban	Incompl...	Riche
447	19	831324	1240	922	Tigray	Urban	Tigray	Urban	Incompl...	Riche

Figure 2.1: Ethiopia DHS, 2011 (in SPSS format)

data structure of the survey in SPSS data format. Computation of important variables such as Mother's age and one year ASFR has been made for all women in each region. Table 2.1 reveals roughly the summarized data template employed in the entire work of the analysis. In this case, $i = 1, \ldots, I = 11$ stands for the i^{th} regions in the country. As mentioned, since the number of regions are 11 in number, then, $I = 11$ and are coded for analysis purpose as: $1 =$ Tigray (Region-1), $2 =$ Affar (Region-2), $3 =$ Amhara (Region-3), $4 =$ Oromiya (Region-4), $5 =$ Somali (Region-5), $6 =$ Benshangul Gumuz (Region-6), $7 =$ SNNP (Region-7), $8 =$ Gambela (Region-12), $9 =$ Harari (Region-13), $10 =$ Addis Ababa (*Region-14*), and,$11 =$ Dire Dawa (Region-15); $j = 1, \ldots, n_i$ is the j^{th} woman(mother) in the region i^{th}; Likewise, Y_{ij} refers to the age of the j^{th} mother in the reproductive age living in the i^{th} region in the study period; and f_{ij} corresponds to the ASFR of the j^{th} mother between 15-49 years old in the i^{th} region in the country during the study period.

[7] Reproductive or childbearing age:*is the reproductive period or span of a woman, assumed for statistical purposes to be 15-49 years of age.*

Mother (j)	Region (i)	Region Code (i)	Age (Y_{ij})	ASFR (f_{ij})
1	Tigray		$y_{1(1)}$	$f_{1(1)}$
2	Tigray		$y_{1(1)}$	$f_{1(1)}$
$\vdots$	$\vdots$	1 (Region-1)	$\vdots$	$\vdots$
n_1	Tigray		$y_{n_1(1)}$	$f_{n_1(1)}$
1	Affar		$y_{1(2)}$	$f_{1(2)}$
2	Affar		$y_{2(2)}$	$f_{2(2)}$
$\vdots$	$\vdots$	2 (Region-2)	$\vdots$	$\vdots$
n_2	Affar		$y_{n_2(2)}$	$f_{n_2(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	Dire Dawa		$y_{1(11)}$	$f_{1(11)}$
2	Dire Dawa		$y_{2(11)}$	$f_{2(11)}$
$\vdots$	$\vdots$	11(Region-15)	$\vdots$	$\vdots$
n_{11}	Dire Dawa		$y_{n_{11}(11)}$	$f_{n_{11}(11)}$

Table 2.1: **Data Layout**

2.3 Pattern of ASFRs of Ethiopia

The age-specific fertility rate curve is, in general, a bell-shaped unimodal curve which first rises slowly and then sharply in the age group 15-19, attains its modal value somewhere between ages 20-29, declines first slowly and then steeply till it approaches zero around the age of 50 years even though some developed countries has already started showing a deviation from this classical bell shaped curve. The magnitude of this pattern is often influenced by different socioeconomic factors, like education, occupation, religion, contraceptive practice, etc and demographic factors, like age at marriage, present family size, gender preference, et.

To address the first objectives of the study, we first took advantage of real data set from 2000, 2005, 2011 Demographic and Health Survey (DHS) of Ethiopia, from which we extracted one-year age specific fertility rate (ASFR) and examined its pattern at both national and regional levels. The plot at national level, which is depicted in Figure 2.1, indicates the age specific fertility patterns are characterized by reversed V-shape. Besides, the 2011 pattern confirms that the ASFRs declined in this year much more than in the previous surveys, especially in all of the women younger than age 44. This pattern rises to reach the peak in the late twenties (age group 25-29) before it starts dropping rapidly. This indicates that there was a shift in fertility towards younger age groups in 2011. It is also observed that the highest ASFRs are in the age group 25-29. Therefore, this age group is the most fertile period of Ethiopia women in the stated year. On the other hand, the 2005's pattern starts with a relatively higher rate and peaks in the early twenties (age group 20-24). It does not demonstrate a rapid drop after the peak. In contrast to the 2005

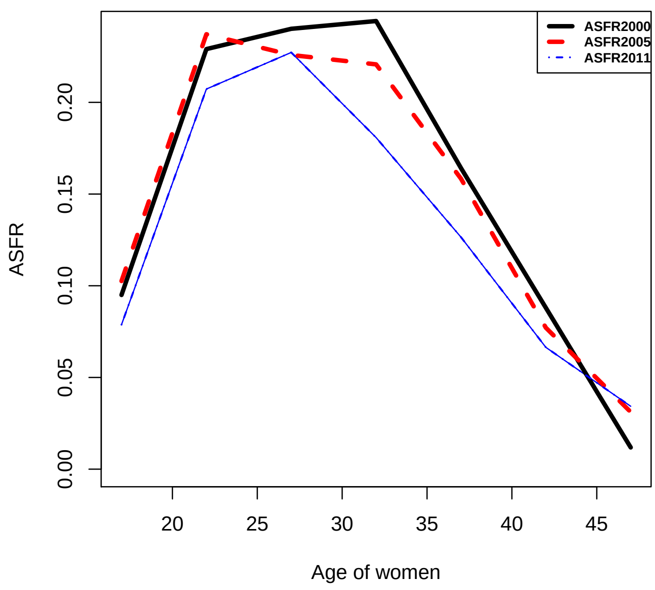


Figure 2.2: Trend in ASFRs of Ethiopia in 2000, 2005 and 2011

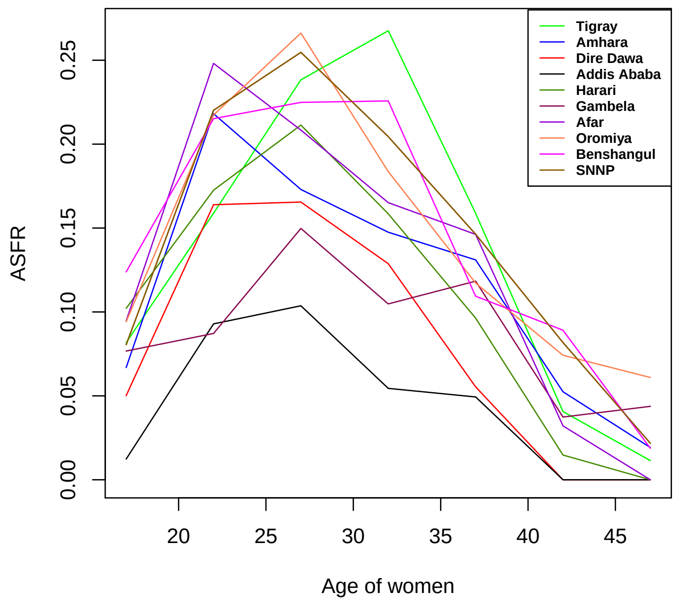


Figure 2.3: Pattern in ASFRs across the regions, 2011 EDHS

and 2011 fertility pattern, the pattern in 2000 depicts a broad peak shape that extends from the early twenties to the early thirties before a rapid drop. It is also noticed that the age group 30-35 was the most fertile years in 2000 while 20-24 age group was the most fertile in 2005. The study also scrutinized how the pattern looks like at regional level. As shown in Figure 2.3, Age-specific fertility levels markedly differ by regions. For instance, mothers of aged 15-19 years have highest fertility rate in Affar region and lowest in Addis Ababa showing a marked differences of 89 between maximum and minimum figures. Both regions also show considerable deviation from the fertility level in other regions. It is also learnt that the age specific schedules seems to have a similar pattern as of national level for all regions except for Somali and Gambella, which could be due to the effect of small sample size. Furthermore, we observe ASFRs of Addis Ababa was much smaller than any other regions in the country in all age groups,which could be owning to the presence of more socio-economically development on this region than the rest. In contrary, all other regions have high ASFRs in some age groups and lower in other. There are also clear inter-regional disparity in factors of fertility. For instance, 3% of young women in Addis Ababa have started childbearing by age 19, compared with 21% of young women in Gambela.
All in all, considerable variation in the pattern of ASFRs of women is reflected at regions vis-á-vis national level. The figure also reveals not only fertility intensity (TFR) but also ASFR shapes varies across regions. This variation in the pattern and other fertility indicators revealed at inter-regional level in Ethiopia calls for designing a more flexible fertility model with which we can provide a good fit and parameter estimation in all regions in the country.

2.4 Summary on Fertility Models (Literature Review)

An easy way of finding out different characteristics of a fertility distribution is to build a suitable model for fertility distribution and derive various important characteristics of fertility distribution from the fitted model. While building any model we have to keep the following points in our mind. Models are alternative means for describing a given data. The basic objective of any modeling is to reduce confusing mass of numbers to a few intelligible basic parameters. Experts appreciate those models where number of parameters are as small as possible and are interpretable in physical terms and are also good enough to approximate all the relevant variations that are observable in the data (Pasupuleti and Pathak, 2010).

A large number of parametric and nonparametric models have been proposed in demographical literature for modelling the age specific fertility curves of many populations, but fitting these models to curves of Ethiopian data has not been undertaken yet. Among these models applied to ASFR data of different countries are:

2.4.1 Parametric Fertility Models

For estimating mortality and fertility patterns several parametric and non-parametric techniques have been proposed. The parametric ones are non linear models that represent the mortality pattern as a function of age and a number of parameters

★ **Hadwiger function**: the Hadwiger function (Hadwiger, 1940; Gilje, 1972) is expressed by,

$$\mathtt{f(x)} = \frac{\mathtt{ab}}{\mathtt{c}}\left(\frac{\mathtt{c}}{\mathtt{x}}\right)^{\frac{3}{2}} \exp\left\{-\mathtt{b}^2\left(\frac{\mathtt{c}}{\mathtt{x}} + \frac{\mathtt{x}}{\mathtt{c}} - 2\right)\right\} \tag{1}$$

where, $\mathtt{f(x)}$ is the ASFR of mothers in the study at age $\mathtt{x}$, and $\mathtt{a}$,$\mathtt{b}$,$\mathtt{c}$ are the three parameters of the model to be estimated, in which case, the parameter a refers to total fertility,the parameter $\mathtt{b}$ determines the height of the curve, the parameter c is related to the mean age of motherhood, while the term $\frac{\mathtt{ab}}{\mathtt{c}}$ is related to the maximum age-specific fertility rate (or modal age-specific fertility rate)(Chandola et al., 1999).

★ **Gamma function**: the Gamma function (Hoem et al., 1981) is given by,

$$\mathtt{f(x)} = \mathtt{R}\frac{1}{\Gamma(\mathtt{b})\mathtt{c}^{\mathtt{b}}}(\mathtt{x}-\mathtt{d})^{\mathtt{b}-1}\exp\left\{-\left(\frac{\mathtt{x}-\mathtt{d}}{\mathtt{c}}\right)\right\},\ \mathtt{for}\ \mathtt{x} > \mathtt{d} \tag{2}$$

where, $\mathtt{f(x)}$ is the ASFR of mothers in the study at age $\mathtt{x}$, $\mathtt{d}$ represents the lower age at childbearing, while the parameter $\mathtt{R}$ determines the level of fertility. The parameters $\mathtt{b}$, $\mathtt{c}$ have no direct demographic interpretation, but Hoem et al. (1981) have substituted these by the mode $\mathtt{m}$, the mean μ and the variance δ^2 of the density, for $\mathtt{c} = \mu - \mathtt{m}$ and $\mathtt{b} = \frac{\mu-\mathtt{d}}{\mathtt{c}} = \frac{\delta^2}{\mathtt{c}^2}$.

★ **Beta function**: the Beta function proposed by Hoem et al. (1981) which is given by the formula,

$$\mathtt{f}(\mathtt{x}) = \mathtt{R}\frac{\Gamma(\mathtt{A}+\mathtt{B})}{\Gamma(\mathtt{A})\Gamma(\mathtt{B})}(\beta-\alpha)^{-(\mathtt{A}+\mathtt{B}-1)}(\beta-\mathtt{x})^{(\mathtt{A}-1)}(\mathtt{x}-\alpha)^{(\mathtt{B}-1)}, \text{ for } \alpha < x < \beta \tag{3}$$

where $\mathtt{f}(\mathtt{x})$ is the ASFR of mothers in the study at age $\mathtt{x}$, and $\mathtt{R}$ determines the level of fertility. Hoem et al. (1981) showed that $\mathtt{A}$ and $\mathtt{B}$ are related to the mean $\mathtt{v}$ and the variance τ^2 through the relations

$$\mathtt{B} = \left\{\frac{(\mathtt{v}-\alpha)(\beta-\mathtt{v})}{\tau^2} - 1\right\}\frac{\beta-\mathtt{v}}{\beta-\alpha} \quad \text{and} \quad \mathtt{A} = \mathtt{B}\frac{\mathtt{v}-\alpha}{\beta-\mathtt{v}}$$

but not in a simple, easily interpretable way. He also stated that α and β represent lower and upper age limits of fertility, but Peristera and Kostaki (2007) showed that in several cases the value of β far exceeds the maximum age.

★ **Peristera and Kostaki Model:** Peristera and Kostaki (2007) noted that the form of the fertility curve has changed in recent years in various countries, as did Chandola et al. (1999) before them. To this end, they proposed flexible models that capture both the standard (classical) age-specific fertility pattern and the distorted, as the pattern shown in countries such as the United Kingdom, Ireland, and Spain.

⇒ Their basic model resembles the normal distribution but is asymmetrical, as the spread before and after the peak differs, and is expressed:

$$\mathtt{f}(\mathtt{x}) = \mathtt{c}_1\exp\left[-\left(\frac{\mathtt{x}-\mu_1}{\sigma_1(\mathtt{x})}\right)^2\right] \tag{4}$$

where $\mathtt{f}(\mathtt{x})$ is the ASFR of mothers in the study at age $\mathtt{x}$; $\mathtt{c}_1$ and μ are parameters to be estimated while

$$\sigma(x) = \begin{cases} \sigma_{11} & \text{if } \ x \leq \mu_1, \\ \sigma_{12} & \text{if } \ x > \mu_1, \end{cases}$$

The parameter $\mathtt{c}_1$ describes the base level of the fertility curve and is associated with the total fertility rate, μ reflects the location of the distribution, *i.e.* the modal age, while σ_{11}, σ_{12} reflect the spread of the distribution before and after its peak, respectively.

⇒ An alternative version of this model, which captures the distorted shape of the fertility pattern, is

$$\mathtt{f}(\mathtt{x}) = \mathtt{c}_1\exp\left[-\left(\frac{\mathtt{x}-\mu_1}{\sigma_1(\mathtt{x})}\right)^2\right] + \mathtt{c}_2\exp\left[-\left(\frac{\mathtt{x}-\mu_2}{\sigma_2}\right)^2\right] \tag{5}$$

where,

$$\sigma_1(x) = \begin{cases} \sigma_{11} & \text{if } x \leq \mu_1, \\ \sigma_{12} & \text{if } x > \mu_1, \end{cases}$$

where $\mathtt{f(x)}$ is the ASFR of mothers in the study at age $\mathtt{x}$, and $\mathtt{c}_1$, $\mathtt{c}_2$, μ_1, μ_2, σ_{11}, σ_{12} are parameters to be estimated. Moreover, the parameters $\mathtt{c}_1$ and $\mathtt{c}_2$ reflect the level of fertility at the first and second peak respectively; μ_1 and μ_2 are related to the mean age of the two subpopulations, the one with earlier fertility and the other with fertility at later ages; and the parameters σ_{11}, σ_{12} reflect the spread of the distribution of the most intense hump before and after each peak, and σ_2 reflects the variance of the less intense one.

★ **Skew Normal Model**: Mazzuco and Scarpa (2015) have introduced a different model based on skew-normal density function (Azzalini, 1985), which has 4 parameters. Because of the skewness parameter, that model has been suitable (flexible) for almost all types of fertility patterns, not only unimodal but also bimodal fertility patterns. The one which is flexible for unimodal is given as defined in Equation 6:-

$$\mathtt{f}\left(\mathtt{x};\xi,\omega,\alpha\right) = \frac{2\mathtt{R}}{\omega}\phi\left(\frac{\mathtt{x}-\xi}{\omega}\right)\Phi\left(\alpha\frac{\mathtt{x}-\xi}{\omega}\right),\ \mathtt{x}\in\mathbb{R},\quad \xi\in\mathbb{R},\quad \omega>0,\quad \alpha\in\mathbb{R} \tag{6}$$

where, ϕ and Φ are respectively the *pdf* and *cdf* of standard normal distribution; $\mathtt{f}\,(\mathtt{x})$ is ASFR at age x of a mother, $\mathtt{R}$ is the **TFR parameter** of the model; ξ and α are the **location parameter** and the **shape parameter** of the model respectively while ω is the **scale parameter** of the model. More detail discussion on this model is given in the forthcoming part, subsection 3.2 of Bayesian Analysis.

★ **Gompertz Curve**: the Gompertz model is widely used in demography and in various branches of science. Initially, the model was developed by Gompertz (1825) to describe age patterns of mortality

$$\begin{aligned} \mathtt{f}\,(\mathtt{x}) &= F(x+1) - F(x) \\ \mathtt{F(x)} &= Ra^{b^{(x-x_o)}} \end{aligned} \tag{7}$$

where $\mathtt{f}\,(\mathtt{x})$ is the age specific fertility rate at age x and $\mathtt{F}\,(\mathtt{x})$ is the average number of children born by exact age x. The parameter R is the total fertility rate (TFR), a is the proportion of total fertility attained by age x_o, b is the intrinsic rate of growth of cumulative age specific fertility rate by age. Wunsch (1966), Martin (1967) , Murphy and Nagnur (1972), Farid (1973), Brass (1981); have suggested using Gompertz curve to model fertility distributions.

2.4.2 Non-Parametric Fertility Models

Instead of specifying a statistical model, one may use a non-parametric model for smoothing age patterns of fertility. The structure of a non-parametric model is not specified a priori but is determined from the data. Non-parametric does not mean that the model does not include parameters, but that the number of parameters is not fixed in advance and that the parameters lack a clear statistical interpretation. There are several approaches to estimating non-parametric models. The most widely applied are local polynomial regression and smoothing splines (Fox, 2000). To the best of my knowledge, local polynomial regression is not applied to fitting fertility schedules. Quadratic and cubic splines are very flexible and so may provide an accurate fit of various types of fertility curves

★ **Quadratic Spline**: Schmertmann (2003) proposes fitting splines by choosing as knots particular ages that can be interpreted. Schmertmann fits the age pattern of fertility by a quadratic spline including four knots, which means that the age schedule is described by five quadratic pieces. A quadratic spline is a piecewise quadratic function that can be described by

$$\mathtt{f}(\mathtt{x}) = \mathtt{a} + \mathtt{b}(\mathtt{x} - \mathtt{m}) + \sum_{\mathtt{j}=1}^{\mathtt{n}} \mathtt{c}_{\mathtt{j}} \left(\mathtt{x} - \mathtt{m} - \mathtt{k}_{\mathtt{j}}\right)^2 \mathtt{D}_{\mathtt{j}} \tag{8}$$

where $\mathtt{D}_\mathtt{j} = 0$ if $\mathtt{x} - \mathtt{m} \leq \mathtt{k}_\mathtt{j}$ and $\mathtt{D}_\mathtt{j} = 1$ otherwise, $\mathtt{m}$ is the minimum age, $\mathtt{x} \geq \mathtt{m}, \mathtt{k}_\mathtt{j}$ are the knots, $\mathtt{n}$ is the number of knots, and $\mathtt{a}$, $\mathtt{b}$, and $\mathtt{c}_\mathtt{j}$ are the coefficients to be estimated(de Beer, 2011).

★ **Cubic Splines**: are very flexible (McNeil et al., 1977; Gilks, 1986) and can be described by:

$$\mathtt{f}(\mathtt{x}) = \mathtt{a} + \mathtt{b}(\mathtt{x} - \mathtt{m}) + \mathtt{c}(\mathtt{x} - \mathtt{m})^2 + \sum_{\mathtt{j}=1}^{\mathtt{n}} \mathtt{d}_{\mathtt{j}} \left(\mathtt{x} - \mathtt{m} - \mathtt{k}_{\mathtt{j}}\right)^3 \mathtt{D}_{\mathtt{j}} \tag{9}$$

where $\mathtt{D}_\mathtt{j} = 0$ if $\mathtt{x} - \mathtt{m} \leq \mathtt{k}_\mathtt{j}$ and $\mathtt{D}_\mathtt{j} = 1$ otherwise, $\mathtt{m}$ is the minimum age, $\mathtt{x} \geq \mathtt{m}, \mathtt{k}_\mathtt{j}$ are the knots, $\mathtt{n}$ is the number of knots, and $\mathtt{a}$, $\mathtt{b}$, $\mathtt{c}$, and $\mathtt{d}_\mathtt{j}$ are the coefficients to be estimated(de Beer, 2011).

A detailed descriptions of the mathematical formulae of most of the models can be found in (de Beer, 2011), (Hoem et al., 1981) and (Peristera and Kostaki, 2007).

To accurately model fertility patterns at country and regional levels in Ethiopia and make inter-regional fertility analyses, a new mathematical model that is both easily used, and provides good fit for the data is required. Such a model could reveal important parameters which need to be taken into account when comparing fertility between regions and across time. This undoubtedly would increase our understanding of fertility scenarios in the regions of the country. In this study, we proposed to use the four parameter Skew Normal Model (Mazzuco and Scarpa, 2015), given in Equation 6.

2.5 (Proposed)Fertility Model: Skew Normal Model in Brief

As categorically elaborated by Singh et al. (2015), fertility analysis of human population is done usually by two ways:-

(*i*). The first way is primarily concerned with the estimation of parameters of standard measures of fertility. Demographic factors like age at marriage, present family size, gender preference (Mahadevan, 1979; Bhasin and Bhasin, 1990; Asari and John, 1998; Chachra and Bhasin, 1998; Bhasin and Nag, 2002) and socioeconomic factors like education, occupation, religion, contraceptive practice, etc. (Bhatia, 1970; Asari and John, 1998) are the determinants of desired family size and the generally considered as cause of the variation in fertility. Place of residence, family type, mass media exposure are some factor that indirectly plays an important role and it affects the fertility. The couple's decision about their next child affects the birth interval and indirectly this affect the fertility.

(*ii*). The second way of fertility analysis deals with the fertility pattern through mathematical curves. Due to its simplicity, the modeling of fertility pattern through different mathematical curves have attracted the interest of demographers and still the researcher who were working in the field of demography are using different mathematical curve to graduate the trend of fertility analysis. Among other reasons, the interest in fertility modeling through curves is due to the fact that it is helpful in population projection, which is very useful for government planning. Fertility analysis plays an important role to measure the intensity of population growth.

Fertility studies in Ethiopia date back to decades and have examined a wide range of topics on fertility: *Determinants* (Alemayehu et al., 2010), *Differentials of fertility* (Alene and Worku, 2008), *Level and Differentials of fertility* (Gebremedhin, 2006; Gebremedhin and Betre, 2009), *Fertility preferences and the demand* (Short and Kiros, 2002), *Differentials of fertility* (Fitaw et al., 2003), *Impact of child mortality and fertility preferences* (Fitaw et al., 2004), *The influence of socio-demographic factors* (Eshetu and Habtamu, 1998), *Fertility decline driven by poverty* (Gurmu and Mace, 2008), *Fertility transition driven by poverty* (Gurmu, 2005), *The quiet revolution* (Kinfu, 2001), *The demographic component of fertility decline*(Lindstrom and Woubalem, 2003), *The Proximate Determinants*(Sibanda et al., 2003) to name the important ones.

Despite their remarkable contribution on the overall study of fertility, those previous studies have also a lot major similar features (limitations) in common. Of those, some are:-

(*a*). Almost all study done thus far used the approach explicated in (i) to make analysis of fertility. In other words, none of the works done so far had dealt with ASFR and analysis of fertility using mathematical modeling, which is one of the targets of this thesis.

(b). Furthermore, many of the researches at country, regional and district levels were focused on the pooled data rather than multilevel and were also conducted using only non-Bayesian setting, which is the second major gap this thesis need to bridge.

In view of this, in this thesis, we propose a four parametric skew normal mathematical model of Mazzuco and Scarpa (2015), described in Equation 6, to model the pattern of ASFR, at national and regional level, and fit the parameters in the model by means of simple Bayesian and Hierarchical Bayesian analyses, as discussed more in the subsequent chapters. The merit behind using mathematical model in Bayesian setting to make fertility analysis is that:

(a). **Its simplicity**: Due to its simplicity, the modeling of fertility pattern through different mathematical curves have attracted the interest of demographers, and still the researcher who were working in the field of demography are using different mathematical curve to graduate the trend of fertility analysis.

(b). **Its applicability**: Among other reasons, the interest in fertility modeling through curves is due to the fact that it is helpful in population projection, which is very useful for government planning. Fertility analysis plays an important role to measure the intensity of population growth. As we are using multilevel regional level data, such analysis will have a remarkable role to make regional-specific inferences.

Analogously, using simple Bayesian analysis in general and Hierarchical Bayesian analyses in particular to estimate the parameters of the proposed model have also some interesting plus points and issues to resolve particularly from multilevel data perspective.

(a). **Issue of Biasedness and accuracy:** since we have regions of varying sample sizes (for more see: Table 1.1), as outlined in Griffiths et al. (1987), using non-Bayesian methods, particularly, the MLE will result in a significant bias, particularly for these regions whose small sample size is very low, such as Somali (see Figure 2.3), Dire Dawa, Harari, Gambela, etc; and therefore, our inference based on classical approach will be not accurate when we have regions having small observation (Zellner and Rossi, 1984).Therefore, one of the reason behind using Bayesian approach to make inference on our model parameters is, hence,that it is a good way out to overcome such problem.

(b). **Issue of Borrowing information/ strength:**
Simple Bayesian analysis, we estimate our parameters assuming that each of the 11 regions has an identical fertility model parameters, such as all regions have the same

Figure 2.4: Patter of Age-Specific Fertility Rates for Somali Region, 2011

fertility rate, R, the same fertility shape parameter, α, etc. The assumption of a common fertility model parameters, however, is quite strong, and may be inappropriate even in Ethiopian regions. Suppose that instead of a common fertility model parameters, we consider the fertility model parameters for each region, as a sample from an *overall population distribution*. Thus, by treating the fertility model parameters, as exchangeable, we develop the following hierarchical model. This model allows the mothers for each region to have its own fertility model parameters, but it also models each fertility model parameters, as coming from a *common population distribution*. Thus, *hierarchical structure allows us to borrow strength for the estimation of each fertility model parameters*. In particular, the estimation of each fertility model parameters is improved by using the fertility data from the other regions.

Bayesian analysis is less dependent on the asymptotic assumptions and is able to produce reliable results with smaller sample sizes (Lee, 2007) with prior distributions having a significant role when sample sizes are small or moderate (Lee, 2007).

To sum up, thoroughly assessing the patterns of ASFR at country and regional level, developing a model and fitting these model with the aid of Hierarchical Bayesian Method, which are the main purposes of thesis, not only possess the aforementioned merits but also are a new approach that opens a new avenue to Ethiopian demographic research development activities. Such a work will also shade some light on how integrate demographic data with Bayesian statistics. In-depth discussion on the model, the Bayesian procedures and on how these can be implemented with the data, the ASFR mothers in the study, and to estimated model parameters is given in subsequent sections

2.6 Fitting the model to Real Data

2.6.1 Preliminary Analysis

In order to determine the performance of the proposed model, we conducted some preliminary analysis of fitting the model along with ten other commonly used parametric and non-parametric models, described in subsection 2.7, namely: the Quadratic spline (QS), Cubic spline (CS), Beta function (BF), Gamma function (GF), Hadwiger function (HF), Skew Normal (SN), Gompertez curve (GC), Adjusred Error Model(AEM), Polynomial function (PF) and Model-1 of Peristera et.al.

The purpose of fitting various models is to compare the performance of the proposed model, Skew Normal Distribution, with other models mentioned thus far. The criterion followed in fitting these models is *Nonlinear Regression with nonlinear least squares* estimation, obtained by minimizing the following sum of square residual (SSQ) equation:-

$$\mathtt{SSQ}(\mathtt{x};\mathtt{R},\theta_2,\cdots,\theta_{\mathtt{r}-1}) = \sum_{\mathtt{x}} \left\{ \mathtt{f}(\mathtt{x};\mathtt{R},\theta_2,\cdots,\theta_{\mathtt{r}-1}) - \widehat{\mathtt{f}}_{\mathtt{x}} \right\}^2 \tag{10}$$

where

- √ $f(x;R,\theta_2,\cdots,\theta_{r-1})$ is an analytical function, considered to model the single year age specific fertility schedule, related as

$$\mathtt{f}(\mathtt{x};\mathtt{R},\theta_2,\cdots,\theta_{\mathtt{r}-1}) = \mathtt{R}\cdot\mathtt{h}(\mathtt{x};\mathtt{R},\theta_2,\cdots,\theta_{\mathtt{r}-1}), \tag{11}$$

  with $h(x;R,\theta_2,\cdots,\theta_{r-1})$ is a probability density function *(pdf)* on the real line with $r-1$ parameter and can be one of the above fertility models, such as SN, Beta, Gamma, etc,

- √ $\widehat{f_x}$ is the estimated of f_x , and

- √ R is the $r^{th}(\equiv\theta_k)$ parameter in Equation 10 and refers to the total fertility rate (TFR) of our data.

The majority of these fertility models described above are *nonlinear* in nature. In general, if f is nonlinear in θ, it is not usually possible analytically to obtain an explicit solution for the *MLE estimate* and to conduct fitting the model to the data, especially when the model, f, involves many parameters and is highly non-linear. Instead an iterative procedure is needed. In such situations, the *MLE* estimate and model fitting must be sought numerically using nonlinear optimization algorithms. This is true for these models. In our case, the `R` function `nls` was employed for the purpose of fitting these models, plotting the empirical values of the models with the observed curves.

2.6.2 Model selection

A measure of the relative quality of a statistical model, known as, *Akaike Information Criterion* (AIC, for short), has been instrumented so as to compare the performance of these candidate fertility models and select which, the proposed or the other fertility models aforementioned above, is the best description of the current data under scrutiny. The AIC is a criterion for comparing various functions by adjusting the SSQ for the number of observations and the number of parameters in the model. The criterion can be used to decide if the improved fit is worth with the decreased degrees of freedom and the increased complexity of the function caused by the addition of another parameter to a model. The variant of the AIC used in this paper is given by

$$\mathtt{AIC} = 2 * \mathtt{k} + \mathtt{n} * \mathtt{ln}\left\{\frac{\mathtt{SSQ}}{\mathtt{n} - \mathtt{k}}\right\} \tag{12}$$

where $\mathtt{k}$ is the number of parameters in the model, $\mathtt{SSQ} = \sum_{\mathtt{x}}(\mathtt{f}_{\mathtt{x}} - \widehat{\mathtt{f}}_{\mathtt{x}})^2$ is the sum of square residuals in which case $\mathtt{f}_{\mathtt{x}}$ is the empirical fertility rate at age $\mathtt{x}$ and $\widehat{\mathtt{f}}_{\mathtt{x}}$ is its estimate, and $\mathtt{n}$ is the number of observations. When comparing the performances of a number of models, the model with least AIC is usually preferred as the best. The AIC procedure has been discussed and successfully used to identify model with best fit by Akaike (1976), Tong (1977); Ozaki (1977), Larimore and Mehra (1985), and Koehler and Murphree (1988), among others.

2.7 Result of the fit

We have fitted the functions given in subsection 2.7 to the one-year ASFR data extracted from 2011 EDHS, at national level and for each regions. Table 2.2 recaps the result of the fit at country level, and for visual comparison, the result is further given in Figure 2.5. A similar result for two regions has been given in Figure 2.6 & Figure 2.7. In all cases, the observed(empirical) age specific fertility rates for the study populations have been listed in column 2. In columns 3 to 10, we have given the fitted function values. One remarkable observation from the results obtained is that the AIC of SN is smaller at national level data and for A.A., Tigray, Harar, Dire Dawa, Oromiya and SNNP.

Table 2.2: **Empirical and fitted values for the ASFR at national level with values of minimization and model selection criterion**

Age	DHS	Gamma	Quad.S	CubicS	Beta	PeristeraK	Gomp	Hadwiger	SkewN
15-19	0.0787964	0.07941014	0.11433305	0.08149314	0.07877457	0.10468448	0.08250132	0.08432734	0.1037879
20-24	0.2072576	0.20540353	0.16846207	0.20130199	0.20747701	0.18113166	0.20109639	0.19978583	0.18096631
25-29	0.2272192	0.22973613	0.19356891	0.22640882	0.22621386	0.22427694	0.23274250	0.23139501	0.22460378
30-34	0.1807907	0.18213480	0.18965356	0.18965356	0.18332391	0.19872525	0.18305005	0.18509950	0.19893728
35-39	0.1264974	0.11955640	0.15671604	0.12387612	0.12274915	0.12600844	0.11782620	0.11952747	0.12610677
40-44	0.0664233	0.06964088	0.09475633	0.06191641	0.06957800	0.05717740	0.06835614	0.06751603	0.05739148
45-49	0.0342798	0.03735370	0.00377444	0.03661436	0.03306453	0.01856638	0.03757527	0.03490573	0.01881487
AIC		-49.7963	-20.8742	-45.1797	-53.9988	-27.3542	-46.2284	-46.4085	-55.5031
RSS		7.99e-05	0.0066253	0.00015458	3.29e-05	0.0019728	0.00017259	0.0019728	3.54e-05

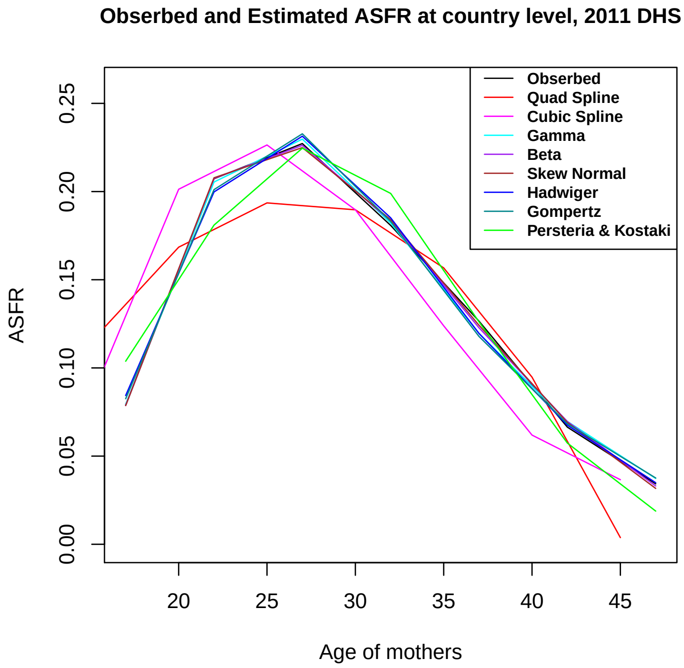


Figure 2.5: *Plot of Observed and Estimated ASFR at country level, 2011 DHS.*

Table 2.3: **Empirical and fitted values for *Addis Ababa* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	1.233490e-02	0.03773830	0.0152599786	0.05484471	0.012073362	0.020961944	0.012179910
20-24	9.288160e-02	0.06446411	0.0869424238	0.11450746	0.093431208	0.086993854	0.093080214
25-29	1.036736e-01	0.07654115	0.0990194666	0.12935400	0.099068819	0.104542567	0.101127084
30-34	5.446270e-02	0.07396942	0.0739694238	0.10918384	0.066857306	0.065694543	0.064346043
35-39	4.935040e-02	0.05674893	0.0342706120	0.07768684	0.032241838	0.028244347	0.031778451
40-44	4.420000e-10	0.02487966	0.0024013480	0.04939944	0.011110913	0.009550461	0.012241331
45-49	7.540000e-10	-0.02163837	0.0008399484	0.02900817	0.002736142	0.002749299	0.003677994
AIC		-24.93042	-34.81144	-15.98306	-35.69982	-35.84763	-35.96744
RSS		0.0037115	0.0006799	0.01001338	0.000599	0.0007804	0.0005762

Table 2.4: **Empirical and fitted values for *Harari* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	1.021462e-01	0.07592061	0.096193645	0.08330303	0.101653339	0.08379656	0.100657698
20-24	1.725509e-01	0.10469356	0.189036670	0.19446002	0.173690715	0.19296910	0.174473906
25-29	2.113519e-01	0.11867410	0.200077112	0.20417907	0.208278303	0.20583789	0.208553080
30-34	1.584689e-01	0.11786221	0.157761457	0.14674206	0.166063092	0.14736166	0.165187648
35-39	9.633450e-02	0.10225791	0.090536188	0.08444879	0.083174387	0.08372554	0.083795796
40-44	1.474320e-02	0.07186118	0.026847789	0.04198894	0.026169145	0.04115787	0.026522879
45-49	5.090000e-09	0.02667203	-0.004857255	0.01882348	0.005172171	0.01838065	0.005140335
AIC		-26.0029	-35.25019	-26.39688	-38.53939	-28.91044	-38.85986
RSS		0.003184297	0.0006385693	0.002261967	0.0003991529	0.002101956	0.0003813

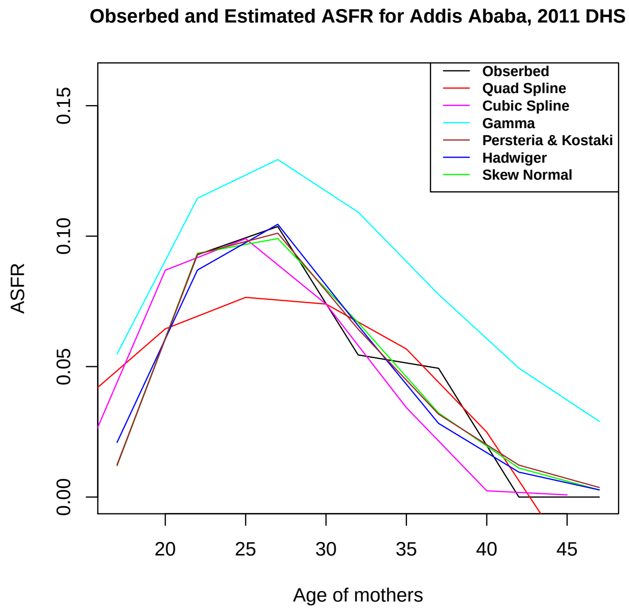


Figure 2.6: Plot of Observed & Estimated ASFR for Addis Ababa, 2011 DHS.

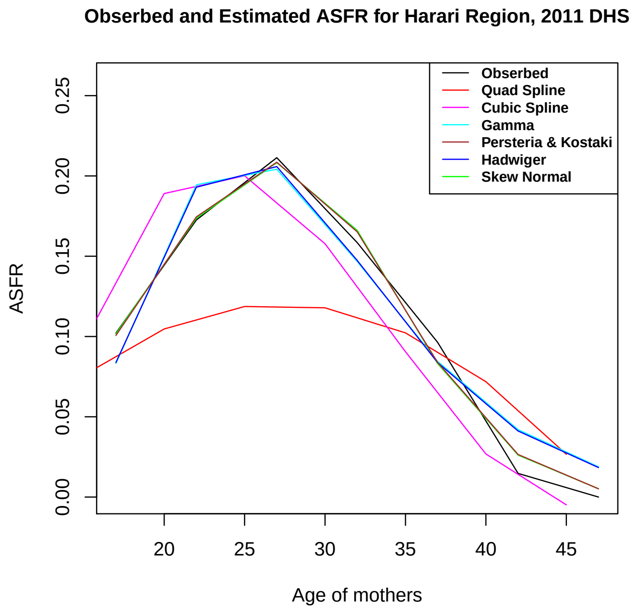


Figure 2.7: Plot of Observed & Estimated ASFR at Harari, 2011 DHS.

2.8 Conclusions

Results obtained from this preliminary analysis reveal that the values of the AIC for the proposed model, Skew Normal (SN), is lowest in the capital, Addis Ababa, Dire Dawa, Harari, Affar, Gambela, Benshangul-Gumuz and country level data as well. On the contrary, its value is higher some of the models and lower the rest on the remain regions, namely Tigray, Oromiya, Amhara, Somali and SNNP, indicating SN is not the best in fitting (capturing the available information on ASFR curve) in these region compared to the rest. However, since still the values of AIC for SN model is lowest not only at country level but also for majority regions, i.e., in 6 of the 11 regions , we conclude that the proposed model is better able to reproduce the empirical fertility data of Ethiopia and its regions than the other existing models considered. In the next two sections, we will present Skew Normal Distribution in-depth to estimating ASFR, which serve as the foundation for the hierarchical Bayesian model developed herein.

3 BAYESIAN MODELING

3.1 Introduction

A Micro-demographic variables, namely *fertility* is not only the main drivers of demographic dynamics but are also key elements to describe the behaviour of populations, both at national and regional levels. Both regional population forecasting and inter-regional comparisons to support policy making require the accurate data on such variables, and good model and its estimation method (Castro et al., 2015). As described in the previous

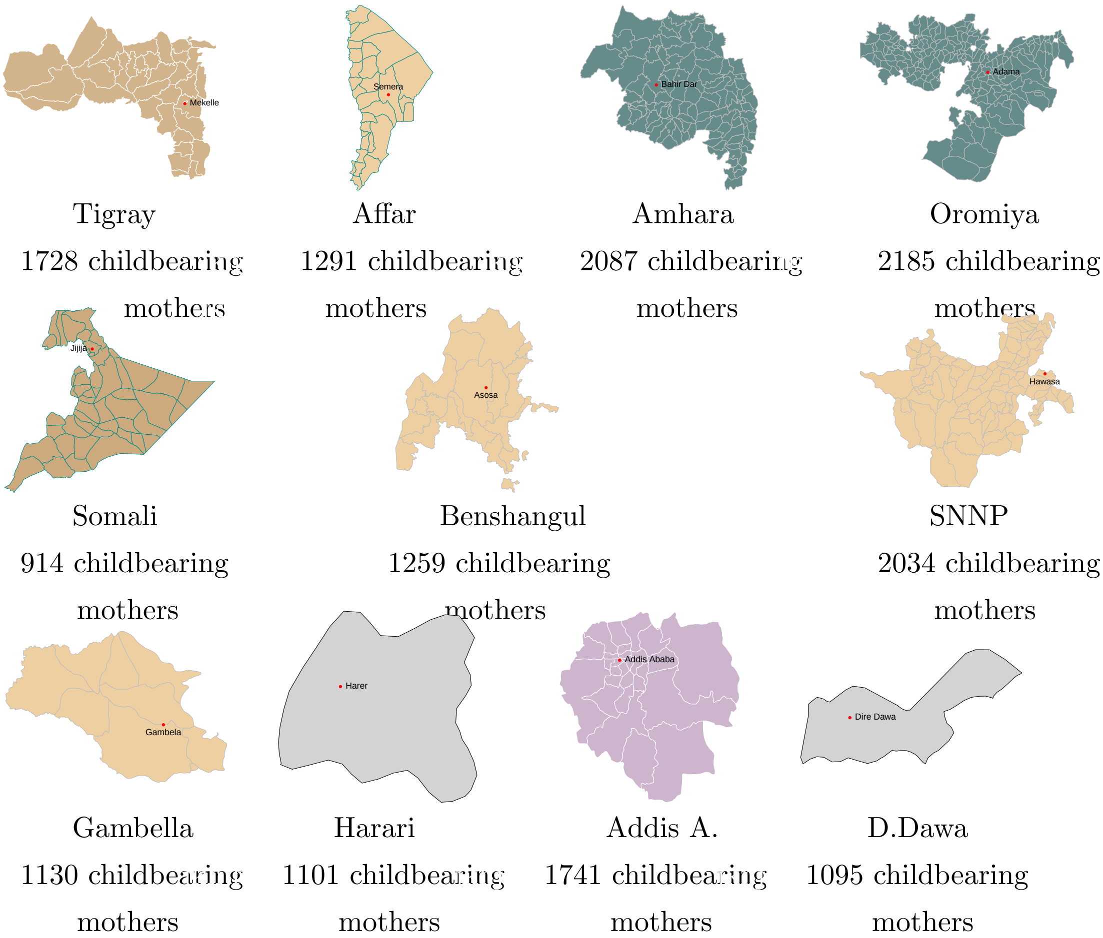


Figure 3.1: Ethiopian regions

chapter, this study uses the 2011 EDHS data amassed from different regions, possessing varying sample sample sizes, some large and others small number of observations (childbearing mothers), as expounded in Figure 3.1. The whole objective of this chapter is to make local-level inference and prediction on the fertility parameters of *Skew Normal model* (Mazzuco and Scarpa, 2015) proposed this far for fitting the Age-specific fertility rate data at country and regional levels in Ethiopian. Griffiths et al. (1987) pointed out that inference based on the conventional classical statistics via MLE, often leads to significant bias and hence, is not accurate in a situation where the subject under scrutiny has small sample size. Therefore, in this chapter Bayesian inference is primarily adopted to make inference and prediction on model parameters. By and large, Bayesian methods are a statistical approach for making inferences from data using probability models for quantities (parameters) underlying the study system (Gelman et al., 2014). Naturally, using Bayesian inference with this model would allow any additional information about the processes to be incorporated as prior information. Despite some intense debate between Bayesian supporters and "frequentists", Bayesian methods are increasingly popular among scientists of all fields, from medical (Spiegelhalter et al., 2004) and social (Snijders and Bosker) sciences, to ecology (Latimer et al., 2009) and fisheries (Millar, 2002; Fernández et al., 2002; Chen and Fournier, 1999; Fernández et al., 2010), among many others. Several reasons lead to the popularity of Bayesian methodologies. Most of those reasons lie within the nature of Bayesian inference as it accommodates the increasing need for maximizing all the information available (Gelman et al., 2014; McCarthy, 2007) and also the need for a more flexible and holistic approach to data analysis, including the estimation of uncertainties in key parameters that in conventional analyses are often forced to be constant for analytical tractability (Gelman et al., 2014). Additionally, Bayesian methods provide a natural framework for accounting for missing values without the need to rely on ad hoc imputation (Lunn et al., 2000). A review of this debate is beyond the scope of this thesis. However, a brief overview of the Bayesian framework and a discussion of its main advantages are given in order to clarify the choice of methodology used and give context to some of the analysis performed in this thesis.

To wrap up, as a new contribution of this chapter, we first developed a fertility model based on whether or not each mother involved in the study gives rise a birth at a specific age, *i.e.*, taking mother‘s birth status (*Have birth*, *No birth*) during a specific age as major classification criterion. Given the result obtained in previous chapter, we assume mother's birth status as a *Bernoulli random variable* having a probability distribution of *the proposed model*. Furthermore, the number of births at this particular age, which is a binary data with non-negative valued random variable, can be modeled using Binomial distribution having a probability, the value of proposed model at that age, and number of mother at that particular age as model parameters. Eventually, we estimated this model using a Bayesian approach using non-informative priors for the model parameters. Nevertheless, one stumbling block encounter in using this methodology was computational intractability. That is, the joint posterior distributions was non-linear, and too complex & intractable to easily drive the full conditional in standard/closed form. Data *Augmentation strategy* (*latent variable method*) has, hence, been instrumented as possible remedy in this respect. Therefore, as another new contribution, in this study a skew-normal (Mazzuco and Scarpa,

2015) latent variable methodology has been implemented in our Bayesian model developed from Binomially distributed fertility (birth) data so as to overcome this computational plight. Detail discussion of the model formulation and the use of Bayesian analysis on it is provided on subsection 3.5. Hence, the fertility data from each of the 11 regions are modeled using a Bayesian model with non-informative prior distributions, *i.e.*, the priors are constructed by assuming there is no information available about the process apart from the data. Furthermore, the maximum likelihood is discussed briefly to compare these two methods. The results and techniques developed in this section are used directly in analyzing the subsequent hierarchical Bayesian models.

In a nutshell, this chapter begins introducing about skew normal model and its family distribution in subsection 3.2. Following this, the fundamental concepts behind Bayesian inference and the computational methods used with it are recapped in subsection 3.3 and subsection 3.4 respectively. Formulation of models is one part of Bayesian inference. Thus, modeling ASFR in Bayesian Analysis is dealt in subsection 3.5. Furthermore, subsection 3.6 also introduces pertaining how to estimate model parameters and subsection 3.7 gives details of two MCMC techniques which are used in the remainder of this thesis. The results of this chapter are then discussed in subsection 3.8.

3.2 Skew Normal Model

3.2.1 More on Skew Normal Model

The normal distribution is symmetric and enjoys many important properties, such as its analytical simplicity, associated Central Limit Theorem, its multivariate extension-both the marginals and conditionals being normal, additivity and other properties. That is why it is widely used in practice. However, there are numerous situations, where the Gaussian distribution assumption may not be valid. Asymmetry in data is one situation where the normality assumption fails, and in such a condition, we need to device alternative models to fit the data. In literature, there are various near normal distributions that have been proposed by many authors (Dey, 2010). The skew-normal is one in this regard. Formally, first introduced by (Azzalini, 1985), the skew-normal (SN) distribution attracted a great deal of attention in the literature because of their flexibility in modeling skewed data, mathematical tractability and inclusion of the normal distribution as a special case (Azzalini and Capitanio, 2014). While the normal distribution with its symmetry has only location and scale parameters, the skew normal distribution has an additional shape parameter describing the skewness. From practical standpoint, this is a very desirable property, where in many real life situations, some skewness is always present in the data. In addition, the skew normal distribution shares many important properties of the normal distribution: for example, the skew normal densities are unimodal, their support is the real line, and the square of a skew normal random variable has the Chi-square distribution with one degree of freedom.

By definition, a random variable X is said to have a *Standard SN distribution* with shape or skewness parameter α, denoted by $X \sim \mathcal{SN}(\alpha)$, if its probability density function (pdf) is of the form:

$$g(x;\alpha) = 2\phi(x)\,\Phi(\alpha x), \quad x \in \mathbb{R}, \quad \alpha \in \mathbb{R} \tag{13}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the *probability density function(pdf)* and the *cumulative density function (cdf)* of standard normal distribution respectively. One of the benefits of this distribution is that the skewness can be introduced by a single parameter α. This parameter controls the shape of the distribution.

For instance, when $\alpha = 0$, $f(x;\alpha)$ corresponds to the standard normal distribution. Plots of the univariate density (44) for $\alpha = -5, -2, 0, 2, 5$, given in Figure 3.2, illustrate the effects of changing α on the shape of the density.

In general, a more flexible random variable can be built by incorporating location and scale parameters, ξ and ω, respectively such as

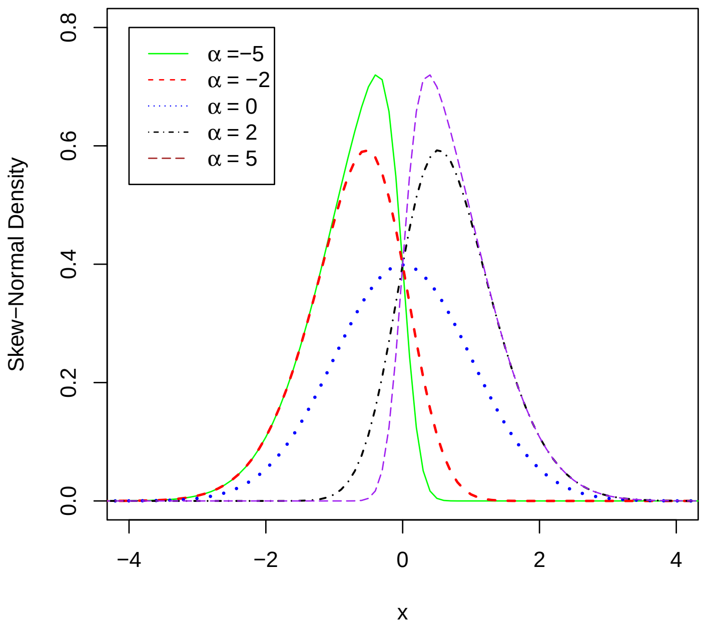


Figure 3.2: SN density function when $\alpha = -5, -2, 0, 2, 5$

$$Y = \xi + \omega X, \quad \text{where} \quad X \sim \mathcal{SN}(\alpha)$$

and this random variable Y is said to follow a skew normal distribution, written as $Y \sim \mathcal{SN}(\xi, \omega^2, \alpha)$, the density of which is given as:-

$$g(y;\theta) = g\left(y;\xi,\omega^2,\alpha\right) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\alpha\frac{y-\xi}{\omega}\right), \tag{14}$$

where, here again, as outlined thus far, ϕ and Φ are respectively the *pdf* and *cdf* of standard normal distribution; $y \in \mathbb{R}$; and $\theta = (\xi, \omega^2, \alpha)$ is vector of model parameters with

$\xi, \alpha \in \mathbb{R}$ and $\omega^2 > 0$ [8].

An alternative representation of the skew-normal that is especially popular in modeling Bayesian analysis is its stochastic representation discussed by Azzalini (1986); Henze (1986). The idea is that if $Z \backsim \mathcal{TN}_{[0,\infty)}(0,1)$ and $\varepsilon \backsim \mathcal{N}(0,1)$ are independent, and $\delta \in (-1,1)$, then the stochastic representation for the skew normal random variable X is given by

$$X_i = \delta Z_i + \sqrt{1-\delta^2}\,\varepsilon \tag{15}$$

and, analogously, the stochastic representation for the skew normal random variable Y, where $Y = \xi + \omega X$ for $X \backsim \mathcal{SN}(x;\alpha)$, $Z \backsim \mathcal{TN}(0,1)\mathbf{I}\{Z > 0\}$, $\varepsilon \backsim \mathcal{N}(0,1)$ with $Z \perp \varepsilon$, and $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$, is

$$Y_i = \xi + \omega X_i = \xi + \omega\left(\delta Z_i + \sqrt{1-\delta^2}\varepsilon\right) = \xi + \omega\delta Z_i + \omega\sqrt{1-\delta^2}\varepsilon \tag{16}$$

Some properties of this distribution includes:

$$E[Y] = \xi + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\omega \quad ; \tag{17}$$

and

$$V[Y] = \left\{1 - \frac{2\alpha^2}{\pi\left(1+\alpha^2\right)}\right\}\omega^2 \tag{18}$$

As far as this study is concerned, as outlined in subsubsection 2.4.1 and subsection 2.5 , a special case of this skew normal model of Mazzuco and Scarpa (2015) is applied to model one year age specific fertility data and thereby estimate the parameters in the model. The model in a more explicit way is:

$$\mathtt{f}\left(\mathtt{y};\xi,\omega,\alpha\right) = \frac{2\mathtt{R}}{\omega}\phi\left(\frac{\mathtt{y}-\xi}{\omega}\right)\Phi\left(\alpha\frac{\mathtt{y}-\xi}{\omega}\right) \tag{19}$$

where $\mathtt{f}\left(\mathtt{y}\right)$ is fertility rate at age y of a mother, $\mathtt{R}$ is the **TFR parameter** of the model, takes values in the interval $(\mathtt{0}, \mathtt{16})$. As paraphrased in subsubsection 2.4.1, ξ and α are the **location parameter** and the **shape parameter** of the model and they assume values in the interval $(-\infty, \infty) = \mathbb{R}$ while ω is the **scale parameter** of the model and it takes values in the interval $(\mathtt{0}, \infty)$. It should be noted that ξ is not the mean of the distribution (so it cannot be interpreted, as one might be tempted to do, as the average age at childbearing) but it is a function of it, as shown by (Arellano-Valle and Azzalini, 2008) for the skew normal. Similarly, ω is not the variance of the distribution but it is proportional to it (Mazzuco and Scarpa, 2015). Apparently, if $\alpha < \mathtt{0}$ the distribution has a negative skewness and if $\alpha > \mathtt{0}$ the skewness is positive.

[8] The square of ω^2 in $\mathcal{SN}\left(\xi,\omega^2,\alpha\right)$ is for analogy with ω^2 in the notation $\mathcal{N}\left(\xi,\omega^2\right)$

3.2.2 Unified Skew Normal Distribution (SUN)

The study of SN class of distribution has been a resumption of interest because of two reasons: first, it opened the door for robustness study. Second, it includes the normal density, and has very similar properties as that of normal density. However, because of the popularity of this class of distribution there have been intense developments in the theory of this class of distribution. Some among them are: the closed skew-normal (CSN) of Gonzalez-Farias et al. (Dominguez-Molina et al., 2003), the hierarchical skew-normal (HSN) of Liseo and Loperfido (Liseo and Loperfido, 2003) and the unified skew normal (SUN) of Azzalini Arellano (Arellano-Valle and Azzalini, 2006). With this view in mind, Arellano-Valle and Azzalini (2006) developed a skew normal model and named it unified skew normal model with the acronym SUN. They showed that this multivariate skew normal model includes or at least is equivalent to the earlier versions of skew normal models.

In general, suppose $(\boldsymbol{U_o}, \boldsymbol{U_1})$ is a multivariate normal vector of dimension $m+d$ with the density

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U_0} \\ \boldsymbol{U_1} \end{pmatrix} \backsim \boldsymbol{N_{m+d}(0, \Omega^*)}; \quad \boldsymbol{\Omega}^* = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Delta}^T \\ \boldsymbol{\Delta} & \overline{\Omega} \end{pmatrix}$$

where Ω^* is the correlation matrix, and $\boldsymbol{\Omega} = \boldsymbol{\omega}\overline{\Omega}\boldsymbol{\omega}$ is the covariance matrix with $\boldsymbol{\omega}$, a $d \times d$ diagonal matrix. Now, suppose Ω^* is positive definite and consider the distribution of $\boldsymbol{Z} = (\boldsymbol{U_1}|\boldsymbol{U_0} + \boldsymbol{\gamma} > \boldsymbol{0})$. Then the density of $y = \boldsymbol{\mu} + \boldsymbol{\omega} Z$ is

$$f(\mathbf{y}) = \phi_d(\boldsymbol{y} - \boldsymbol{\mu}; \boldsymbol{\Omega}) \frac{\Phi_m\left(\boldsymbol{\gamma} + \boldsymbol{\Delta}^T\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}[\boldsymbol{y} - \boldsymbol{\mu}]; \boldsymbol{\Gamma} - \boldsymbol{\Delta}^T\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\right)}{\Phi_m(\boldsymbol{\gamma}; \boldsymbol{\Gamma})} \tag{20}$$

for $\boldsymbol{y} \in \Re^{\boldsymbol{d}}$. The notation $\boldsymbol{\phi_d(y - \mu; \Omega)}$ is used to denote the d dimensional multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance Ω, $\boldsymbol{\Phi_d(y - \mu; \Omega)}$ denotes the corresponding distribution function. This density is called **SUN** (acronym for *unified skew normal distribution*) and is denoted by

$$\mathbf{y} \backsim \mathrm{SUN}_{d,m}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \overline{\boldsymbol{\omega}}, \boldsymbol{\Omega}^*)$$

, where $\boldsymbol{\Omega} = \boldsymbol{\omega 1_d}$.

Note that if Δ equal to zero, then the density reduces to the d dimensional multivariate normal distribution. The derivation of the **SUN** density was given in Arellano-valle and Azzalini (2006).

3.3 Basic Theory: Bayesian Inference

Statistics involves the collection, analysis and interpretation of data for the purpose of making statements or inferences about one or more physical processes that give rise to the data. *Statistical inference* concerns unknown parameters that describe certain population characteristics such as the true mean efficacy of a treatment for cancer or the probability of experiencing an adverse event. Inferences are made using data and a statistical model that links the data to the parameters. The statistical model might be very simple such that, for example, the data are normally distributed with some unknown but true population mean, μ say, and known population variance, σ^2 say, so that our objective is to make inferences about μ through a sample of data. In practice, statistical models are much more complex than this.

There are two main and distinct approaches to inference, namely *frequentist* and *Bayesian statistics*, although most people, when they first learn about statistics, usually begin with the frequentist approach (also known as the classical approach).

Bayesian inference is a process of fitting a probability model to the data set and summarizing the posterior probability distribution on model parameters and on unobserved quantities. Instead of producing maximum likelihood estimates for unknowns totally based on the sample data, Bayesian methods use the probability for quantifying uncertainty in inferences based on the statistical data analysis (Congdon, 2003).

A Bayesian analysis synthesizes two sources of information about the unknown parameters of interest. The first of these is the *sample data*, expressed formally by the *likelihood function (sampling distribution)*. The second is the *prior distribution*, which represents additional *(external) information* that is available to the investigator (Figure 3.3). If we

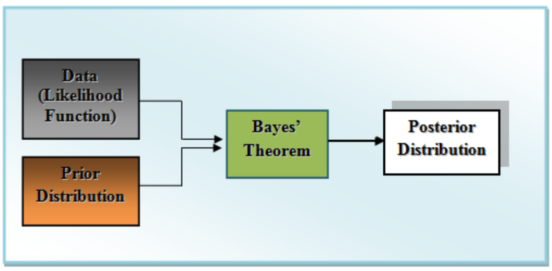


Figure 3.3: The Bayesian method.

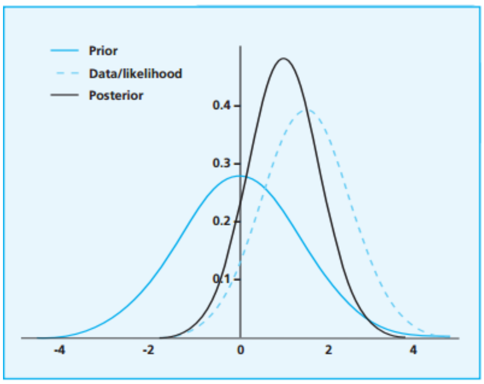


Figure 3.4: Example of a triplot.

represent the data by the symbol $\boldsymbol{y}$ and denote the set of unknown parameters by $\boldsymbol{\theta}$, then the likelihood function is $p(\boldsymbol{y}|\boldsymbol{\theta})$, the distribution of the data or the probability of observing the data $\boldsymbol{y}$ being conditional on the values of the parameter $\boldsymbol{\theta}$. If we further represent the prior distribution for $\boldsymbol{\theta}$ by $p(\boldsymbol{\theta})$, giving the probability that $\boldsymbol{\theta}$ takes any particular value based on whatever additional information might be available to the investigator, then, the joint distribution or full probability model can be written as

$$p(\boldsymbol{\theta}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{21}$$

and with the application of Bayes's theorem, an elementary result about conditional probability named after the **Reverend Thomas Bayes**, we have:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} \tag{22}$$

where $p(\boldsymbol{y})$ is the probability distribution of $\boldsymbol{y}$, called *marginal likelihood* for the data, $\boldsymbol{y}$, used as a *normalizing constant or normalization factor* to ensure the posterior density proper, *i.e.*, $\int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} = 1$ or $\sum_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = 1$ (Lynch, 2007) and is given by

$$p(\boldsymbol{y}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \text{ if } \theta \text{ is continuous, and}$$

$$p(\boldsymbol{y}) = \sum_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \text{ if } \boldsymbol{\theta} \text{ is discrete.}$$

Here, Equation 22, $p(\boldsymbol{\theta}|\boldsymbol{y})$ is called *posterior distribution of* $\boldsymbol{\theta}$, the distribution of the parameter after the data is observed. It expresses what is now known about $\boldsymbol{\theta}$ based on both the sample data and prior information.

Since marginal likelihood, $p(\boldsymbol{y})$, does not depend on $\boldsymbol{\theta}$, it can be considered as a constant with fixed $\boldsymbol{y}$. Therefore, Equation 23 can be given up to normalizing constant as,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{23}$$

that is, posterior distribution is *proportional to*[9] the product of the likelihood function and the prior distribution. Using numerical methods described in the next section, we can work with Equation 23 for model estimation and avoid computing the normalizing constant which is not easily obtained.

The posterior distribution for $\boldsymbol{\theta}$ is a weighted compromise between the prior information and the sample data. In particular, if for some value of $\boldsymbol{\theta}$ the likelihood in the right-hand side of Equation 23 is small, so that the data suggests that this value of $\boldsymbol{\theta}$ is implausible, then the posterior distribution will also give small probability to this $\boldsymbol{\theta}$ value. Similarly, if for some value of $\boldsymbol{\theta}$ the prior distribution in the right-hand side of Equation 23 is small, so that the prior information suggests that this value of $\boldsymbol{\theta}$ is implausible, then, again, the posterior distribution will also give small probability to this $\boldsymbol{\theta}$ value. In general, the posterior probability will be high for some $\boldsymbol{\theta}$ only when both information sources support that value.Therefore, the choice of prior distribution function will affect the posterior distribution function; thus we need to take into account all prior knowledge of the experiment and

[9] The symbol $\propto$ stands for "(directly)proportional to"

assign an appropriate prior distribution function(Gelman et al., 2014).

Within the Bayesian framework, *prediction* is possible through the *predictive distribution*, which describes how likely different outcomes of a future experiment are. Let $\tilde{\boldsymbol{y}}$ denote a future/new observation with probability density function $p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})$, and $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior distribution of θ given the data $\boldsymbol{y}$. Then,

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{y}) = \int_{\theta} p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\theta \tag{24}$$

is the predictive distribution of $\tilde{\boldsymbol{y}}$ given $\boldsymbol{y}$.

Bayesian inference offers several potentially attractive advantages over classical inference. *First*, it offers *more intuitive and meaningful inferences.* This is to say, Bayesian approach enables direct probability statements to be made about parameters of interest, whereas frequentist methods make indirect inferences by considering the data and more extreme but unobserved situations conditional on the null hypothesis being true; that is, p-values. *Second*, it gives the *ability to tackle more complex problems.* Statistical modeling can often generate quite complex problems and these can quickly become difficult to deal with or to construct exact test statistics from using a frequentist approach. Often it is necessary to rely on *large-sample approximations* by assuming **asymptotic normality**. In contrast, Bayesian inferences can be used to compute approximate values in highly complex situations; and *third*, it allows *the incorporation of prior information in addition to the data.* The use of prior information in addition to sample data is fundamental to the Bayesian approach. Prior information of some degree almost always exists and can make important contributions to strengthen inferences about unknown parameters and/or to reduce sample sizes. So, *if a suitable prior distribution can be specified, there are good reasons to choose Bayesian procedures.* One reason for rejecting these Bayesian procedures is the difficulty in *computing the integral* in Equation 24. This problem can be overcome by using simulation based techniques such as **Markov chain Monte Carlo (MCMC)** to simulate realizations of the posterior distribution. Estimates of the posterior distribution could then be obtained from the simulated sample.

3.4 Bayesian Computation: *Methods*

3.4.1 Rejection Sampling

Rejection sampling uses one distribution to sample from another by exploiting information relating the two; Von Neumann (1951) provided an algorithm for this method. Suppose that we want to sample from a *target distribution* F with probability density function $f(x)$. Suppose further that we can sample from a *proposal distribution* Q whose probability density function $q(x)$ can be scaled by a constant factor M to provide an upper bound on $f(x)$, *e.g.*, ***we might be able to scale a normal distribution so that our distribution of interest lies below it everywhere***. If we satisfy these requirements, then we can use rejection sampling to generate proposals from Q that we stochastically accept or reject according to the relative difference between $Mq(x)$ and $f(x)$. Specifically, to produce one sample:

Step-1: Generate a proposal x from the proposal distribution $q(x)$.

Step-2: Draw a sample u uniformly from the interval $[0, q(x)]$.

Step-3: If $u < (1/M) * f(x)$, accept x. Otherwise, reject x and return to Step 1.

Rejection sampling is most efficient in the limit where the scaled proposal density equals the target density, in which case all proposals are accepted. More generally, in expectation, this procedure accepts proposals at a rate given by

$$\int \frac{f(x)}{Mq(x)} dx \leq 1$$

Succinctly, the algorithmic representation of the *Accept-Reject method* is as follows

Algorithm 1:Accept-Reject Algorihm

Step-1: Generate x from $q(x)$

Step-2: Generate u uniform on $[0, q(x)]$

Step-2: If $u < (1/M) * f(x)$, then accept x. Otherwise reject and repeat

3.4.2 Markov Chain Monte Carlo (MCMC) Techniques

In complicated Bayesian models, it is often not easy to obtain the posterior distribution analytically. This analytic bottleneck has been eliminated by the emergence of Markove Chain Monte Carlo (MCMC). MCMC simulates the posterior distribution $p(\theta|y)$, and it is used when it is not possible to sample directly from $p(\theta|y)$. It works by simulating direct draws from probability distributions of interest and then correcting those draws to better approximate the target posterior distribution $p(\theta|y)$. The samples are drawn sequentially and the previous sample values are used to randomly generate the next sample value (hence, forming a Markov chain). This technique is successful because the approximate distributions are improved at every step until convergence to the unknown target distribution (Gelman et al., 2014). As such, in order to achieve full convergence of the

MCMC approximation to the true posterior distribution, it is common to discard the first few samples (a process known as burn-in).

The Gibbs sampler, which is a particular Markov chain algorithm designed to sample from a high dimensional distribution (or joint probability distribution), is crucial in the Bayesian model fitting process because $p(y)$ is extremely difficult to calculate or approximate directly. However, the Gibbs sampler allows us to generate a sample from the parameters of the conditional distributions which will be proportional to, but does not require $p(\theta|y)$. Therefore, the power of Gibbs sampling is that the joint distribution of the parameters will converge to the joint probability of the parameters given the observed data. This means that the Gibbs sampler finds estimates for the parameters of interest in order to determine how well the observable data fits the model of interest, and also whether or not data independent of the observed data fits the model described by the observed data.

MCMC techniques provide a way of simulating from complex distributions by simulating from Markov chains which have the target distributions as their stationary distributions. There are many MCMC techniques of which two are described below. The specific details of how these techniques are used will be given later. There is much literature available on the theory behind MCMC techniques and on applications of the techniques. Introductions to the area are provided by Besag et al. (1995), (Gamerman) and Besag (2001). Below,we outline the specific steps for sampling the fertility parameters.

3.4.3 The Gibbs Sampler

The Gibbs sampler was used by Gelman et al. (2003) for models with the Gibbs distribution and was extended to the general form given here by Gelfand and Smith (1990). The Gibbs sampler enables simulation from multivariate distributions by simulating only from the conditional distributions. So, suppose the density of interest is $p(\theta)$, where $\theta = (\theta_1, \cdots, \theta_d)'$ and the full conditionals are given by

$$\pi(\theta_i|\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_d) = \pi(\theta_i|\theta_{-i}), \quad i = 1, \cdots, d \tag{25}$$

If it is possible to simulate from the full conditionals then the Gibbs sampler can be used by using the following algorithm:

1. Initialise the counter to $j = 1$ and the state of the chain to $\theta^0 = (\theta_1^0, \cdots, \theta_d^0)'$

2. Obtain a new value θ^j from $\theta^{(j-1)}$ by successive simulation from the full conditionals:

$$\begin{aligned}
&\theta_1^j \backsim \pi(\theta_1|\theta_2^{(j-1)}, \cdots, \theta_d^{(j-1)}) \\
&\theta_2^j \backsim \pi(\theta_2|\theta_1, \theta_3^{(j-1)}, \cdots, \theta_d^{(j-1)}) \\
&\quad\vdots \\
&\theta_d^j \backsim \pi(\theta_1|\theta_2^{(j-1)}, \cdots, \theta_{d-1}^{(j-1)})
\end{aligned}$$

3. Increase counter from j to $j+1$ and return to step 2.

If it is possible to simulate from the full conditionals of the posterior distribution Equation 22, then it is also possible to simulate from the posterior itself. The Gibbs sampler should be run after initializing the sampler somewhere in the support of θ. The resulting chain will converge, after an initial "burn-in" period, to the posterior distribution.

3.4.4 Metropolis-Hastings Sampling

The Gibbs sampler provides a way of simulating from multivariate distributions provided that the full conditional distributions can be simulated from. It may not be straightforward to simulate from these full conditionals but Metropolis-Hastings schemes provide a way. These schemes come from work by Metropolis et al. (1953) and Hastings (1970). Given a distribution of interest, f, a reversible Markov chain, which has this distribution as its stationary distribution, can be constructed. Simulating from such a Markov chain will result in values from the distribution of interest.

The procedure is to construct a transition kernel $p(\theta, \phi)$ such that the equilibrium distribution of the chain is f. This transition kernel is made up of two elements; an arbitrary transition kernel $q(\theta, \phi)$ also known as the *proposal distribution*, and an acceptance probability $\rho(\theta, \phi)$. The acceptance probability

$$\rho(\theta, \phi) = \min\{\mathbf{R}, 1\} = \min\left\{\frac{\mathtt{f}(\phi)}{\mathtt{f}(\theta)}\frac{\mathtt{q}(\theta|\phi)}{\mathtt{q}(\phi|\theta)}, 1\right\}$$

was suggested by Hastings (1970). The algorithm below can then be followed to obtain a chain with limiting distribution f.

The particular type of MCMC method used in this thesis is based on simulation of a *random walk chain*. Here, the proposed value ϕ at point j is $\phi = \theta^{(j-1)} + \omega_j$. The ω_j are IID random variables and have density $f(\cdot)$. Supposing $f(\cdot)$ is easy to simulate from, an *innovation*, ω_j, can be simulated. The *candidate* point is then set to $\phi = \theta^{(j-1)} + \omega_j$ and the transition kernel is given by $q(\theta, \phi) = f(\phi - \theta)$. This is then used to calculate the acceptance probability. The variance of the innovation affects the acceptance probability: if the variance is too low most proposals will be accepted, resulting in very slow convergence, and if it is too high very few will be accepted and the moves in the chain will often be large

Algorithm 2:Metropolis-Hastings

- $\surd$ **Step-1:** Start with ***arbitrary*** $\theta^{(0)}$ from the support of **target distribution**, f

- $\surd$ **Step-2:** At stage $\mathtt{j}, \mathtt{j} = 1, \cdots, \mathtt{Nsim}$, generate **proposal value**, $\phi^{\mathtt{j}}$ from the ***proposal density*** $\mathtt{q}$, given the **current state** $\theta^{(\mathtt{j})}$. That is,

$$\phi^{\mathtt{j}} \backsim \mathtt{q}\left(\phi^{\mathtt{j}}|\theta^{(\mathtt{j})}\right),\ \mathtt{t} = 1, \cdots, \mathtt{Nsim}$$

- $\surd$ **Step-3:** Take, then, $\theta^{[\mathtt{j}+1]} = \phi^{\mathtt{j}}$ (*meaning:* if we accepted, take the next value, $\theta^{[\mathtt{j}+1]}$ being the proposed value, $\psi^{\mathtt{j}}$) with probability of $\rho(\theta^{(\mathtt{j})}, \phi^{\mathtt{j}})$. Otherwise, take $\theta^{[\mathtt{j}+1]} = \theta^{(\mathtt{j})}$ (*meaning:* if we don't accepted, take the next value, $\theta^{[\mathtt{j}+1]}$ being the current value, $\theta^{(\mathtt{j})}$), that is:-

$$\theta^{[\mathtt{j}+1]} = \begin{cases} \phi^{\mathtt{j}} & \text{with probability of } \rho(\theta^{(\mathtt{j})}, \phi^{\mathtt{j}}) \\ \theta^{(\mathtt{j})} & \text{with probability of } 1 - \rho(\theta^{(\mathtt{j})}, \phi^{\mathtt{j}}) \end{cases}$$

where,

$$\rho(\theta, \phi) = \min\{\mathbf{R}, 1\} = \min\left\{\frac{\mathtt{f}(\phi)}{\mathtt{f}(\theta)}\frac{\mathtt{q}(\theta|\phi)}{\mathtt{q}(\phi|\theta)}, 1\right\}$$

, called **MH-Acceptance Probability**.*This ensures that our probability is a number between 0 and 1*, and

$$\mathbf{R} = \frac{\mathtt{f}(\phi)}{\mathtt{f}(\theta)}\frac{\mathtt{q}(\theta|\phi)}{\mathtt{q}(\phi|\theta)}$$

also called **Acceptance Rate**. This random acceptance is done by generating a uniform on $(0, 1)$ random variable U and accepting the proposal ϕ^{j} if $U \leq \rho(\theta^{(\mathtt{j})}, \phi^{\mathtt{j}})$

- $\surd$ **Step-4:** Increase n and return to **Step-2**.

3.5 Model Development:Bayesian Inference of Binary Fertility Data using Skew Normal Latent Variable

The aim of this section is to develop fertility model for which inference and learning can be carried out in a unified way, without the need for making potentially conflicting assumptions. Given the status of mothers(*Have birth*, *No birth*) in the study, one can develop a new model which use to estimate the fertility parameters of the model proposed in the previous chapter. Therefore, in this section, a Binomial model of number of mothers, n_x, who gave rise a birth at particular age x with fertility probability of π_x, where, $\pi_x = R \cdot SN(x; \xi, \omega^2, \alpha)$, during the study period is formulated for each of the 11 regions for use along with appropriate non-informative priors in our Bayesian analysis. Nevertheless, one stumbling block encounter in using this methodology was computational intractability. That is, the joint posterior distributions was non-linear, and too complex & intractable to easily drive the full conditional in standard/closed form. Data *Augmentation strategy* (*latent variable method*) has, hence, been instrumented as possible remedy in this respect. This section starts with a detailed discussion of developing Binomial based model. This is followed by detailed description of the choice of priors subsubsection 3.5.2 with particular reference to issues that arise in analyses involving fertility data. Discussions on formulations of model likelihood and evaluation of the resulting posterior distribution will be addresses respectively in subsubsection 3.5.3 and subsubsection 3.5.4 methods for its computation (3.4.3). Further issues that arise in Bayesian implementations such as the use of data augumentation approach and its computational mechanism are outlined in the next consecutive sections. Eventually, discussion on the results obtained is presented on subsection 3.8.

3.5.1 Model Specification: *Modeling Fertility Data*

To begin with, consider a woman, i, found in the year **[x,x+1)**, in which case $\{x : 15 \leq x \leq 49\}$, is mother's age, and $i = 1, 2, \ldots, n$, with n is the total number of mothers of year 15-49 in the study. On the top of this, we assume a random variable $W_i(x)$ in $\{0, 1\}$ in such a way that $W_i(x)$ denotes **1 if the i^{th} woman of age 15-49 (reproductive age) gives birth to a child during the year [x,x+1)**, and **0 otherwise**. Symbolically, this indicator random variable W_i is then explicated as:

$$W_i(x) = \begin{cases} 1 & \text{if the } i^{th} \text{ woman of age 15-49 gives} \\ & \text{birth to a child during the year } [x, x+1), \\ 0 & \text{elsewhere} \end{cases} \tag{26}$$

for $i = 1, 2, \ldots, n$. In the same fashion, if we further suppose $\pi_i(x)$ is the corresponding ***fertility probability*** of success at this particular age x, *i.e.*, the probability that **a woman in the reproductive age has given birth to a child** during **the year [x,x+1)**, (say: when a mother is 24 years old), then $(1 - \pi_i(x))$ will be the ***probability*** when **this woman failed to deliver** during **this stated year,[x,x+1)**. This is to say, we have:-

$$\begin{aligned} P\left(W_i(x) = 1 | X = x\right) &= \pi_i(x), \\ P\left(W_i(x) = 0 | X = x\right) &= 1 - \pi_i(x) \end{aligned} \tag{27}$$

Taking Equation 26 and Equation 27 all together, it can be shown that the status of mother W_i, for $1 \leq i \leq n$, or $W_1, \cdots, W_n$, are *independent and identically* **Bernoulli Distributed**

Random Variable[10] with **fertility probability** π_i for the **value 1** and $(1-\pi_i)$ for the **value 0**, *i.e.*,

$$W_i(x) \backsim \mathsf{Bernoulli}\,(\pi_i(x)) \tag{28}$$

for $i = 1, 2, \ldots, n$, as shown being recapped in Table 3.1 below. Moreover, if we take a closer look at the vector $\mathbf{w} = (w_1, w_2, w_3, \cdots, w_n)$ of all indicator variables, and assume independence among the status of mothers, W_i, the distribution of $\mathbf{W}$ can be written as

$$P\,(W = w|X = x) = \prod_{i=1}^{n} \pi_x^{w_i}\,(1-\pi_x)^{1-w_i} \tag{29}$$

Table 3.1: **Total number of women in the reproductive age group and their corresponding indicator function**

Women	Indicator function	Pr. of giving birth at age $[x, x+1)$
1	$W_1(x) = \begin{cases} 1 & \text{if the } 1^{st} \text{ woman has} \\ & \text{birth in the year } [x, x+1) \\ 0 & \text{elsewhere} \end{cases}$	$\begin{cases} \pi_1(x) & \text{if } W_1 = 1 \\ 1-\pi_1(x) & \text{if } W_1 = 0 \end{cases}$
$\vdots$	$\vdots$	$\vdots$
i	$W_i(x) = \begin{cases} 1 & \text{if the } i^{th} \text{ woman has} \\ & \text{birth in the year } [x, x+1) \\ 0 & \text{elsewhere} \end{cases}$	$\begin{cases} \pi_i(x) & \text{if } W_i = 1 \\ 1-\pi_i(x) & \text{if } W_i = 0 \end{cases}$
$\vdots$	$\vdots$	$\vdots$
n	$W_n(x) = \begin{cases} 1 & \text{if the } n^{th} \text{ woman has} \\ & \text{birth in the year } [x, x+1) \\ 0 & \text{elsewhere} \end{cases}$	$\begin{cases} \pi_n(x) & \text{if } W_n = 1 \\ 1-\pi_n(x) & \text{if } W_n = 0 \end{cases}$

Without loss of generality, if every woman of the same age x (all women at age 24, for instance) has the same probability of giving birth to a child, say $\pi_i(x)$, then it is useful to summarize these women of equal age x and to introduce the **scalar random variable** $W_x = \sum_{x=j}^{n_x} W_j(x)$, the sum of n_x - independent Bernoulli random variables, denoting *the number of women delivered births during this specified age,* x. In doing so, we assume the total number of women at age x by n_x, as outlined previously and in the Table 3.2 as well, and since W_x is, as explained, a sum of n_x-independent Bernoulli random variables at age x, then it follows that W_x is *Binomially distributed random variable* with parameters n_x and π_x, *i.e.*, $W_i(x) \sim \mathsf{Bin}\,(n_x, \pi_x)$. Thus,

$$p\,(W_x = d_x|X = x) = p\,(W_x = d_x) = p\,(d_x) = \binom{n_x}{d_x} \pi_x^{d_x}\,(1-\pi_x)^{n_x-d_x} \tag{30}$$

[10] The Bernoulli random variable Y , is one with binary outcomes chosen from $0, 1$ and its probability density function is $f_Y(y) = p^y(1-p)^{1-y}$, where $p = P(Y = 1)$ is the pr.of success

Table 3.2: **Total number of women in the reproductive age group and their corresponding indicator function**

Age[a]	TW at age x[b]	R.Variable[c]	TWDB at age x[d]	Pr.[e]
15	n_{15}	$W_1(15), \cdots, W_{n_{15}}(15)$	$W_{15} = \sum_{x=j}^{n_{15}} W_j(15) = d_{15}$	π_{15}
16	n_{16}	$W_1(16), \cdots, W_{n_{16}}(16)$	$W_{16} = \sum_{x=j}^{n_{16}} W_j(16) = d_{16}$	π_{16}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x	n_x	$W_1(x), \cdots, W_{n_x}(x)$	$W_x = \sum_{x=j}^{n_x} W_j(x) = d_x$	π_x
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
49	n_{49}	$W_1(49), \cdots, W_{n_{49}}(49)$	$W_{49} = \sum_{x=j}^{n_{49}} W_j(49) = d_{49}$	π_{49}
	$n = \sum_{x=15}^{49} n_x$	$\mathbf{W}$	$\mathbf{d}$	

[a] Women's age $x, 15 \leq x \leq 49$; [b] Total number of women at age x; [c] Random Variables at age x; [d] Total number of women delivered births at age x; [e] Probability of giving birth at age x

Further more, if $\mathbf{d}$ is a vector with elements d_x, $x = 0, 1, \ldots, n$, where n is the total number of mothers in the study, then the distribution in Equation 30 can be extended to account the distribution of $\mathbf{d}$ and this joint probability is described by:-

$$p(\mathbf{W} = \mathbf{d}) = p(\mathbf{d}) = \prod_{x=0}^{n} \binom{n_x}{d_x} \pi_x^{d_x} (1 - \pi_x)^{n_x - d_x} \tag{31}$$

Donating the fertility intensity for age x as f_x, the skew normal curve is usually defined as

$$f_x\left(x; \xi, \omega^2, \alpha, R\right) = R \cdot g\left(x; \xi, \omega^2, \alpha\right) = 2R\omega^{-1}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right) \tag{32}$$

3.5.2 Prior Specification/ Elicitation

Prior elicitation is one of the most crucial issues in Bayesian data analysis. It is the most debated topic in theoretical research and is also a challenging issue to practitioners. Opponents of Bayesian approach criticize the arbitrariness in the choice of prior, whereas proponents praise it as a manageable way of introducing flexibility in Bayesian analysis (Kass and Wasserman, 1996) . Berger (1985) noted that whenever a practitioner can summarize historical or subjective information about the unknown parameter, an informative prior should be used. On the other hand, more often either historical or subjective information is unavailable, or incorporating such information into a prior distribution is difficult for a real problem, thus noninformative prior[11] distributions are needed. Bayesian analysis with noninformative priors preserves the appearance of objectivity, and is being increasingly recognized by classical statisticians.

Therefore, to perform a Bayesian analysis, we first have to select a prior for the model parameters. It should be noted that, in general, the prior distribution has to be selected carefully in the Bayesian modeling. First of all, it is not possible to choose an improper prior because this leads to an improper posterior density (see,e.g. Frühwirth-Schnatter, 2006, 2006, Section 3.2). Furthermore, as noted by Jennison (1997), one should avoid trying to be as "*noninformative as possible*" by *choosing large prior variances* because the choice of the prior of the parameters strongly affects the posterior. For this reason, we adopt the priors introduced by Canale and Scarpa (2013) in the context of students' performance in university examinations. Canale and Scarpa (2013) suggested that proposed a joint prior distribution for the location, scale and shape parameters in the case of skew normal model. Therefore, we introduce two noninformative prior distributions for the scalar shape parameter α. The first is simply a *normal* and the second proposal is a skew-normal distribution since from previous studies we came to understand the shape of ASFR follows in most cases either of the patterns mentioned.

However, to specify priors for the *location* and *scale* parameter, ξ and ω^2 respectively, we used the *stochastic representation* considered in Equation 16. Such representation offers several advantages. First, a conditionally conjugate prior[12] for ξ and ω^2 is available (see subsection 3.6) and, second, straightforward estimation using a Gibbs sampler becomes feasible. Therefore, it allows to us to sample those parameters jointly from a closed-form posterior. Eventually, we make use of the demographic interpretation to assume a *gamma prior* for the total fertility rate parameter, R. In conclusion, the priors for the parameters and hyperprior for the hyperparameters are given as follows.

(a). Prior for R:

$$R \sim \text{Gamma}\,(a, b)\,,\ \text{where } a, b > 0$$

(b). Prior for ξ:

$$\xi \sim \text{N}\left(\xi_o, \kappa\omega^2\right),\ \text{where } \xi_o \in \Re \text{ and } \omega^2 > 0$$

[11] *Noninformative priors*: are also called *automated priors, default priors, vague priors*, or *priors of ignorance*

[12] P, a family of prior distributions, is a *conjugate prior* of F, a family of likelihood functions, if for any likelihood function $f \in F$ and for any prior distribution $p \in P$, the corresponding posterior distribution p^* satisfies $p^* \in P$,i.e., the posterior is the same family of the prior, and moreover we have an explicit formula for for the posterior hyperparameters. The conjugacy is popular because of its mathematical ease, once the conjugate pair likelihood/prior is found, the posterior is found easily

(c). Prior for ω^2:

$$\omega^2 \sim \text{InvGamma}\,(c, d)\,, \text{ where } a, d > 0$$

(d). Prior for α: we assume two prior cases

(i). Case-1: when α is supposed to to be normally distributed parameter, *i.e.*,

$$\alpha \sim \text{N}\left(\alpha_o, \psi_o^2\right), \quad \text{or} \quad , \text{ where } \ \alpha_o \in \Re \text{ and } \ \psi_o^2 > 0$$

(ii). Case-2: when α is supposed to have a skew normal distribution, *i.e.*,

$$\alpha \sim \ \text{SN}\left(\alpha_o, \psi_o^2, \lambda_o\right), \text{ where } \ \alpha_o, \lambda_o \in \Re \text{ and } \ \psi_o^2 > 0$$

The hyperparameters in our model a, b,ξ_o,c, d, α_o,ψ_o, and λ_o are all assumed known.

3.5.3 Model Likelihood Function

As outlined in Equation 31 above, we have n independent binary random variables, $W_1.\cdots, W_n$ are observed, where mother's status, W_i, is distributed Bernoulli with fertility probability,π_x. That is:

$$d_x|\pi_x \sim \text{Bin}\,(n_x, \pi_x) \tag{33}$$

for d_x, $x = 0, 1, \ldots, n$ .Therefore, the model likelihood is given as

$$p\,(\mathbf{d}|\pi) \;=\; \prod_{x=0}^{n} \binom{n_x}{d_x} \pi_x^{d_x}\,(1-\pi_x)^{n_x - d_x} \tag{34}$$

Taking the fact that:

$$\begin{aligned} \pi_x = \widehat{f_x}\left(x; \xi, \omega^2, \alpha, R\right) &= R \cdot \ \ g(x; \xi, \omega^2, \alpha, R) \\ &= 2R\omega^{-1}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right) \in [0,1] \end{aligned} \tag{35}$$

we can re-write the model likelihood in Equation 34 as

$$\begin{aligned} p\left(\mathbf{d}|\xi, \omega^2, \alpha, R\right) = \Bigg(\prod_{x=0}^{n}\binom{n_x}{d_x}\left[R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]^{d_x} \\ \left\{1 - \left[R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]\right\}^{n_x - d_x}\Bigg) \end{aligned} \tag{36}$$

Note that both the right and the left hand sides of the Equation 36 assume values in the interval $[0, 1]$as π_x is a probability value and hence is in the interval$[0, 1]$, and this holds for the right hand side expression as well for it is the product of a $R \in [0, 1]$ and the density $g(x; \xi, \omega^2, \alpha, R) \in [0, 1]$.

3.5.4 Posterior Distribution

The posterior distribution summarizes the current state of knowledge about all the *uncertain quantities* (including *unobservable parameters* and also *missing, latent, and unobserved potential data*) in a Bayesian analysis (Gelfand, 2014). Analytically, the posterior density is the product of the prior density and the likelihood. Thus, in our case, denoting $p(\theta^*)$ the prior distribution for θ^*, where $\theta^* = (R, \xi, \omega^2, \alpha)$, it follows that for $\mathbf{d} = d_x, x = 0, 1, \ldots, n$, the resulting posterior distribution is of the form:-

$$\begin{aligned} p\left(R, \xi, \omega^2, \alpha|\mathbf{d}\right) = &\left(\prod_{x=0}^{n} \binom{n_x}{d_x} \left[R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]^{d_x}\right. \\ &\left.\left\{1-\left[R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]\right\}^{n_x-d_x}\right) \\ &\times p\left(R, \xi, \omega^2, \alpha\right) \end{aligned} \tag{37}$$

Assuming prior independence of model parameters, this joint posterior distribution can be re-written as in what follows:

$$\begin{aligned} p\left(R, \xi, \omega^2, \alpha|\mathbf{d}\right) &\propto p\left(\mathbf{d}|R, \xi, \omega^2, \alpha\right) \cdot p(R) \cdot p(\xi) \cdot p\left(\omega^2\right) \cdot p(\alpha) \\ &= \left(\prod_{x=0}^{n} \binom{n_x}{d_x} \left\{R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right\}^{d_x}\right. \\ &\left.\left\{1-\left[R \cdot \ \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]\right\}^{n_x-d_x}\right) \\ &\times R^{a-1}e^{-bR} \times \frac{1}{\omega}e^{-\frac{1}{2\kappa\omega^2}(\xi-\xi_o)^2} \times \left(\omega^2\right)^{-c-1} e^{-d/\omega^2} \times p_i(\alpha) \end{aligned} \tag{38}$$

, where *i=1, 2* and, hence:

$$p_1(\alpha) \propto \phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right), \text{and } p_2(\alpha) \propto \phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)\Phi\left(\lambda_o\frac{\alpha-\alpha_o}{\psi_o}\right)$$

, which is complex and intractable.

3.6 Bayesian Inference using Data Augmentation Approach

As seen in Equation 38, the joint posterior distribution, $p(R, \xi, \omega^2, \alpha|\mathbf{d})$, is nonlinear with respect to the model parameters, $\theta^* = (R, \xi, \omega, \alpha)$. Thus, no matter what prior structure we choose for θ^*, the conditional distributions are intractable, difficult to evaluate or derive; and even if derived, none of them are in closed or standard functional forms (e.g. gamma, multivariate normal, etc). Consequently, the Gibbs sampler cannot be used directly to draw samples from the given distributions and the resulting conditional distributions are,therefore, difficult or impossible to simulate.

Recent advances in Bayesian simulation, however, have shown that Gibbs sampling algorithms based on the *method of data augmentation* can overcome such computational plights and provide reliable model fitting (Geweke et al., 1994). In what follows, we, therefore, introduce n independent latent variables, $Z_1, \cdots, Z_n$ into the problem, where Z_i is distributed $R \cdot SN(\xi, \omega^2, \alpha)$. The advantage of the strategy is straightforward: those principal observations, combined with the tool of Gibbs Sampling, will allow us to simulate from the posterior distribution of, $\theta^* = (R, \xi, \omega^2, \alpha)$.

3.6.1 Data Augmentation Approach: *using Special SN latent variable model*

The term data augmentation (DA) refers to methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables. It was first proposed by (Tanner and Wong, 1987), and is an important development in the field of Markov Chain Monte Carlo algorithms. As mentioned, when it is combined with the pioneer works of (Metropolis et al., 1953) and (Hastings, 1970), it makes the Bayesian analysis of more complex models possible and suitable for Gibbs Sampler. In principle, different augmented-data models can be used to construct different DA algorithms with the different properties. In this work, we choose the Skew normal model (Mazzuco and Scarpa, 2015) proposed so far as latent variable model to derive the full conditionals from the augmented joint posterior model and implement the Gibbs sampling with ease.

It is easy to show that the Bernoulli random variable given in Equation 26, *i.e.*, the status of the i^{th} childbearing age mother (*Have birth*; *No birth*), where, $i = 1, 2, \cdots, n$; n, total number of childbearing age mothers in the study,

$$W_i(x) \sim \text{Bern}\left(\pi_i(x)\right) \text{ where } \pi_i(x) = \Pr\left(W_i(x) = 1 | R, \xi, \omega^2, \alpha\right) = R \cdot \text{SN}(x; \xi, \omega^2, \alpha) \quad (39)$$

is equivalent to considering that

$$W_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i \leq 0 \end{cases} \quad (40)$$

where $\boldsymbol{z} = (z_1, \cdots, z_n)'$ is a vector of unknown latent (*unobserved/auxiliary*) variable corresponding to $\boldsymbol{W}$, distributed as:

$$z_i \sim R \cdot \ \text{SN}(\xi, \omega^2, \alpha) \quad (41)$$

for $i = 1, 2, \cdots, n$. As to this study is concerned, this auxiliary variable has no any practical meaning, yet adopted here in our model for the sake of computational convenience, *i.e.*, so as to avoid working with observed data likelihood model in Equation 37, which led to intractable posterior distribution in Equation 38.

From Equation 41, it is apparent the sampling distribution of z_i conditional on R, ξ, ω, and α is:-

$$p\left(z_i | R, \xi, \omega^2, \alpha\right) \propto R \cdot \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right)$$

and, this, in turn, yields its joint density of the form:-

$$p\left(\boldsymbol{z}|R,\xi,\omega^2,\alpha\right) \propto \prod_{i=1}^{n} R\cdot\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right) \tag{42}$$

As for conditional for W_i in Equation 40 given R,ξ,ω,α,z_i, *viz.*, $p\left(\mathrm{w}_i|z_i,R,\xi,\omega,\alpha\right)$, it is essential first to recall and note that when $z_i > 0$, then W_i will equal to 1 while when $z_i \leq 0$, then W_i assumes a value zero. In other words, the sign of z_i perfectly predicts the value of W. Hence, we write the conditional for W as:-

$$p\left(\mathrm{w}_i|z_i,R,\xi,\omega^2,\alpha\right) = \mathbb{I}\left(z_i>0\right)\mathbb{I}\left(W_i=1\right)+\mathbb{I}\left(z_i\leq 0\right)\mathbb{I}\left(W_i=0\right) \tag{43}$$

with $\mathbb{I}$ denoting the indicator function which assumes the value 1 if the statement on the parenthesis is true, and 0,otherwise. Then, the joint conditional distribution is

$$p\left(\boldsymbol{w}|z,R,\xi,\omega^2,\alpha\right) = \prod_{i=1}^{n}\left\{\mathbb{I}\left(z_i>0\right)\mathbb{I}\left(W_i=1\right)+\mathbb{I}\left(z_i\leq 0\right)\mathbb{I}\left(W_i=0\right)\right\} \tag{44}$$

Thus, we can now "augment" the observed data $\boldsymbol{w} = (w_1,\cdots,w_n)'$ with latent data $\mathbf{z} = (z_1,\cdots,z_n)'$ to form the *complete-data* or *augmented-data vector*,$(\boldsymbol{w},\mathbf{z})$,and its likelihood as provided in the subsequent subsection.

3.6.2 Complete-data Likelihood Model

As the name indicates, putting the pieces in Equation 42 and Equation 44 together , we obtain the likelihood or the joint density of the complete data vector, $\{\mathbf{w},\mathbf{z}\}$, referred to *complete-data likelihood function* or *augmented data joint density*, $L\left(R,\xi,\omega^2,\alpha|\boldsymbol{w},\boldsymbol{z}\right)$, which is given as:-

$$\begin{aligned}
L\left(R,\xi,\omega^2,\alpha|\boldsymbol{w},\boldsymbol{z}\right) &= p\left(\boldsymbol{w},\boldsymbol{z}|R,\xi,\omega,\alpha\right) \qquad => \text{In general, } L\left(\theta|y,z\right)=p\left(y,z|\theta\right)=p\left(y|z,\theta\right)p\left(z|\theta\right)\\
&= p\left(\boldsymbol{w}|\boldsymbol{z},R,\xi,\omega^2,\alpha\right)p\left(\boldsymbol{z}|R,\xi,\omega^2,\alpha\right)\\
&= \prod_{i=1}^{n}\left\{\mathbb{I}\left(z_i>0\right)\mathbb{I}\left(W_i=1\right)+\mathbb{I}\left(z_i\leq 0\right)\mathbb{I}\left(W_i=0\right)\right\}\\
&\cdot\prod_{i=1}^{n}\frac{R}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\\
&= \prod_{i=1}^{n}\Bigg[\left[\mathbb{I}\left(z_i>0\right)\mathbb{I}\left(W_i=1\right)+\mathbb{I}\left(z_i\leq 0\right)\mathbb{I}\left(W_i=0\right)\right]\\
&\cdot\frac{R}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\Bigg]
\end{aligned} \tag{45}$$

The complete data likelihood is not of much direct use since the latent data is not observed. Where the complete data likelihood comes into use is in the design of algorithms to estimate $\theta^* = (\xi,\omega^2,\alpha,R)$

3.6.3 Joint Posterior Distributions of the augmented data

Having used the notion of Bayes' theorem in and assumption of prior independence, the full posterior distribution corresponding to the proposed model parameters R, ξ, ω^2, α and the auxiliary variable $\mathbf{z}$ is proportional to the product of the complete-data likelihood in Equation 45, and the prior specifications given in subsubsection 3.5.2 up to normalizing constant,*i.e.*,

$$
\begin{aligned}
p\left(R, \xi, \omega^2, \alpha, z|\mathrm{w}\right) \propto & \left(\text{ Complete-data } \text{ Likelihood }\right) \times \left(\text{ Prior }\right) \\
& \propto p\left(\mathrm{w}, z|R, \xi, \omega^2, \alpha\right) p\left(R, \xi, \omega^2, \alpha\right) \\
& \propto \prod_{i=1}^{n} \Bigg[\left[\mathbb{I}\left(z_i > 0\right) \mathbb{I}\left(W_i = 1\right) + \mathbb{I}\left(z_i \leq 0\right) \mathbb{I}\left(W_i = 0\right)\right] \\
& \cdot \frac{R}{\omega} \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right) \Bigg] \times p\left(R, \xi, \omega, \alpha\right) \\
& \propto \prod_{i=1}^{n} \Bigg[\left[\mathbb{I}\left(z_i > 0\right) \mathbb{I}\left(W_i = 1\right) + \mathbb{I}\left(z_i \leq 0\right) \mathbb{I}\left(W_i = 0\right)\right] \\
& \cdot \frac{R}{\omega} \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right) \Bigg] \times p\left(R\right) \times p\left(\xi\right) \times p\left(\omega^2\right) \times p_i\left(\alpha\right) \qquad (46)
\end{aligned}
$$

, where $i = 1, 2$ and

$$
\begin{aligned}
p\left(R\right) &= \tfrac{b^a}{\Gamma(a)} R^{a-1} e^{-bR} \propto R^{a-1} e^{-bR}, \\
p\left(\xi\right) &= \tfrac{1}{\sqrt{2\pi\kappa\omega^2}} e^{-\frac{1}{2\kappa\omega^2}(\xi - \xi_o)^2} \propto \left(\omega\right)^{-1} e^{-\frac{1}{2\kappa\omega^2}(\xi - \xi_o)^2}, \\
p\left(\omega^2\right) &= \tfrac{d^c}{\Gamma(c)} \left(\omega^2\right)^{-c-1} e^{-d/\omega^2} \propto \left(\omega^2\right)^{-c-1} e^{-d/\omega^2}, \\
p_1\left(\alpha\right) &= \tfrac{1}{\sqrt{2\pi\psi_o^2}} e^{-\frac{1}{2\psi_o^2}(\alpha - \alpha_o)^2} \propto e^{-\frac{1}{2\psi_o^2}(\alpha - \alpha_o)^2} \propto \phi\left(\tfrac{\alpha - \alpha_o}{\psi_o}\right), \text{ and} \\
p_2\left(\alpha\right) &= \tfrac{2}{\psi_o} \phi\left(\tfrac{\alpha - \alpha_o}{\psi_o}\right) \Phi\left(\lambda_o \tfrac{\alpha - \alpha_o}{\psi_o}\right) \propto \phi\left(\tfrac{\alpha - \alpha_o}{\psi_o}\right) \Phi\left(\lambda_o \tfrac{\alpha - \alpha_o}{\psi_o}\right)
\end{aligned}
$$

3.6.4 Full Conditional Distributions of Augmented data

Evidently, it is not easy to draw (independent) samples directly from the joint posterior described in Equation 46 above. Thus, we use the Gibbs sampler[13], a class of MCMC algorithm to make (slightly) dependent, approximate draws from this target(posterior)

[13] Gibbs sampling is applicable when the joint distribution is not known explicitly or is difficult to sample from directly, but the conditional distribution of each variable is known and is easy (or at least, easier) to sample from.

distribution and thereby effectively approximate it. To implement Gibbs sampler, however, we have to have the full conditionals[14] of each of the parameters stated in the model.In our case, driving those conditionals from the posterior in Equation 46 is a simple but tedious matter in a sense that we need only to select the terms containing the parameter of interest,then discard all other multiplicative terms as proportionality constants, and at last, rearrange what's left to determine the resulting distribution. Accordingly, what follows are the conditionals of each parameter given in the joint posterior distribution of Equation 46.

Sampling Z:

Given ξ, ω^2, α and R, the variables z_i are independent, and $z_i|\xi, \omega^2, \alpha, R$ is distributed as the $SN(\xi, \omega^2, \alpha)$ truncated at the left by 0 if $w_i = 1$ and truncated at the right by 0 if $w_i = 0$, for $i = 1, \cdots, n$

$$\begin{aligned} p(z_i|R, \xi, \omega^2, \alpha, W) &\propto \left\{ \left[\mathbb{I}\left(z_i > 0\right) \mathbb{I}\left(W_i = 1\right) + \mathbb{I}\left(z_i \leq 0\right) \mathbb{I}\left(W_i = 0\right) \right] \cdot \frac{R}{\omega} \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right) \right\} \\ &\propto \begin{cases} \frac{1}{\omega} \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right) \mathbb{I}\left(z_i > 0\right) & \text{if } \mathrm{w}_i = 1 \\ \frac{1}{\omega} \phi\left(\frac{z_i - \xi}{\omega}\right) \Phi\left(\alpha \frac{z_i - \xi}{\omega}\right) \mathbb{I}\left(z_i \leq 0\right) & \text{if } \mathrm{w}_i = 0 \end{cases} \\ &\propto \begin{cases} \mathrm{SN}\left(z_i; \xi, \omega^2, \alpha\right) \mathbb{I}\left(z_i > 0\right) & \text{if } \mathrm{w}_i = 1 \\ \mathrm{SN}\left(z_i; \xi, \omega^2, \alpha\right) \mathbb{I}\left(z_i \leq 0\right) & \text{if } \mathrm{w}_i = 0 \end{cases} \end{aligned} \tag{47}$$

Therefore,

$$z_i|R, \xi, \omega^2, \alpha, W \backsim \begin{cases} \mathrm{SN}\left(z_i; \xi, \omega^2, \alpha\right) \mathbb{I}\left(z_i > 0\right) & \text{if } \mathrm{w}_i = 1 \\ \mathrm{SN}\left(z_i; \xi, \omega^2, \alpha\right) \mathbb{I}\left(z_i \leq 0\right) & \text{if } \mathrm{w}_i = 0 \end{cases} \tag{48}$$

Since z_i distributed truncated Skew normal and the simulation need to be performed on its density, normalization the conditional in Equation 48 should, therefore, be made so that it becomes a density on its support and will also integrate to 1. Therefore, the conditional distribution in Equation 48 can be rewritten

$$z_i|R_j, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b \backsim \begin{cases} C_1 \mathrm{SN}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right) & \text{if } z_i > 0 \ \& \ \mathrm{w}_i = 1 \\ \\ C_2 \mathrm{SN}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right) & \text{if } z_i \leq 0 \ \& \ \mathrm{w}_i = 0 \end{cases} \tag{49}$$

in which case, C_1 and C_2 are normalization constants provided as

$$C_1 = \frac{1}{\int_0^{+\infty} \mathrm{SN}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right) dz_i} = \frac{1}{cdf_{\mathrm{SN}}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right)}$$

and

$$C_2 = \frac{1}{\int_{-\infty}^0 \mathrm{SN}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right) dz_i} = \frac{1}{1 - \int_0^{\infty} \mathrm{SN}\left(z_i; \xi_j, \omega_j^2, \alpha_j\right) dz_i} = \frac{1}{1 - cdf_{\mathrm{SN}}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right)}$$

[14] Full Conditional Distributions/Full Conditional Posterior Distributions:-this is the conditional distribution of one parameter conditional on the known information and all the other parameters, $p(\theta_j|\theta_{-j}, y)$ for any parameter θ in the model

We can sample from this univariate truncated skew-normal distribution (40) in a similar fashion as the Accept-Reject procedure, along with Gibbs sampling algorithm described in (Griggio, 2013/2014; Canale and Scarpa, 2013) to sample from truncated distributions.

Sampling R:

Analogously, the relevant terms in the conditional posterior for fertility rate parameter, R are

$$
\begin{aligned}
p(R|z_i, \xi, \omega^2, \alpha, W) &\propto \left[\prod_{i=1}^{n} R\right] \cdot R^{a-1} e^{-bR} \\
&\propto R^n \cdot R^{a-1} e^{-bR} \\
&\propto R^{n+a-1} e^{-bR}
\end{aligned} \tag{50}
$$

Therefore,

$$
R|z_i, \xi, \omega^2, \alpha, W \backsim \text{Gamma } (n+a, b) \tag{51}
$$

In a similar manner, sampling (ξ, ω^2) and α can be done using the following conditional distributions

$$
\begin{aligned}
p\left(\xi, \omega^2|\alpha, Z, W\right) &\propto \left\{\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_i - \xi}{\omega}\right)\Phi\left(\alpha\frac{z_i - \xi}{\omega}\right)\right\} \times p\left(\xi\right) \times p\left(\omega^2\right) \\
&\propto \left\{\prod_{i=1}^{n} \text{SN}\left(z_i; \xi, \omega^2, \alpha\right)\right\} \times \text{N}\left(\xi; \xi_o, \kappa\omega^2\right) \times \text{Inv.Gamma}\left(\omega^2; c, d\right)
\end{aligned} \tag{52}
$$

and

$$
\begin{aligned}
p\left(\alpha|\xi, \omega^2, Z, W\right) &\propto \begin{cases} \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\right] \times p_1\left(\alpha\right), & \text{where } p_1\left(\alpha\right) \text{ is a normal prior} \\ \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\right] \times p_2\left(\alpha\right), & \text{where } p_2\left(\alpha\right) \text{ is a SN prior} \end{cases} \\
&\propto \begin{cases} \left\{\prod_{i=1}^{n} \text{SN}\left(z_i; \xi, \omega^2, \alpha\right)\right\} \times \text{SN}\left(\alpha; \alpha_o, \psi_o^2\right) \\ \left\{\prod_{i=1}^{n} \text{SN}\left(z_i; \xi, \omega^2, \alpha\right)\right\} \times \text{SN}\left(\alpha; \alpha_o, \psi_o^2, \lambda_o\right) \end{cases}
\end{aligned} \tag{53}
$$

Sampling from (ξ, ω^2):

We can observe from Equation 41 that

$$
z_i \backsim \text{SN}(\xi, \omega^2, \alpha), \quad i = 1, 2, \cdots, n \tag{54}
$$

Note that z_i are latent variable introduced previously. In addition, using the stochastical representation for the skew normal distribution (Henze, 1986) we can write

$$
z_i = \xi + \omega\delta\eta_i + \omega\sqrt{1-\delta^2}\epsilon_i, , \quad i = 1, 2, \cdots, n \tag{55}
$$

where, $\epsilon_i \backsim N(0,1)$, the standard normal distribution and $\eta_i \backsim TN_{[0,\infty)}(0,1)$, the half normal distribution.It follows that the conditional distribution $z_i|\eta_i$ is a normal distribution with mean $\xi+\omega\delta\eta_i$ and variance $\omega\sqrt{1-\delta^2}$ where $\delta=\alpha/\sqrt{1+\alpha^2}$.By considering this result, a hierarchical formulation of the model is given as follow:

$$\bigstar Z|\eta_i,\omega^{-2},\xi,\delta \backsim N\left(\xi+\omega\delta\eta_i,\omega\sqrt{1-\delta^2}\right)$$
$$=> p(Z|\eta_i,\omega^{-2},\xi,\delta) \propto \prod_{i=1}^{n}\frac{1}{\omega\sqrt{1-\delta^2}}e^{-\frac{1}{2\omega\sqrt{1-\delta^2}}(z_i-\xi-\omega\delta\eta_i)^2} \tag{56}$$
$$\bigstar \eta_i \backsim TN_{[0,\infty)}(0,1)$$
$$=> p(\eta_i) \propto e^{-\frac{1}{2}\eta_i^2}\mathbb{I}_{\eta_i>0} \tag{57}$$

Putting together Equation 56 and Equation 57 with the prior specification given thus far, we obtain the complete posterior conditional distributions given by

$$p(\eta_i,\omega^{-2},\xi,\delta|Z) \propto \left\{\prod_{i=1}^{n}\frac{1}{\omega\sqrt{1-\delta^2}}e^{-\frac{1}{2\omega\sqrt{1-\delta^2}}(z_i-\xi-\omega\delta\eta_i)^2}\times e^{-\frac{1}{2}\eta_i^2}\mathbb{I}_{\eta_i>0}\right.$$
$$\left.\times\frac{1}{\sqrt{\kappa\omega^2}}e^{-\frac{1}{2\kappa\omega^2}(\xi-\xi_o)^2}\times(\omega^{-2})^{c-1}e^{-d/\omega^{-2}}\right\} \tag{58}$$

Note that all of the full conditional distributions for Gibbs sampling are straightforward to derive and are given as follows:

$$\blacktriangle\quad \eta_i|\omega^{-2},\xi,\delta,Z \backsim \mathrm{TN}\left(\frac{\delta(z_i-\xi)}{\omega^2},1-\delta^2\right) \tag{59}$$

$$\blacktriangle\quad \xi|\eta_i,\omega^{-2},\delta,Z \backsim \mathrm{N}\left(\widehat{\mu},\widehat{\kappa}\omega^2\right) \tag{60}$$

where,

$$\widehat{\mu}=\frac{\kappa\sum_{i=1}^{n}(z_i-\delta\eta_i)+\xi_o(1-\delta)^2}{n\kappa+(1-\delta^2)} \quad\text{and}\quad \widehat{\kappa}=\frac{\kappa(1-\delta^2)}{n\kappa+(1-\delta^2)}$$

$$\blacktriangle\quad \omega^{-2}|\eta_i,\xi,\delta,Z \backsim \mathrm{InvGamma}\left(c+\frac{1}{2}(n+1),d+\widehat{d}\right) \tag{61}$$

$$\widehat{d}=\frac{1}{2\omega^2(1-\delta^2)}\sum_{i=1}^{n}(z_i-\xi-\omega\delta\eta_i)^2+\frac{1}{2\kappa\omega^2}(\xi-\xi_o)^2$$
$$=\frac{1}{2(1-\delta^2)}\left\{\delta^2\sum_{i=1}^{n}\eta_i^2-2\delta\sum_{i=1}^{n}\eta_i(z_i-\xi)+\sum_{i=1}^{n}\eta_i(z_i-\xi)^2+\frac{(1-\delta^2)}{\kappa}(\xi-\xi_o)^2\right\}$$

Sampling from α:

We have instrumented pretty much the same approach as employed by (Canale and Scarpa, 2013) so as to update and draw samples for skewed parameter, α, from Equation 41. To this end,without loss of generality, we assume the distribution of latent variables $z_1,\cdots,z_n$ in Equation 41 as standard skew normal distribution .Putting it differently, we considered the fact that ξ and ω in Equation 41 are known, and $\xi=0$ and $\omega=1$.Following this, the expression in Equation 41, then, reduces to

$$p(\alpha|\boldsymbol{z}) \propto \begin{cases} \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\right] \times p_1(\alpha), & \text{where } p_1(\alpha) \text{ is a normal prior} \\ \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_i-\xi}{\omega}\right)\Phi\left(\alpha\frac{z_i-\xi}{\omega}\right)\right] \times p_2(\alpha), & \text{where } p_2(\alpha) \text{ is a SN prior} \end{cases}$$

$$\propto \prod_{i=1}^{n} \phi(z_i)\,\Phi(\alpha z_i) \times p_i(\alpha), \quad i = 1, 2 \tag{62}$$

where,

$$p_1(\alpha) = \frac{1}{\psi_o}\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right) \text{ and } p_2(\alpha) = \frac{2}{\psi_o}\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)\Phi\left(\lambda_o\frac{\alpha-\alpha_o}{\psi_o}\right)$$

, which are two noninformative prior distributions of α(Canale and Scarpa, 2013). Having made simplification on the derivations, the full conditional posterior distribution, $\alpha|\xi, \omega^2, \boldsymbol{z}, \boldsymbol{w}$, under the two prior cases is given as follows:

When using Normal Prior, $p_1(\alpha) = \frac{1}{\psi_o}\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)$

$$\begin{aligned} &\alpha|z \backsim \mathrm{SUN}_{1,n}\left(\alpha_o, \Delta_1\alpha_o/\psi_o, \psi_o, \Omega^*\right), \quad \text{or} \\ &\alpha|z \backsim \mathrm{SUN}_{1,n}\left(\alpha_o, \Delta_1\alpha_o/\psi_o, \psi_o, 1, \Delta_1, \Gamma_1\right) \end{aligned} \tag{63}$$

and

When using Skew Normal Prior, $p_2(\alpha) = \frac{2}{\psi_o}\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)\Phi\left(\lambda_o\frac{\alpha-\alpha_o}{\psi_o}\right)$

$$\begin{aligned} &\alpha|z \backsim \mathrm{SUN}_{1,n+1}\left(\alpha_o, \gamma_2, \psi_o, 1, \Delta_2, \Gamma_2\right) \\ &\alpha \backsim \pi(\alpha|v^*) \end{aligned} \tag{64}$$

where $y^* = (y_i - \xi)/\omega$ for $i = 1, \cdots, n$ and $\pi(\alpha|y), j = 1, 2$. The details on the derivation of those conditional distributions are given on the appendix.

3.7 Bayesian Computation: *Markov Chain Monte Carlo Implementation*

We use a combination of the Accept-Reject and the Gibbs sampler algorithm to draw samples from each full posterior conditional distribution, and thereby, estimate the parameters, $\theta_i^* = (\xi_i, \omega_i^2, \alpha_i)$ of the proposed ASFR model for each region. We utilize the the Accept-Reject Sampling (aka: rejection sampling) method here to obtain posterior samples for the shape parameter, α, from the truncated Skew-Normal distribution. Here, we used R-optimization command, `nlminb` to determine upper bound (global maximum). Unlike to α_i, Gibbs Sampling algorithm was rendered to draw posterior samples for all other model parameters of each region from their respective full conditionals. Algorithm 4 below recapitulates all the conditions of the model parameters used in evaluation of posterior samples.

We set, $z_{ij}^{(t+1)}$ equal to a random draw using accept-reject algorithm from a $\texttt{TSN}(\xi_j^{(t)}, \omega_j^{2(t)}, \alpha_j^{(t)})$ distribution based on $\texttt{U}(j-1/2, j+1/2)$. Then, for $(t+1)^{st}$ iteration of MCMC, for $j = 1, \cdots, 11$, we set $R_j^{(t+1)}$ equal to a random draw from $\texttt{Gamma}(n + a^{(t)}, b^{(t)})$ distribution given in *(ii)* above, with the conditioning arguments , $\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b$, respectively. Following this, we set we set $\eta_{ij}^{(t+1)}$ equal to a random draw from TN $\left(\frac{\delta_j^{(t)}(z_{ij}^{(t)} - \xi_j^{(t)})}{\omega_j^{2((t))}}, 1 - \delta_j^{2(t)}\right)$, where, $\delta_j^{(t)} = \alpha_j^{(t)} / \sqrt{1 - \alpha_j^{(t)}}$ distribution given in *(ii)* above, with the conditioning arguments , $\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b$, respectively. Similarly, to sample $\xi_j^{(t+1)}$, we set we set $\xi_j^{(t+1)}$ equal to a random draw from N $\left(\widehat{\mu}, \widehat{\kappa}\omega_j^{2(t)}\right)$ distribution, where

$$\widehat{\mu} = \frac{\kappa \sum_{i=1}^{n_j} (z_{ij}^{(t+1)} - \delta_j^{(t)} \eta_{ij}^{(t+1)}) + \xi_o (1 - \delta_{ij})^2}{n\kappa + (1 - \delta_j^{2(t)})} \quad \text{and} \quad \widehat{\kappa} = \frac{\kappa(1 - \delta_j^{2(t)})}{n\kappa + (1 - \delta_j^{2(t)})}$$

We also set $\omega_j^{-2(t+1)}$ equal to a random sample from InvGamma $\left(c + \frac{1}{2}(n+1), d + \widehat{d}\right)$ distribution given in *(v)* above, where,

$$\widehat{d} = \frac{1}{2(1 - \delta_j^{2(t)})} \left\{ \left[\delta_j^{2(t)} \sum_{i=1}^{n_j} \eta_{ij}^{2(t+1)} - 2\delta_j^{(t)} \sum_{i=1}^{n_j} \eta_j^{(t+1)} (z_{ij}^{(t+1)} - \xi_j^{(t+1)}) \right] \right.$$
$$\left. + \sum_{i=1}^{n_j} \eta_{ij}^{(t+1)} (z_{ij}^{(t+1)} - \xi_j^{(t+1)})^2 + \frac{(1 - \delta_j^2(t))}{\kappa} (\xi_j^{(t+1)} - \xi_o)^2 \right\}$$

Summary of full conditions for Posterior simulation

$$(\boldsymbol{i}).\quad z_{ij}|R_j,\xi_j,\omega_j^2,\alpha_j,\boldsymbol{w},a,b \backsim \begin{cases} C_1\text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right) & \text{if } z_{ij}>0 \ \&\ \text{w}_{ij}=1 \\ C_2\text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right) & \text{if } z_{ij}\leq 0 \ \&\ \text{w}_{ij}=0 \end{cases}$$

where, C_1 and C_2 are normalization constants given by

$$C_1 = \frac{1}{cdf_{\text{SN}}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)} \quad \text{and} \quad C_2 = \frac{1}{1-cdf_{\text{SN}}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)}$$

$$(\boldsymbol{ii}).\quad R_j|\boldsymbol{z},\xi_j,\omega_j^2,\alpha_j,\boldsymbol{w},a,b \backsim \text{Gamma}\ (n+a,b) \quad \text{for} \quad j=1,\cdots,J=11$$

$$(\boldsymbol{iii}).\quad \eta_{ij}|\omega_{ij}^{-2},\xi_j,\delta_{ij},\boldsymbol{z} \backsim \text{TN}\left(\frac{\delta(z_{ij}-\xi_j)}{\omega_j^2},1-\delta_{ij}^2\right),\ i=1,2,\cdots,n_j;j=1,2,\cdots,J$$

$$(\boldsymbol{iv}).\quad \xi_j|\eta_{ij},\omega_j^{-2},\delta_{ij},\boldsymbol{z} \backsim \text{N}\left(\widehat{\mu},\widehat{\kappa}\omega_j^2\right),\ j=1,2,\cdots,J$$

where,

$$\widehat{\mu} = \frac{\kappa\sum_{i=1}^{n_j}(z_{ij}-\delta_{ij}\eta_{ij})+\xi_o(1-\delta_{ij})^2}{n\kappa+(1-\delta_{ij}^2)} \quad \text{and} \quad \widehat{\kappa} = \frac{\kappa(1-\delta_{ij}^2)}{n\kappa+(1-\delta_{ij}^2)}$$

$$(\boldsymbol{v}).\quad \omega_j^{-2}|\eta_{ij},\xi_j,\delta_{ij},\boldsymbol{z} \backsim \text{InvGamma}\left(c+\frac{1}{2}(n+1),d+\widehat{d}\right),j=1,2,\cdots,J$$

$$\widehat{d} = \frac{1}{2(1-\delta_{ij}^2)}\left\{\delta_{ij}^2\sum_{i=1}^{n_j}\eta_{ij}^2-2\delta_{ij}\sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j)+\sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j)^2+\frac{(1-\delta_{ij}^2)}{\kappa}(\xi_j-\xi_o)^2\right\}$$

$$(\boldsymbol{vi}).\quad \alpha_j|z \backsim \begin{cases} \text{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,\Omega^*\right) \\ \\ \text{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,1,\Delta_1,\Gamma_1\right) \end{cases} \tag{65}$$

Following this, we set we set $\eta_{ij}^{(t+1)}$ equal to a random draw from TN $\left(\frac{\delta_j^{(t)}(z_{ij}^{(t)}-\xi_j^{(t)})}{\omega_j^{2((t))}},1-\delta_j^{2(t)}\right)$, where, $\delta_j^{(t)}=\alpha_j^{(t)}/\sqrt{1-\alpha_j^{(t)}}$ distribution given in *(ii)* above, with the conditioning arguments , $\boldsymbol{z},\xi_j,\omega_j^2,\alpha_j,\boldsymbol{w},a,b$, respectively. Similarly, to sample $\xi_j^{(t+1)}$, we set we set $\xi_j^{(t+1)}$ equal to a random draw from N $\left(\widehat{\mu},\widehat{\kappa}\omega_j^{2(t)}\right)$ distribution, where

$$\widehat{\mu} = \frac{\kappa\sum_{i=1}^{n_j}(z_{ij}^{(t+1)}-\delta_j^{(t)}\eta_{ij}^{(t+1)})+\xi_o(1-\delta_{ij})^2}{n\kappa+(1-\delta_j^{2(t)})} \quad \text{and} \quad \widehat{\kappa} = \frac{\kappa(1-\delta_j^{2(t)})}{n\kappa+(1-\delta_j^{2(t)})}$$

We also set $\omega_j^{-2(t+1)}$ equal to a random sample from InvGamma $\left(c+\frac{1}{2}(n+1), d+\widehat{d}\right)$ distribution given in *(v)* above, where,

$$\widehat{d}=\frac{1}{2(1-\delta_j^{2(t)})}\left\{\left[\delta_j^{2(t)}\sum_{i=1}^{n_j}\eta_{ij}^{2(t+1)}-2\delta_j^{(t)}\sum_{i=1}^{n_j}\eta_j^{(t+1)}(z_{ij}^{(t+1)}-\xi_j^{(t+1)})\right]\right.$$
$$\left.+\sum_{i=1}^{n_j}\eta_{ij}^{(t+1)}(z_{ij}^{(t+1)}-\xi_j^{(t+1)})^2+\frac{(1-\delta_j^2(t))}{\kappa}(\xi_j^{(t+1)}-\xi_o)^2\right\}$$

Comparison between maximum method and Bayesian method is organized and given the analysis section, a practical data set is used to compare the two methods.

3.8 Empirical Results:

3.8.1 Sensitivity of priors:

Sensitivity refers to how sensitive a model's performance are to minor changes in the model parameters. Prior sensitivity analysis plays an important role in applied Bayesian analyses as it aids to assess the robustness of the model using different prior distributions. Thus, with regards our model, it is important to examine the sensitivity of our results to the prior information for ξ_o, k, α_o, ψ_o^2, λ_o, a and b. To this effect, we first use noninformative priors of:

$$\begin{aligned}&Gamma\,(m_o,n_o)\,,\ \ where\ \ m_o=0.01,\quad m_o=0.01\\&Gamma\,(r_o,s_o)\,,\ \ where\ \ r_o=0.01,\quad s_o=0.01\\&\mathrm{N}\left(\xi_o,k\omega_j^2\right),\ \ where\ \ \xi_o=20,\quad k=0.6\\&\mathrm{SN}\left(\alpha_o,\psi_o^2,\lambda_o\right),\ \ where\ \ \alpha_o=1,\quad \psi_o^2=4,\quad \lambda_o=2\end{aligned}$$

However, to examine whether or not our estimates are robust to reasonable changes in the prior, we repeated the analysis for alternative choices of:

$$\begin{aligned}&Gamma\,(m_o,n_o)\,,\ \ where\ \ m_o=0.001,\quad m_o=0.001\\&Gamma\,(r_o,s_o)\,,\ \ where\ \ r_o=0.001,\quad s_o=0.001\\&\mathrm{N}\left(\xi_o,k\omega_j^2\right),\ \ where\ \ \xi_o=3,\quad k=0.9\\&\mathrm{SN}\left(\alpha_o,\psi_o^2,\lambda_o\right),\ \ where\ \ \alpha_o=2,\quad \psi_o^2=5,\quad \lambda_o=3\end{aligned}$$

The resulting posterior findings for our model parameters in each region, $\theta_i^*=(R_i,\xi_i,\omega_i^2,\alpha_i)$ were assessed. The overall values were similar to what we obtained previous. Finally, we repeated the analysis for a flatter prior

$$\begin{aligned}&Gamma\,(m_o,n_o)\,,\ \ where\ \ m_o=0.1,\quad m_o=0.1\\&Gamma\,(r_o,s_o)\,,\ \ where\ \ r_o=0.1,\quad s_o=0.1\\&\mathrm{N}\left(\xi_o,k\omega_j^2\right),\ \ where\ \ \xi_o=4,\quad k=2\\&\mathrm{SN}\left(\alpha_o,\psi_o^2,\lambda_o\right),\ \ where\ \ \alpha_o=0.5,\quad \psi_o^2=5,\quad \lambda_o=-2\end{aligned}$$

and, our the posterior findings were determined. The result obtained were almost identical to what determined before, indicating the results are robust to changes in prior information.

3.8.2 Model Adequacy

The posterior predictive distribution provides diagnostics for assessing models as well as statistics for making inferences about the discrepancy between data and model. The posterior predictive distribution assigns probabilities to hypothetical or future values of y, written y^{rep}, integrating over uncertainty about the posterior distribution of the parameters (159-177, Gelman et al., 2014):

$$\begin{aligned} p\left(y^{rep}|y\right) &= P\left[T(y^{rep}) \geq T((y^{rep})|y^{rep}\right] \\ &= \int \underbrace{p\left(y^{rep}|\theta^{*}\right)}_{\substack{\text{Sampling} \\ \text{Distribution}}} \quad \underbrace{p\left(\theta^{*}|y\right)}_{\substack{\text{Posterior} \\ \text{Distribution}}} \quad d\theta \end{aligned} \tag{66}$$

The integral defining the posterior predictive distribution has two parts. The first part gives the probability density of y^{rep} given particular values of θ^{*}. The form of this density is given by the sampling distribution for y. The second part of the integral is the posterior distribution for the model parameters, θ^{*}. The posterior predictive distribution incorporates two kinds of uncertainty:sampling uncertainty about y given θ^{*}, and parametric uncertainty about θ^{*}. The posterior predictive distribution can be compared to the observed data to assess model fit. If a model fits the data well, the observed data are relatively likely under the posterior predictive distribution. On the other hand, large discrepancies between the observed data and the posterior predictive distribution indicate that the model fits poorly. The posterior predictive distribution is straightforward to simulate for Bayesian models estimated with Markov Chain Monte Carlo (MCMC) methods. Given random draws,θ^{*t} from the posterior distribution and the sampling distribution, $p\left(y^{rep}|\theta^{*}\right)$ rep can be generated by a random draw from$p\left(y^{rep}|\theta^{*t}\right)$.

The discrepancy between the model and the data can be assessed with a test statistic, $T\left(y\right)$, that summarize some substantively important feature of y. Model fit can be judged by comparing the observed $T\left(y\right)$ to the distribution of $T\left(y^{rep}\right)$. A Bayesian *p-value* is defined by,

$$p = Pr\left[T(y^{rep}) \geq T(y^{rep})|y^{rep}\right]. \tag{67}$$

The *p-value* describes, conditional on the model, the probability of observing data at least as extreme as that actually observed. An extreme value for p indicates the data are unlikely under the model. Note that the *p-value* does not describe the probability that a particular model is correct, nor nor does the *p-value* provide evidence against a null in favor of an alternative. Instead, small *p-value* indicate the implausibility of the data under the model and the utility of examining other models(Albert and Chib, 1993).

3.8.3 Convergence diagnosis

It is important to establish whether a sequence of Markov chain Monte Carlo iterations has converged, that is, reached its stationary distribution. A number of different *iteration* techniques have been developed to determine whether Markov chain Monte Carlo algorithms have reached convergence. Cowles and Carlin (1996) provided a useful summary of these techniques. It is important to be able to determine when a chain has not converged, but many researchers believe that the available convergence appraisal techniques are essentially unstable because the stationary distributions will never be identified. Cowles and Carlin (1996) suggested techniques for avoiding this problem, proposing that diagnostics should be used with vigilance and that multiple methods should be used rather than just a single method.

In this thesis, three of the most popular diagnostic tests will be used to determine whether the chains of parameters have achieved convergence for the models discussed. Any type of MCMC sampler can be used with these methods which include Geweke (1992), Heidelberger and Welch (1983) and Raftery et al. (1992). The chains of parameters are considered to have reached convergence if the z-score of the Geweke diagnostic test lies between -2 and 2, which is considered to have a 5% significance level (Geweke, 1992). The chains of parameters are also considered to have achieved convergence if the p-value of the Heidelberger and Welch (1983) diagnostic test is lower than 95%, indicating the acceptance of the null hypothesis. Under the Raftery et al. (1992) diagnostic test the chains of parameters are considered to have converged if the dependence factor (I) is lower than 5, indicating that the sample is less correlated Raftery et al. (1992). Informaly, we can also examine the convergence with the help of time series *trace plot*, as displaced in the next section below.

3.8.4 Posterior Findings

Gibbs sampling algorithm in `R` was implemented to draw samples from the full conditionals given above and there by, estimate the parameters, ξ, ω^2, α and R, in the proposed ASFR model for each of the eleven regions in the country. A total of 6,000 iterations were conducted, but only the final 1,000 iterations were of use in determining our posterior findings: posterior point and interval estimates, assessing convergence, etc to mention some.

Table 3.3 presents the Bayesian posterior point estimates, *i.e.*, posterior moments (*pos-*

terior mode, posterior mean, posterior median, standard deviations) and its 95% credibility interval for the model parameters in Somali region: ξ_{Somali}, ω^2_{Somali}, α_{Somali} and R_{Somali}. Beside to these values, maximum likelihood estimates are also determined to make comparison with their Bayesian counterparts, and to trace which approach, Bayesian or maximum likelihood, yields better result in terms of precision. These result are given on the second and third column of the table respectively. From the result in the table, one can extract or draw a lot of worthwhile points. For instance,

=> the posterior mean of location parameter, which is important to recognize the condition of fertility in the region, was 19.92 with 95% credible interval of $(18.87, 20.98)$, which is a strong evidence that the majority of the childbearing mothers in the area gave birth at early age, that is, earlier than 21 year old, mean age. The result from maximum likelihood also strongly shares this notion.

=> So as to have an idea on how the pattern of fertility rate, specially age specific rate, looks like in the region, the shape parameter has also been examined. The underlined result in the table indicates that the shape parameter of this region, α_{Somali}, had a posterior mean of value 3.59 with 95% credible interval of $(3.26, 3.92)$, indicating a strong evidence of positive skewness. The MLE of this parameter was also close the Bayesian value , which is, about 3.00 with 95% credible interval of $(2.85, 3.56)$. These results strongly support the idea that most of the people in the region delivered their first birth at early age. This results also signals an abstract message to the government or any other concerned body so as to carry out much work on the factors that exacerbates and favors this early fertility, among others.

=> Scientific evidence on the rate of fertility is also another crucial element for policy makers and demographers. To this end, the mean fertility rate has been under scrutiny using the two approaches as seen in the table. Both the posterior and the maximum likelihood estimates of the region also confirms the fertility rate, R_{Somali}, in the area was high, with an average children of about 7 per single mother. This result also indicates the posterior mean was much closer to the true value fixed fertility value ($R = 7.1$) than the value of obtained from the maximum likelihood method. Thus, it can be concluded that the Gibbs Sampler algorithm (Bayesian method) gave more consistent fertility estimate than the maximum likelihood approach ***since its true value was also inside the 95% credible interval***. The result in the table also reveals that the credible interval of this fertility rate parameter $(6.49, 7.48)$, was narrower than its maximum likelihood compliment, that is, $(6.19, 7.73)$ suggesting that the use of the Gibbs sampler gave more precise result with lower uncertainty than its maximum likelihood analogue.

Table 3.3: **Bayesian and ML estimates for fertility data set of Somali Region**

Parameters	Bayesian Estimates					ML Estimate		
	Mean	Median	SD	95%[Conf	Int]	MLE	95%[Conf	Int]
ξ	19.923	19.889	0.6234	18.86945	20.97655	19.460	17.311	21.609
ω^2	8.4102	8.3813	0.3168	7.874808	8.945592	8.207	7.301	9.113
α	3.5883	3.5133	0.1941	3.260271	3.916329	3.002	2.853	3.561
R	6.9853	6.7915	0.2941	6.488271	7.482329	6.962	6.193	7.731

As outlined in the previous section, there are two methods to check convergence. One is examining *trace plots of the sample values* and the other is *iteration method*. Figure 3.5 through Figure 3.8 show trace plots for Somali region fertility model parameters: ξ_{Somali}, ω^2_{Somali}, α_{Somali} and R_{Somali} respectively. We can be reasonably confident that convergence has achieved since all the chains appear to be well mixed, suggesting the stability of the model parameters. This is an *informal approach* to convergence diagnosis. A quantitative way of checking convergence is based on an analysis of variance. The results of the MCMC convergence diagnostics conducted using `coda`(`coda package`) (Cowles and Carlin, 1996) in `R` are presented, followed by the results for the estimation of parameters.

Table 3.4: **MCMC convergence diagnostic test for Somali Region fertility model parameters using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.539	*passed*	0.00241	-0.9414	1.36
ω^2	*passed*	0.632	*passed*	0.055	-0.459	3.73
α	*passed*	0.962	*passed*	0.00171	0.1039	2.11
R	*passed*	0.633	*passed*	1.25	1.076	1.31

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

Only the convergence of the estimated ξ_{Somali} was examined. The Geweke, Heidelberger & Welch (H-W), and Raftery & Lewis (R-L) tests were the convergence diagnostic tests used. The results are given in Table 3.4. The z-score for ξ_{Somali} obtained using the Geweke test was -0.9414. As this value lay between -2 and 2 it could be concluded that ξ_{Somali} converged at a 5% significance level. The stationary test for the H-W diagnostic was passed. The p-value obtained was 0.525 which indicated that the null hypothesis for ξ_{Somali} was

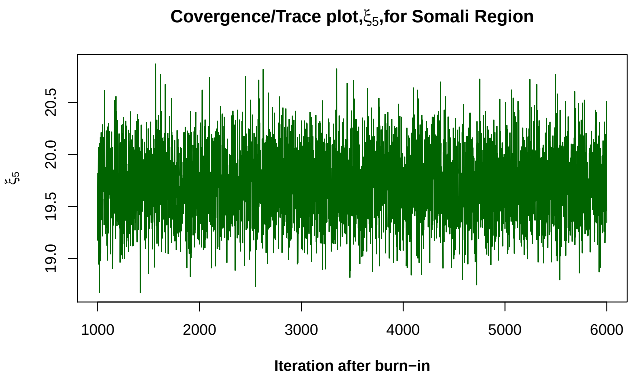


Figure 3.5: MCMC trace plots after burn-in for the fertility parameter of Somali region, ξ_{Somali}

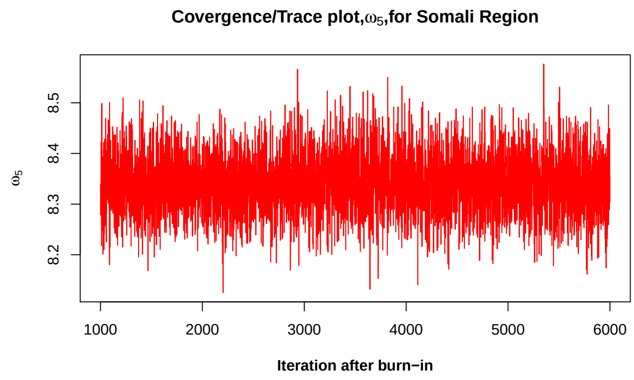


Figure 3.6: MCMC trace plots after burn-in for the fertility parameter of Somali region, ω_{Somali}

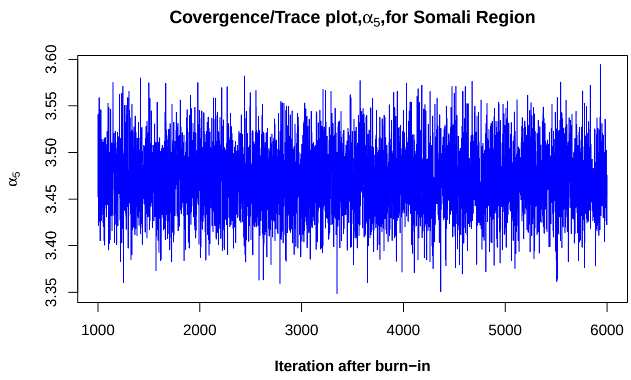


Figure 3.7: MCMC trace plots after burn-in for the fertility parameter of Somali region, α_{Somali}

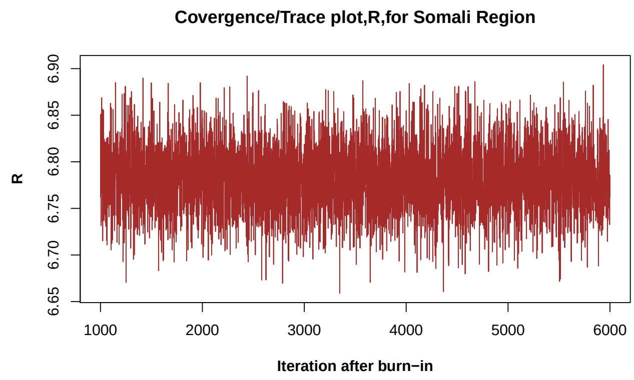


Figure 3.8: MCMC trace plots after burn-in for the fertility parameter of Somali region, R_{Somali}

not rejected. The half-width test was also passed. Furthermore, the dependence factor (I) obtained was 1.36. As the value of I obtained was lower than 5, the sample was less correlated. All of these diagnostic tests indicated that the chains had converged.

In the same way, we can interpret the convergence of the other parameters: ω^2_{Somali}, α_{Somali} and R_{Somali}, whose value is given in the stated table. The overall result in the Table 3.4 indicates us all the parameters are believed to have converged.

Unlike to the rest of Ethiopian region, Addis Ababa is one of the city administrations (the eleven regions) which has extraordinarily low fertility rate, with fertility rate of 1.9 number of children per woman of age 15-49. Evidence on this region has also organized on the table.

=> Despite the condition of fertility in Somali region, Addis Ababa had higher average fertility age, as calulated using the formula. To put it differently, the posterior mean of the location parameter ξ was 33.70 with 95% credible interval of $(33.04, 34.36)$. This figure results a mean age of 32.5 years old, which is a strong evidence that the mean age of mothers in the city to give rise their first birth was 31.6.

=> the standard deviation of ξ_{AA} obtained by Bayesian procedure was also 0.39. As this value was smaller than standard deviation of ξ_{Somali}, which is 0.62, then the spread of the data in Addis Ababa was significantly smaller.

Table 3.5: **Bayesian and ML estimates for fertility data set of Addis Ababa City Administration**

Parameters	Bayesian Estimates					ML Estimate		
	Mean	Median	SD	95%[Conf	Int]	MLE	95%[Conf	Int]
ξ	33.697	33.823	0.3914	33.03553	34.35847	32.654	29.049	36.262
ω^2	8.328	8.536	0.2678	7.846228	8.809772	8.654	7.698	9.610
α	-3.3028	-3.2102	0.1581	-3.570019	-3.035641	-3.117	-3.461	-2.773
R	1.8947	1.9176	.0432	1.821692	1.967708	1.985	1.766	2.204

Unlike to Somali region, the shape parameter of Addis Ababa had a posterior mean of -3.30 with 95% credible interval of $(-3.57, -3.03)$, and its maximum likelihood estimation was also of -3.12, with 95% confidence interval of $(-3.46, -2.77)$. Those figures suggest a strong evidence for skewness, particularly negative skewness. Hence, it can be concluded that there was a strong evidence suggesting most of the people in the capital gave rise birth at latter age of early 30 and on. In spite of their meager variation in values, the credible interval of this skewness parameter was shorter than its confidence interval. Once again, this Bayesian estimate was more precise and had lower uncertainty vis-á-vis its value obtained by maximum likelihood method.

The posterior mean and the maximum likelihood estimate of this parametric value also confirms the fertility rate in the area was very low and the number of birth per mother was also high. The result in the table also reveals that its posterior mean was much closer to the true value fixed fertility value ($R = 1.9$) than the value of obtained from the maximum likelihood method. Thus, it can be concluded that the Gibbs Sampler algorithm (Bayesian method) gave more consistent fertility estimate than the maximum likelihood approach ***since its true value was also inside the 95% credible interval.*** Moreover, the result also indicate that the credible interval of this fertility rate parameter $(1.82, 1.96)$, was narrower than its maximum likelihood compliment, that is, $(1.77, 2.20)$ suggesting that the use of the Gibbs sampler gave more precise result with lower uncertainty than its maximum likelihood analougy. The standard deviations for this parameter obtained using this formula were also smaller than results obtained by other approaches, indicating that the data was less spread out. The trace of the parameters for Addis Ababa are shown in

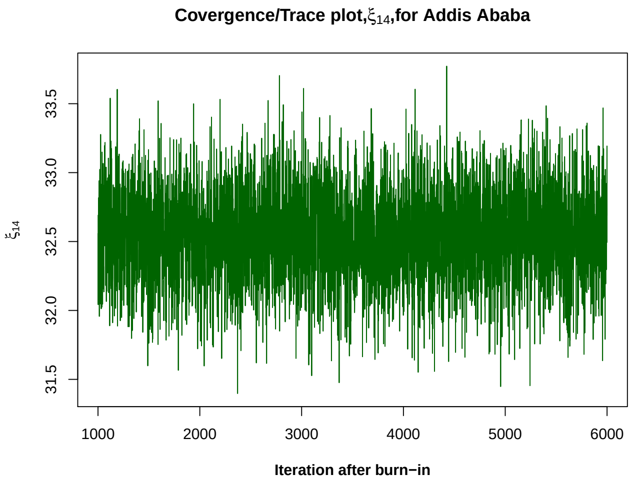


Figure 3.9: MCMC trace plots after burn-in for the fertility parameter of Addis Ababa, ξ_{AA}

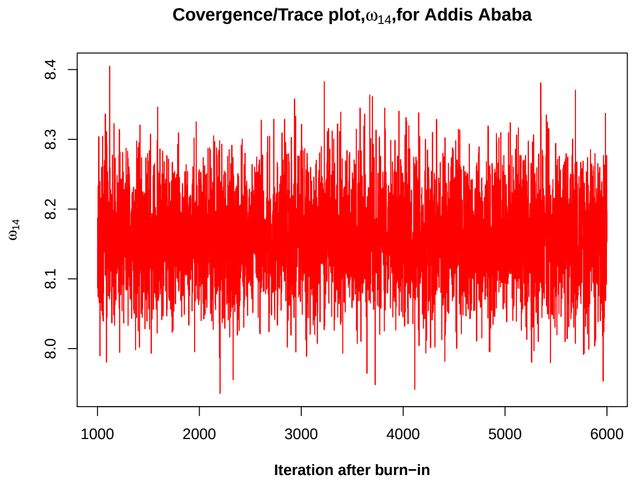


Figure 3.10: MCMC trace plots after burn-in for the fertility parameter of Addis Ababa,ω_{AA}

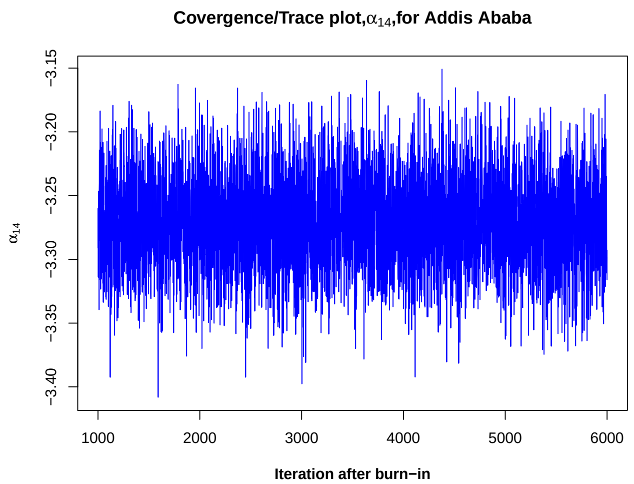


Figure 3.11: MCMC trace plots after burn-in for the fertility parameter of Addis Ababa,α_{AA}

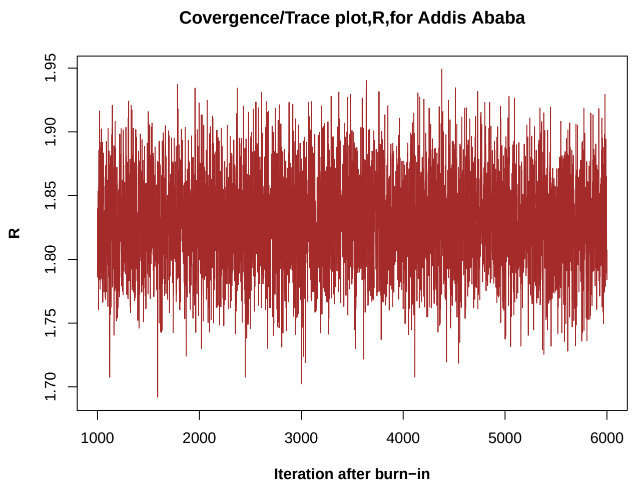


Figure 3.12: MCMC trace plots after burn-in for the fertility parameter of Addis Ababa, R_{AA}

the plot of Figure 3.9 to Figure 3.12, and as the same time results from those various Convergence diagnostic tests mentioned earlier are displayed in Table 3.6, suggesting almost all estimators, ξ_{AA}, ω^2_{AA}, α_{AA}, R_{AA}, showed nice behaviour with very little fluctuations. Consequently, we can deduce that the chain mixed well, once gain suggesting the stability of these parameters. we also render the quantitative way of checking convergence, and the result shows the simulation is believed to have converged. Another important point to mention is that, due to the skewness of the fertility distribution, in both regions: Somali and Addis Ababa, the posterior medians is superior to posterior mean. Therefore, posterior median instead of the mean should be chosen to be the estimated value and even for comparison with with the ML result. This also holds for other skewed region. In addition

Table 3.6: **MCMC convergence diagnostic test for Addis Ababa fertility model parameters using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.372	*passed*	0.00952	1.827	0.951
ω^2	*passed*	0.669	*passed*	0.00171	0.594	1.02
α	*passed*	0.487	*passed*	0.00103	-0.7289	0.982
R	*passed*	0.499	*passed*	0.00097	-0.2828	1.02

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

to those result for Addis Ababa and Somali region, we also managed to determine results for other regions in the study during the study period, as shown in Appendix B. All-in-all, the result for other region revealed that most of the credible intervals for the parameters obtained using the Gibbs sampler formulation were narrower than the credible intervals obtained using maximum likelihood estimation. The results obtained also suggested that the Gibbs sampler algorithm formulation was a better approach for estimating parameters for the ASFR fertility model than the other approach considered, which is maximum likelihood estimation .

Generally, both ML method and Bayesian method can be used to analyze our model, but they can be applicable to different conditions.
ML method can be applied when:

♣ **ML method can be applied when:**

- ♠ precise values of model parameters have been known
- ♠ large sample size can be obtained in the test

and,

♣ **Bayesian method can be applied when:**

- ♠ *uncertainties on the model parameters exist*
- ♠ *prior knowledge on the model parameters are available*
- ♠ *few data is available in the test*

4 HIERARCHICAL BAYESIAN MODELING

4.1 Introduction

In many social science setting, the data available for analysis comprise multiple groups. In these settings it is often plausible that any statistical model we might fit to the data need to be flexible, so as to capture variation across the groups, typically accomplished by letting some or all of the parameters vary across the groups. Examples include *survey data gathered over a set of locations (e.g., states, countries, congressional districts, etc); experimental studies deployed in multiple locations; studies of educational outcomes which the subjects are students, who are grouped in schools, which nest in school districts, which in turn nest in states*. In studies of data of this type, it is potentially of great important to deal with group-level or between-group parameter variation as well because it improves not only our estimates but also our group-level knowledge.

In spite of its importance, many common statistical methods, however, assume either:-

(i). Complete pooling model (*assumes homogeneity across groups, see Figure 4.1 or Figure 4.3 (a)) (e.g., fixed-effects meta-analysis)*; or

(ii). Complete independence across groups[15](as in *No pooling model, which estimates statistical models group-by-group, see Figure 4.2 or Figure 4.3 (c))*

to fit heterogeneous data.

4.1.1 Approach 1: Complete pooling

As its name suggests, the complete pooling approach first averages the data across all groups/regions, and then estimates the model's parameters for the averaged data. In other words, the researcher effectively analyzes the data as if they were generated by a single group/region (Nilsson et al., 2011). By and large, this approach assumes homogeneity between groups, and hence attempts to completely ignore the difference or the *variability* between (across) the groups, and combines every thing together into one big pool to make inference on the data. As to our study is concerned, the concept of complete pooling is meant to think of homogeneity across regions in the study and consider identical fertility model parameters for each of the 11 regions,

$$\theta_1^* = \cdots = \theta_7^* = \theta_{12}^* = \cdots = \theta_{15}^* = \theta^* = (R, \xi, \omega^2, \alpha) \tag{68}$$

This is to say, we assume that all 11 regions in study have the equal or common total fertility rate(R), the equal or common fertility shape pattern (α), the same average age in

[15] *Group*:the terminology *group*is here defined in the context of any number of people or things, and may mean different classes, schools,departments, regions, populations, countries, etc

fertility($\simeq \xi$), the same variability in fertility($\simeq \omega^2$), etc, and infer them with Bayesian analysis.

Nevertheless, this assumption of a common fertility model parameters for all regions,

$$\theta_j^* = \theta^* = (R, \xi, \omega^2, \alpha), \quad \forall j = 1, \cdots, J = 11 \tag{69}$$

is quite strong, and may be inappropriate and is questionable since one would expect some variation in the true rates in general; and that of Ethiopia in particular, where sound and visible disparity in fertility clearly shown among regions, as stated above. Therefore, unfortunately, this approach comes with a serious drawback. Especially, when regions/groups differ, conclusions drawn from the complete pooling approach are known to be potentially misleading and could lead to fallacy: For instance:- *the policy recommendation on fertility that would seem to work for the "average region" could perform poorly or even in a counter-productive way in any given region as it ignores any information conveyed by the region variable, which is wasteful.* In particular, Estes and others have shown that region/group differences, when ignored, can induce averaging artifacts in which the data that are averaged over regions/groups are no longer representative for any of them (Estes, 2002; Heathcote et al., 2000). For example:- *consider a situation in which one half of the women's in a certain locality is risk-seeking on woman's fertility whereas the other half is risk-averse. When complete pooling is fitted to the average data it may support the conclusion that the women are risk-neutral, a conclusion that is correct for none of the individual participant women.* Owning to these facts this approach, which *gives identical estimates for all regions*, is particularly inappropriate for this study, *whose goal is to scrutinize and make region-level inference.*

Note that in this application, the number of sampling units are large (n is large) as this approach pools all the data in all groups/regions into one big pool; and the variability is also high as it assumes homogeneity across groups. Therefore, the resulting estimated means or parameters will have large variance and low bias, which may lead to over-fitting.

4.1.2 Approach 2: Complete Independence

The assumption that all regions/groups are identical is clearly unrealistic, and this is why many researchers now estimate model parameters for each regions/groups separately (e.g.: Brown and Heathcote, 2003; Estes and Maddox, 2005; Haider and Frensch, 2002). This complete independence approach implicitly assumes that each region/group is unique and has variations across/among them; and estimates model parameters group-by-group. By

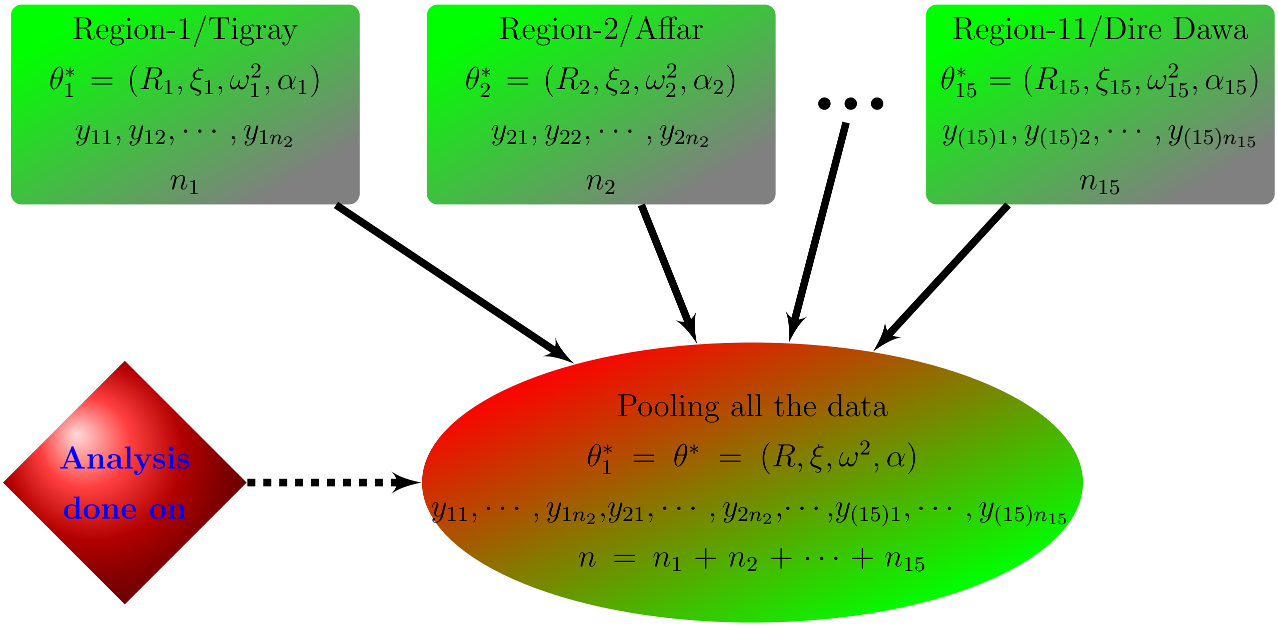


Figure 4.1: *Approach-1: Complete pooling*

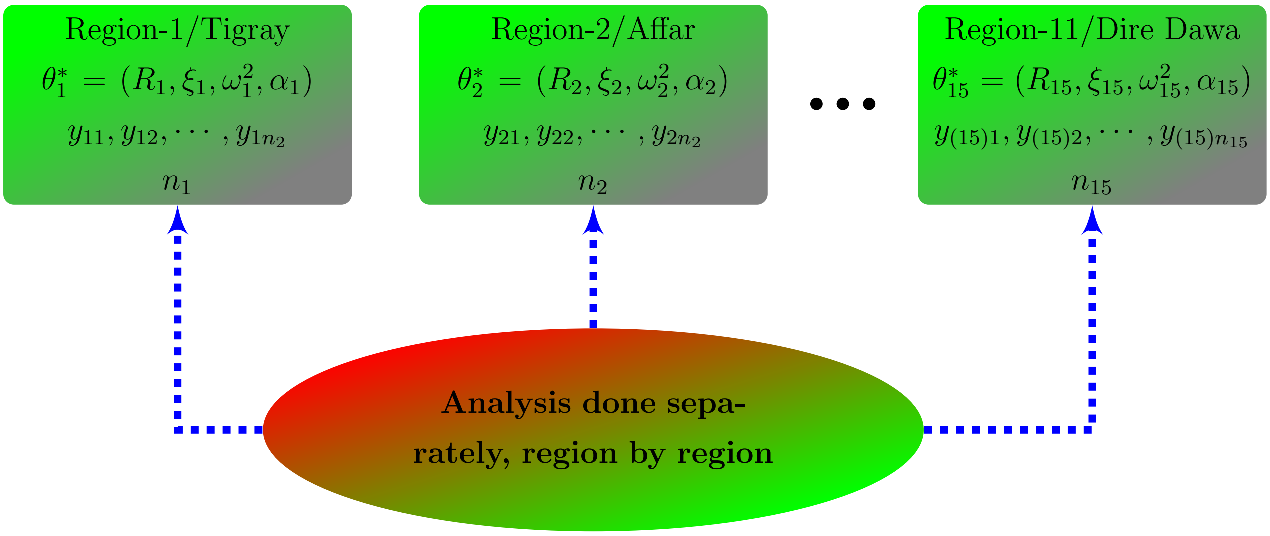


Figure 4.2: *Approach-2:Complete independence (no pooling)*

considering each region/group as a separate unit of analysis, the complete independence approach avoids the averaging artifacts that plague the complete pooling approach, and allows for statistical inferences both for the entire group and for individual region/group (Gelman and Hill, 2007; Scheibehenne and Pachur, 2013).

However, this assumption of estimating statistical models group-by-group, which in our case is region-by-region independently, is often just not feasible, if even it were desirable. The main draw back of this approach is that the amount of data in any given region can sometimes be small such that the within-group or region analysis yields parameter estimates that are relatively too noisy, unreliable and imprecise to be useful (Albert and Chib, 1993; Nilsson et al., 2011). This is because this application provides us estimates which have high bias and low variance, which may cause the problem of underfitted.

To sum up, fundamental problems are associated with all current methods to estimate parameters in ASFR model. The complete pooling model, although robust, may lead to averaging artifacts- participants are not identical. The complete independence model, although it avoids averaging artifact, may lead to noisy and extreme parameter estimates- the price that has to be paid for assuming that each participant is unique (Nilsson et al., 2011)

4.1.3 Approach 3: Hierarchical or partial pooling

As outlined previously, there is a tension here, between the *high bias/low-variance* results that might be obtained with an analysis that ignores between(among) regional-level parameter heterogeneity (*viz.*: *Complete Pooling*), and the *low-bias/high-variance* results from region-specific analyses (*viz.*: *Complete Indepence*). Therefore, hierarchical statistical models offer a principled way to compromise or trade-off this tension between the extremes of complete pooling and complete independence (Gelman and Hill, 2007; Shiffrin et al., 2008).
Roughly speaking, hierarchical, or multilevel, modeling is a statistical method that can be used to quantitatively and coherently combine heterogeneous information. It is commonly used in a variety of research areas, such as educational research, sociological science, biometric research and econometric research. A Hierarchical Data Structure is a multilevel structure consisting of higher (macro) level and lower (micro) levels. Lower levels consisting of individuals are grouped (nested) into higher levels. Higher levels are grouped again into even higher levels. For example, in an educational research context, students as individuals are nested within classes as a lower level are nested within the higher level schools, schools are within districts and so on.

In previous chapter, a "*Complete independent (no pooling) modeling*" technique had been considered to estimate parameters in our fertility model (see **??**). In such application, each of the 11 regions was thought of independent and had its own separate fertility model parameters, *i.e.*,,

$$\theta_j^* = (R_j, \xi_j, \omega_j^2, \alpha_j), \qquad \text{for} \quad j = 1, \cdots, J = 11 \tag{70}$$

(*i.e.,* every region was assumed to have its own fertility rate, R, shape pattern, α, etc) before estimating them separately (region- by-region) by means of simple Bayesian analysis. As pointed out above, one of the major stumbling block with previous method when utilized

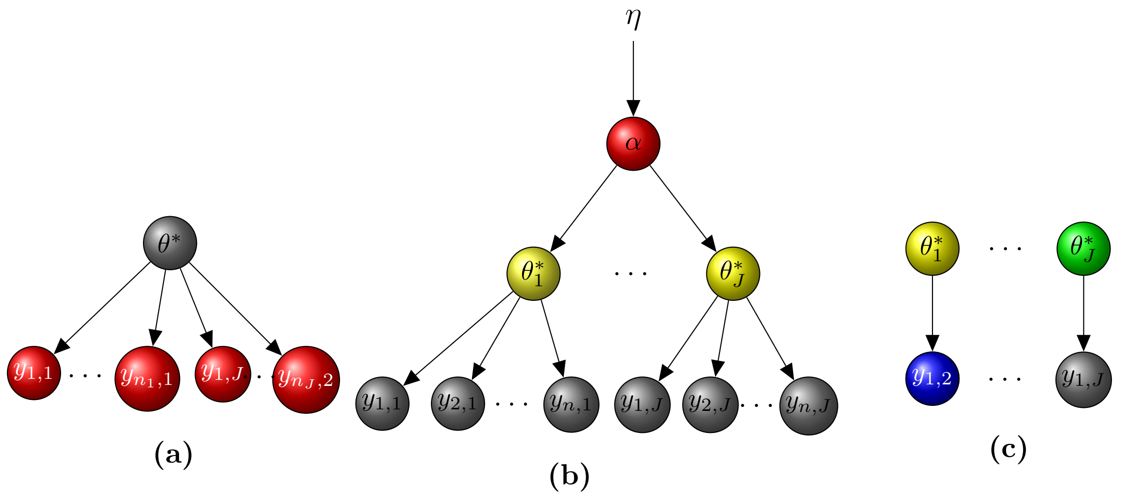


Figure 4.3: *The Classical estimates: complete pooling (a) and no pooling (c); and hierarchical model (b)*

to our dataset is that those region-specific ASFR parameters, $\theta_j^* = (R_j, \xi_j, \omega_j^2, \alpha_j)$ for $j = 1, \cdots, J = 11$, could be in imprecise estimates, particularly for the regions which contain smaller number of observations, number of mothers in the reproductive age, namely Somali region, Affar region, SNNP, etc. Therefore, in the entire of this chapter, Hierarchical Bayesian models will be instrumented as a possible solution to trade-off the predicaments in parametric estimation (over-fitting/under-fitting) in using non-hierarchical Complete-independent and Complete-pooling model reviewed above (Albert and Chib, 1993; Nilsson et al., 2011). The implementation of this modeling approach has a lot of plus points to offer in our fertility model parameter estimation, specifically compared to non-hierarchical alternatives discussed earlier (Schaub and Kéry, 2012). The orthodox and trivial ones are:

(*i*) . **"Borrowing strength" via exchangeability:-** the purpose of hierarchical Bayesian approach is to find an optimal compromise between the extremes of complete pooling and complete independence; the imposed group-level structures simultaneously inform the individual level, such that the individual estimates can borrow strength from the information available about the other individuals in a sample (Gelman et al., 2014). This means, in our case, for larger regions we have good estimates, for smaller regions we may be able to borrow information from other regions to obtain more accurate estimates.

(*ii*) . **Hierarchical Modeling as a "Shrinkage" Estimator:-** In the hierarchical approach, individual parameter estimates that are deemed unlikely given the overall distribution of parameter values (because they are located at the periphery of the distribution) are "corrected" by pulling them closer toward the group mean. This property, which is an attractive feature of borrowing of strength, sometimes referred to as *shrinkage*[16], prevents potentially unreliable information from having a disproportionate influence on the group level (Nilsson et al., 2011) and "dampers" the sample size effects. Therefore, the estimates for regions with large sample sizes are shrunk less than those based on smaller sample sizes (Dey et al., 2000). For these reasons, it has been argued that hierarchical methods provide a more thorough and efficient evaluation of models in many studies (Shiffrin et al., 2008; van Ravenzwaaij et al., 2011).

(***iii***) . **Hierarchical modeling as a "Semi-Pooling" estimator:-** Hierarchical modeling is a compromise between *complete-independent (no pooling)*, which models unique characteristics of each region but ignores shared information, and *complete- pooling*, in which the opposite is the case. Hierarchical Bayesian modeling increased prediction accuracy over *no or complete pooling* by making optimal use of information for regions represented in the estimation set. In particular, the estimation of each fertility model parameters, $\theta_j^* = (R_j, \xi_j, \omega_j^2, \alpha_j)$ is improved by using the fertility data from the other regions. That is, it has potential to provide parameter estimates that are less prone to measurement error, and thus more stable (Atkinson and Nevill, 1998). This advantage is well justified on theoretical grounds (Gelman and Hill, 2007; Rouder and Lu, 2005), and the hierarchical approach has also proved successful when applied to empirical data (e.g., Rouder et al., 2008; Scheibehenne and Studer, 2014) .

(***iv***) . **Hierarchical modeling as regional (group) level comparison:-** Hierarchial Bayesian model allows the mothers for each region to have its own fertility model parameters, $\theta_j^* = (R_j, \xi_j, \omega_j^2, \alpha_j)$, but it also infers each of those parameters, $\theta_j^* = (R_j, \xi_j, \omega_j^2, \alpha_j)$ as coming from a common population distribution. Hence, hierarchical modeling techniques might be beneficial for comparisons on the regional-level (Gelman and Hill, 2007), where the goal is not to improve reliability on the individual level but to drive robust estimates for each region (Scheibehenne and Pachur, 2013). Thus, it is extraordinarily significant for this application, one of it principal objectives of which is to scrutinize and make region-level inference, not inference at national level.

(***v***) . **Hierarchical modeling as easing robustness:-**Although parameters are traditionally estimated independently for each single participant, it has recently been proposed that more reliable estimates might be achieved by using hierarchical Bayesian

[16] If the present fertility rate of region, say k is lower than the average mean, its predictive score will be pushed up because this first poor performance may happen due to bad luck. In reverse case, the predictive will be diminished. This effect is known as the *shrinkage effect*

procedures, which exploit group-level distributions to inform individual-level estimation (e.g., Gelman and Hill, 2007; Lee and Webb, 2005). As was pointed out by Nilsson et al. (2011), this borrowing of strength should increase the reliability of parameter estimates for individual participants, and thus provide more robust results (see also van Ravenzwaaij et al., 2011)

Though increasingly popular, Bayesian hierarchical implementations have been developed for only relatively few fertility models in general and ASFR model in particular (*but see:-* Griggio, 2013/2014; Canale and Scarpa, 2015). Below we develop, to our knowledge for the first time, a hierarchical model for estimating ASFR model parameters. Consequently, in this chapter, we will extend all the works done in previous chapter and illustrate a full implementation of Bayesian hierarchical modeling based on data augmentation and Markov chain Monte Carlo (MCMC) sampling, and demonstrate the ease of implementation and accuracy of results. Hence, in the remaining work, general impression on the theoretical framework of Hierarchical Bayesian Inference will be given in subsection 3.2; Brief discussion pertaining to the notion of exchangeable will follow in subsection 3.3; Hierarchical model development and its estimation mechanism will also addressed in subsection 3.4 and subsection 3.5; respectively. Last but not certainly the least, we will discuss the empirical result in subsection 3.6.

4.2 Hirarchical Bayesian Inference

In chapter 2, we introduce Bayesian modeling in which the hyperparameters, *i.e.*, the parameters of prior distribution, are **known**. However in many real-world applications, *such information is not available.* Thus, hierarchical Bayesian model is introduced to resolve this predicament. From a broadest point of view, hierarchical model is a Bayesian statistical model, with many levels and structured in terms of a sequence of conditional distributions with the capacity to cope with high-dimensional complex models typically needed for inferences and predictions (Clark, 2005, 2007; Clark and Gelfand, 2006; Cressie et al., 2009; Wikle, 2003) , where the prior distribution, $p(\boldsymbol{\theta})$, is decomposed into several or multiple conditional levels (*i.e.*, hierarchies) of probability distributions(Robert, 2001), as:-

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1), p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2), p_3(\boldsymbol{\theta}_2|\boldsymbol{\theta}_3), \cdots, p_{n-1}(\boldsymbol{\theta}_{n-1}|\boldsymbol{\theta}_n) \tag{71}$$

that represent relationships between information arising *within* single population or group, as well as relationships between information arising from *different* populations or groups (Kwok and Lewis, 2011).

The parameters $\boldsymbol{\theta}_i$ are called *hyper-parameters*[17] of level i, for $1 \leq i \leq n$. The most common hierarchical Bayesian model is the case when $n = 2$. At the first stage, a distribution

[17] Hyper-parameters are the parameters of prior parameters

for the data given parameters is specified. At the second stage, prior distributions for parameters given hyper-parameters are specified and distributions for hyper-parameters are specified at the third stage. Figure 4.4 and Figure 4.5 show a typical hierarchical Bayesian model of our data set. The "hierarchy" arises because the model for the parameters sits "above" the model for the data and analogously,the model for the hyperparameters sits "above" the model for the parameters.

As outlined in subsection 2.2, suppose that $\boldsymbol{y} = (y_{1(1)}, \cdots, y_{ij}, \cdots y_{n_J(J)})$, where y_{ij} corresponds to the age of the i^{th} mother in the reproductive age(between 15-49 year old) living in the j^{th} region in the study period, where, $j = 1, \ldots, J$, with $J = 11$ referring the total number of regions; and $i = 1, \ldots, n_j$, with n_j, the total sampled mothers between 15-49 taken from the j^{th} region. Let $\boldsymbol{y}$, the data, be *i.i.d.* drawn from a distribution with *unknown parameters* $\boldsymbol{\theta}^* = (\xi_j, \omega_j^2, \alpha_j, R_j)$. The unknown parameters are drawn from a **prior distribution** with *unknown hyperparameters* $\boldsymbol{\phi} = (a, b)$, which themselves are random variables and are drawn from a distribution with parameters η then the joint distribution of the parameter and hyperparameters is

$$p(\boldsymbol{\theta}^*, \boldsymbol{\phi}) = p(\boldsymbol{\theta}^*|\boldsymbol{\phi})p(\boldsymbol{\phi}) \tag{72}$$

and, hence, the posterior density for hierarchical model is

$$\begin{aligned} p(\boldsymbol{\theta}^*, \boldsymbol{\phi}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\theta}^*, \boldsymbol{\phi})p(\boldsymbol{\theta}^*, \boldsymbol{\phi}) \\ &= p(\boldsymbol{y}|\boldsymbol{\theta}^*, \boldsymbol{\phi})p(\boldsymbol{\theta}^*|\boldsymbol{\phi})p(\boldsymbol{\phi}) = p(\boldsymbol{y}|\boldsymbol{\theta}^*, \boldsymbol{\phi})p(\boldsymbol{\theta}^*|\boldsymbol{\phi})p(\boldsymbol{\phi}|\eta) \end{aligned} \tag{73}$$

Table 4.1 shows the form of a general hierarchical model with three levels.

Table 4.1: **Hierarchical model structure with three levels**

Level	Variables	Densities	Description
1(Data level[a])	*Observation*|*Parameters*	$p(\boldsymbol{y}\|\boldsymbol{\theta}, \boldsymbol{\phi})$	Model for the data or Likelihood model
2(Parameter level[b])	*Parameters*|*Hyperparameters*	$p(\boldsymbol{\theta}\|\boldsymbol{\phi})$	Between-group model or "Prior distribution"
3(Hyperparameters[c])	*Hyperparameters*	$p(\boldsymbol{\phi})$	Hyperprior distribution

[a] *aka*:The *observational model* for the data ; [b] *aka*: The *structural model* for the parameters;
[c] *aka*: The *hyperparameter model* for the parameters of the structural model

Bayesian inference allows the flexibility in explicitly modeling hierarchical models. However, one of the common problems in the Bayesian hierarchical models is that the posterior distributions may not tractable algebraically in many cases, as the hierarchical models considered in this study. Moreover, posterior densities for the hierarchical models often lead to nonstandard densities. To overcome such analytical limitations, sampling-based estimation methods have been used. Markov Chain Monte-Carlo (MCMC) methods (Gilks, 1986) using Gibbs sampler and the Metropolis-Hastings algorithm, Rejection Sampling, etc, are widely applied to generate a large number of samples from posterior distributions. Any distribution summary (such as mean, median or quantiles) of the posterior distributions of model parameters or unknowns can then be approximated by their sample analogue

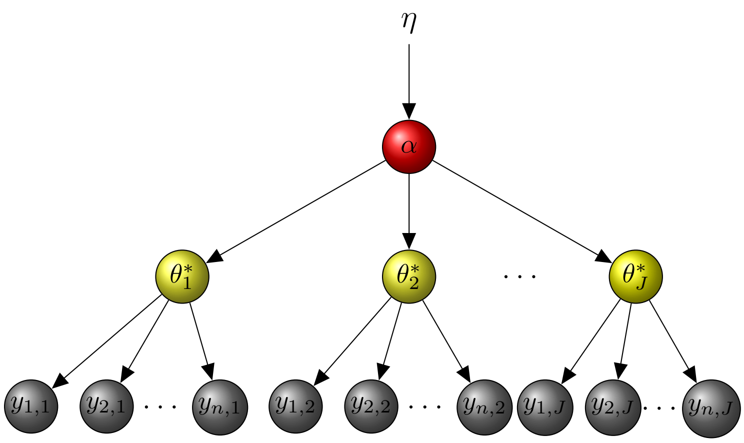


Figure 4.4: A hierarchical model in full Bayesian framework

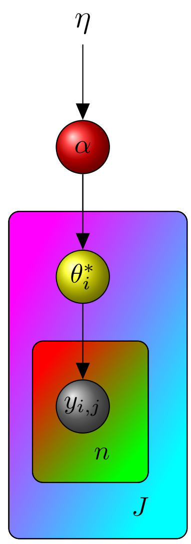


Figure 4.5: An equal model with a plate representation.

4.3 Exchangeability

A tacit assumption in statistic learning is that the $N = \sum_{j=1}^{J} n_j$ observations $y = (y_{11}, \cdots, y_{n_jJ})$ are exchangeable, *i.e.*, the joint distribution $p(y_{11}, \cdots, y_{n_jJ})$ of the data is invariant if the indices of the variables are permuted. Let $\nu = \{\nu(11), \cdots, \nu(n_jJ)\}$ denote a permutation of the indies from 1 to N, the exchangeability assumption yields:

$$p(y_{11}, \cdots, y_{n_jJ}) = p\left(y_{\nu(11)}, \cdots, y_{\nu(n_jJ)}\right) \tag{74}$$

Furthermore, when the number of the variables is infinite, i.e. $N \longmapsto \infty$, the variables are *infinite exchangeable*, if any finite subset of variables are exchangeable. Based on the exchangeability assumption, it is natural to model the data as independently and identically distributed given model parameters θ^*,

$$p(y_{11}, \cdots, y_{n_jJ}|\theta^*) = \prod_{i=1}^{n_j} \prod_{j=1}^{J} p(y_{ij}|\theta^*) \tag{75}$$

The exchangeability relations in a model can be illustrated in a graphical representation, referred to as *plate*, which is a template that allows the subgraphs can be replicated.

Figure 4.4 shows the model discussed in subsection 4.1. Figure 4.5 shows the equal model in a *plate*. In the plate language, variables (not random) are represented directly by their names, e.g. the hyperparameters α. Random variables, e.g. θ, are represented as circles with their names. The n exchangeable variables $\{y_1, \cdots, y_n\}$are represented as a single variable y_i in a rectangle. The number n at the corner specifies the number of the variables. An arrow, e.g. from α to θ denotes that the probability distribution of θ is conditioned on α. Note, that the arrow from θ to y_i specifies each of the n variables y_i depends on θ. The plate representation is often used to illustrate probability models. It clarifies the exchangeability relations in a compact and elegant way.

4.4 Model Development: Hierarchical Bayesian Inference of Binary fertility data using Skew Normal latent Variable

4.4.1 Data Augmentation Approach: *using Special SN latent variable model*

To illustrate hierarchical Bayesian analysis, we extend the Bayesian setting discussed in Chapter 3. In addition, the model development is carried out in the same fashion as the Bayesian analysis of the previous section. Consequently, we consider, once more, the binary random variable given in previous chapter of subsubsection 4.6.1, W_{ij}, the status of the i^{th} childbearing age mother (*Have birth*; *No birth*), where, $i = 1, 2, \cdots, n_j$; n_j is total number of childbearing age mothers in the study, living in the j^{th} region, where, $j = 1, \ldots, J$, with total number of regions in the study $J = 11$, *i.e.*, symbolically:-

$$W_{ij}(x_{ij}) = \begin{cases} 1 & \text{if the } i^{th} \text{ woman of age 15-49 living in } j^{th} \text{ region gives} \\ & \text{birth to a child during the year } [x, x+1), \\ 0 & \text{elsewhere} \end{cases} \tag{76}$$

for $i = 1, 2, \ldots, n_j; j = 1, \ldots, J$. Thus, mother‘s birth status is a Bernoulli distributed random variable with fertility probability of

$$\pi_{ij} = \Pr\left(W_{ij} = 1 | R_j, \xi_j, \omega_j^2, \alpha_j\right) = R_j \cdot \ \mathrm{SN}(x_{ij}; \xi_j, \omega_j^2, \alpha_j)$$

In a nutshell, we have

$$\begin{aligned} &W_{ij} \sim \mathrm{Bern}\left(\pi_{ij}\right), \\ &\pi_{ij} = \Pr\left(W_{ij} = 1 | R_j, \xi_j, \omega_j^2, \alpha_j\right) = R_j \cdot \ \mathrm{SN}(x_{ij}; \xi_j, \omega_j^2, \alpha_j) \end{aligned} \tag{77}$$

However, as mentioned in subsubsection 4.6.1, this consideration is equivalent with defining mother's birth status, W as:-

$$W_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \leq 0 \end{cases} \tag{78}$$

where

$$z_{ij} \sim R_j \cdot \ \mathrm{SN}(\xi_j, \omega_j^2, \alpha_j) \tag{79}$$

which is a latent (unobserved) variable corresponding to W_{ij}.
Thus, the sampling distribution of z_{ij} conditional on $R_j, \xi_j, \omega_j,$ and α_j for $j = 1, \ldots, J,$ with $J = 11$ can be provided as

$$p\left(z_{ij}|R_j, \xi_j, \omega_j^2, \alpha_j\right) \propto R_j \cdot \phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j \frac{z_{ij}-\xi_j}{\omega_j}\right) \tag{80}$$

and, thus, the joint density is

$$p\left(\boldsymbol{z}|R_j, \xi_j, \omega_j^2, \alpha_j\right) \propto \prod_{i=1}^{n_j}\prod_{j=1}^{J} R_j \cdot \phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j \frac{z_{ij}-\xi_j}{\omega_j}\right) \tag{81}$$

In the same fashion, we can write the conditional for $\boldsymbol{W}$ as:-

$$p\left(\mathrm{w}_{ij}|z_{ij}, R_j, \xi_j, \omega_j^2, \alpha_j\right) = \mathbb{I}\left(z_{ij} > 0\right)\mathbb{I}\left(W_{ij} = 1\right) + \mathbb{I}\left(z_{ij} \leq 0\right)\mathbb{I}\left(W_{ij} = 0\right) \tag{82}$$

with $\mathbb{I}$ denoting the indicator function which assumes the value 1 if the statement on the parenthesis is true, and 0, otherwise.Thus, the joint conditional distribution is

$$\begin{aligned} p\left(\boldsymbol{w}|z_{ij}, R_j, \xi_j, \omega_j^2, \alpha_j\right) = \prod_{i=1}^{n_j}\prod_{j=1}^{J}\Big[&\mathbb{I}\left(z_{ij} > 0\right)\mathbb{I}\left(W_{ij} = 1\right) \\ &+ \mathbb{I}\left(z_{ij} \leq 0\right)\mathbb{I}\left(W_{ij} = 0\right)\Big] \end{aligned} \tag{83}$$

By introducing the z_{ij}'s, we are "augmenting" the observed data $\mathbf{y} = \left(y_{11}, \cdots, y_{n_jJ}\right)'$ with latent data $\mathbf{z} = \left(z_{11}, \cdots, z_{n_jJ}\right)'$ to form the *complete-data vector*,$(\boldsymbol{w}, \mathbf{z})$.

4.4.2 Stage I (individual level, within-region model): Complete-data Likelihood Model

This *complete-data likelihood function* or *augmented data joint density*, $L\left(R, \xi, \omega^2, \alpha|\boldsymbol{w}, \boldsymbol{z}\right)$ is obtained by putting together the pieces in Equation 81 and Equation 83, as given in Equation 84.

$$\begin{aligned} L\left(R_j, \xi_j, \omega_j^2, \alpha_j|\boldsymbol{w}, \boldsymbol{z}\right) &= p\left(\boldsymbol{w}, \boldsymbol{z}|R_j, \xi_j, \omega_j^2, \alpha_j\right) \\ &= p\left(\boldsymbol{w}|\boldsymbol{z}, R_j, \xi_j, \omega_j^2, \alpha_j\right) p\left(\boldsymbol{z}|R_j, \xi_j, \omega_j^2, \alpha_j\right) \\ &= \prod_{j=1}^{J}\prod_{i=1}^{n_j}\left\{\mathbb{I}\left(z_{ij} > 0\right)\mathbb{I}\left(W_{ij} = 1\right) + \mathbb{I}\left(z_{ij} \leq 0\right)\mathbb{I}\left(W_{ij} = 0\right)\right\} \\ &\cdot \prod_{j=1}^{J}\prod_{i=1}^{n_j}\frac{R_j}{\omega_j}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha\frac{z_{ij}-\xi_j}{\omega_j}\right) \\ &= \prod_{j=1}^{J}\prod_{i=1}^{n_j}\Big[\left[\mathbb{I}\left(z_{ij} > 0\right)\mathbb{I}\left(W_{ij} = 1\right) + \mathbb{I}\left(z_{ij} \leq 0\right)\mathbb{I}\left(W_{ij} = 0\right)\right] \\ &\cdot \frac{R_j}{\omega_j}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha\frac{z_{ij}-\xi_j}{\omega_j}\right)\Big] \end{aligned} \tag{84}$$

4.4.3 Stage II (between Regions, Second stage/regional-level priors):Prior Distributions

The parameter θ_j^* in the second stage of the hierarchy is assumed exchangeable across the J regions, and also arising from a *Gamma distribution* with hyperparameters (a, b)

$$(\theta_j^*|a, b) \sim \text{Gamma}(a, b) \tag{85}$$

for $j = 1, \cdots, J = 11$, where $a > 0$ and $b > 0$ are pre-specified by the investigator.

(a). Prior for R_j:

$$R_j \sim \text{Gamma}\,(a, b)\,, \text{ where } \; a, b > 0 \tag{86a}$$

(b). Prior for ξ_j:

$$\xi_j \sim \text{N}\left(\xi_o, \kappa\omega_j^2\right)\,, \text{ where } \; \xi_o \in \Re \; \text{ and } \; \omega_j^2 > 0 \tag{86b}$$

(c). Prior for ω_j^2:

$$\begin{aligned} &\omega_j^2 \sim \text{InvGamma}\,(c, d)\,, \text{ where } \; c, d > 0 \\ &\text{and each hyperparameter is small } (0.01 \text{ or } 0.001) \end{aligned} \tag{86c}$$

(d). Prior for α_j : we assume two prior cases:-

Case-1: when α_j is supposed to to be normally distributed parameter, i.e.

$$\alpha_j \sim \text{N}\left(\alpha_o, \psi_o^2\right), \text{where } \alpha_o \in \Re \text{ and } \psi_o^2 > 0 \text{ or} \tag{86d}$$

Case-2: when α is supposed to have a skew normal distribution, i.e.,

$$\alpha_j \sim \; \text{SN}\left(\alpha_o, \psi_o^2, \lambda_o\right)\,, \text{ where } \; \alpha_o, \lambda_o \in \Re \; \text{ and } \psi_o^2 > 0 \tag{86e}$$

4.4.4 Stage III (common across all regions):Hyperparameter Distribution

One extension to Bayesian model deemed in Chapter-II is to consider a set of hyperprior distributions for the parameters of our model. The model given in Equation 84 has eight hyperparameters, a, b, ξ_o, c, d, α_o, ψ_o, and λ_o. Here, we assume all the hyperparameters, except a and b, are fixed and known, and hence, must be specified before analysis. Therefore, in this thesis we concern ourselves only with the first two of these, which are the hyperparameter of the prior R_j. The reason behind why hierarchical structure is preferred only on this regional fertility rate parameter is that the other parametric vectors were not varying as large as this fertility rate parameter across regions (*For more, see the result of Bayesian analysis, Chapter-II*).

Priors for these hyperparameters, a and b, are often taken to be *gamma distributions* with parameters (m_o, n_o) and (r_o, s_o) respectively

$$(a|m_o, n_o) \sim \text{Gamma}(m_o, n_o) \tag{87a}$$

$$(b|r_o, s_o) \backsim \text{Gamma}(r_o, s_o) \tag{87b}$$

where all of the parameters m_o, n_o, r_o and r_o take on positive values, with $m_o, n_o, r_o, s_o \geq 1$. The *gamma distributions* hyperprior is chosen for the hyperparameter a because of conjucacy. Various non-informative prior distributions for b have been suggested in Bayesian literature and software, including an improper uniform density on b (Gelman et al., 2003), proper distributions such as $b \backsim inverse - gamma(0.001, 0.001)$ Spiegelhalter et al. (2004) , and distributions that depend on the data-level variance (Box and Tiao, 1973). In this thesis, we explore and make recommendations for prior distributions for b a similar *gamma distribution*. The main difference between the two- and three-stage models is that the third stage provides a hyperprior over the R_j parameters that should provide for improved estimation when the R's are distributed as assumed

4.4.5 Joint Posterior Distributions of the augmented data

Taking in to account the assumption of prior independence and making use of Bayes' theorem, the joint posterior distribution of the model parameters and the augmented data,

$$p\left(R_j, \xi_j, \omega_j^2, \alpha_j, a, b, z|\mathrm{w}\right)$$

is proportional to the product of the complete-data likelihood representation in Equation 81, and the prior and hyperprior specifications given in Equation 82 and Equation 87a-b up to normalizing constant,*i.e.*,

$$\begin{aligned}
p\left(R_j, \xi_j, \omega_j^2, \alpha_j, a, b, z|\mathrm{w}\right) &\propto (\text{ Complete-data Likelihood }) \times (\text{ Prior }) \times (\text{ Hyperprior }) \\
&\propto p\left(\mathrm{w}, z|R_j, \xi_j, \omega_j^2, \alpha_j, a, b\right) p\left(R_j, \xi_j, \omega_j^2, \alpha_j|a, b\right) p\left(a, b\right) \\
&\propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\left[\left[\mathbb{I}\left(z_{ij}>0\right)\mathbb{I}\left(W_{ij}=1\right)+\mathbb{I}\left(z_{ij}\leq 0\right)\mathbb{I}\left(W_{ij}=0\right)\right]\right. \\
&\left.\cdot\frac{R_j}{\omega}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j\frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \times p\left(R_j, \xi_j, \omega_j^2, \alpha_j|a, b\right) p\left(a, b\right) \\
&\propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\left[\left[\mathbb{I}\left(z_{ij}>0\right)\mathbb{I}\left(W_{ij}=1\right)+\mathbb{I}\left(z_{ij}\leq 0\right)\mathbb{I}\left(W_{ij}=0\right)\right]\right. \\
&\left.\cdot\frac{R_j}{\omega_j}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j\frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \\
&\times p\left(R_j\right) \times p\left(\xi_j\right) \times p\left(\omega_j^2\right) \times p_k\left(\alpha\right) \times p\left(a\right) \times p\left(b\right) \qquad (88)
\end{aligned}$$

, where $k = 1, 2$ and the following prior and hyperprior distributions

$$\begin{aligned}
&\underline{Prior\ Distributions} : -\\
&p\left(R_j\right) = \frac{b^a}{\Gamma(a)} R_j^{a-1} e^{-bR_j} = \frac{b^a}{\Gamma(a)} R_j^{a-1} e^{-bR_j},\ a, b > 0\\
&p\left(\xi_j\right) = \frac{1}{\sqrt{2\pi\kappa\omega_j^2}} e^{-\frac{1}{2\kappa\omega_j^2}(\xi_j - \xi_o)^2} \propto (\omega_j)^{-1} e^{-\frac{1}{2\kappa\omega_j^2}(\xi_j - \xi_o)^2},\ c, d > 0\\
&p\left(\omega_j^2\right) = \frac{d^c}{\Gamma(c)} \left(\omega_j^2\right)^{-c-1} e^{-d/\omega_j^2} \propto \left(\omega_j^2\right)^{-c-1} e^{-d/\omega_j^2},
\end{aligned}$$

$$\begin{aligned}
&\underline{Prior\ Distributions} : -\\
&p\left(\omega_j^2\right) = \frac{d^c}{\Gamma(c)} \left(\omega_j^2\right)^{-c-1} e^{-d/\omega_j^2} \propto \left(\omega_j^2\right)^{-c-1} e^{-d/\omega_j^2}, && (89a)\\
&p_1\left(\alpha_j\right) = \frac{1}{\sqrt{2\pi\psi_o^2}} e^{-\frac{1}{2\psi_o^2}(\alpha_j - \alpha_o)^2} \propto e^{-\frac{1}{2\psi_o^2}(\alpha_j - \alpha_o)^2} \propto \phi\left(\frac{\alpha_j - \alpha_o}{\psi_o}\right), \text{and}\\
&p_2\left(\alpha_j\right) = \frac{2}{\psi_o} \phi\left(\frac{\alpha_j - \alpha_o}{\psi_o}\right) \Phi\left(\lambda_o \frac{\alpha_j - \alpha_o}{\psi_o}\right) \propto \phi\left(\frac{\alpha_j - \alpha_o}{\psi_o}\right) \Phi\left(\lambda_o \frac{\alpha_j - \alpha_o}{\psi_o}\right)\\
&\underline{Hyperprior\ Distributions} : -\\
&p\left(a\right) = \frac{n_o^{m_o}}{\Gamma(m_o)} a^{m_o - 1} e^{-(n_o)\cdot a} \propto a^{m_o - 1} e^{-(n_o)\cdot a},\ a, m_o, n_o > 0 && (89b)\\
&p\left(b\right) = \frac{s_o^{r_o}}{\Gamma(r_o)} b^{r_o - 1} e^{-(s_o)\cdot b} \propto b^{r_o - 1} e^{-(s_o)\cdot b},\ b, r_o, s_o > 0
\end{aligned}$$

4.4.6 Full Conditional Distributions of Augmented data

Our aim is to construct an MCMC Algorithm to get samples from the joint posterior distribution of $R_j, \xi_j, \omega_j^2, \alpha_j, a, b, z$ and w. To do this we follow the approach of a Gibbs sampler, that is sequentially updating the parameters $R_j, \xi_j, \omega_j^2, \alpha_j, a,$ and b by using the full conditional distributions, *i. e.*, the distribution of one parameter given all others and the data. In this section, we derive the full conditional distribution of the parameters.

$(\boldsymbol{i})$.**Sampling Z**:

Given $\xi_j, \omega_j^2, \alpha_j, R_j, a, b$ and $\boldsymbol{w}$, the variables z_i are independent, and $z_{ij}|\xi_j, \omega_j^2, \alpha_j, R_j, a, b,$ and $\boldsymbol{w}$ is distributed as the $SN(\xi_j, \omega_j^2, \alpha_j)$ truncated at the left by 0 if $w_{ij} = 1$ and truncated at the right by 0 if $w_{ij} = 0$, for $i = 1, \cdots, n_j$ and $j = 1, \cdots, J$. That is,

$$
\begin{aligned}
p(z_{ij}|R_j, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b) &\propto \Bigg[\left[\mathbb{I}\left(z_{ij} > 0\right) \mathbb{I}\left(W_{ij} = 1\right) + \mathbb{I}\left(z_{ij} \leq 0\right) \mathbb{I}\left(W_{ij} = 0\right)\right] \\
&\quad \cdot \frac{R_j}{\omega_j} \phi\left(\frac{z_{ij} - \xi_j}{\omega_j}\right) \Phi\left(\alpha_j \frac{z_{ij} - \xi_j}{\omega_j}\right) \Bigg] \\
&\propto \begin{cases} \frac{1}{\omega_j} \phi\left(\frac{z_{ij} - \xi_j}{\omega_j}\right) \Phi\left(\alpha_j \frac{z_{ij} - \xi_j}{\omega_i}\right) \mathbb{I}\left(z_{ij} > 0\right) & \text{if } \mathrm{w}_{ij} = 1 \\ \\ \frac{1}{\omega_j} \phi\left(\frac{z_{ij} - \xi_j}{\omega_j}\right) \Phi\left(\alpha_j \frac{z_{ij} - \xi_j}{\omega_j}\right) \mathbb{I}\left(z_{ij} \leq 0\right) & \text{if } \mathrm{w}_{ij} = 0 \end{cases} \\
&\propto \begin{cases} \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) \mathbb{I}\left(z_{ij} > 0\right) & \text{if } \mathrm{w}_{ij} = 1 \\ \\ \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) \mathbb{I}\left(z_{ij} \leq 0\right) & \text{if } \mathrm{w}_{ij} = 0 \end{cases} \qquad (90)
\end{aligned}
$$

Therefore,

$$
z_{ij}|R_j, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b \backsim \begin{cases} C_1 \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) & \text{if } z_{ij} > 0 \;\&\; \mathrm{w}_{ij} = 1 \\ \\ C_2 \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) & \text{if } z_{ij} \leq 0 \;\&\; \mathrm{w}_{ij} = 0 \end{cases} \qquad (91)
$$

after normalizing the marginal distribution. In this case, C_1 and C_2 are normalization constants and provided as

$$
C_1 = \frac{1}{\int_0^{+\infty} \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) dz_{ij}} = \frac{1}{cdf_{\mathrm{SN}}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right)}
$$

and

$$
C_2 = \frac{1}{\int_{-\infty}^0 \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) dz_{ij}} = \frac{1}{1 - \int_0^{\infty} \mathrm{SN}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right) dz_{ij}} = \frac{1}{1 - cdf_{\mathrm{SN}}\left(z_{ij}; \xi_j, \omega_j^2, \alpha_j\right)}
$$

Like the previous chapter, the Accept-Reject procedure has been rendered to simulate this conditional from truncated skew-normal distribution in Equation 91.

(***ii***).**Sampling R**:

This is one of the important parameters in our model, which can use to make fertility within and between regions in our model. In order to draw samples for this parameter, we have to determine its conditional distribution, given as follows:

$$
\begin{aligned}
p(R_j|\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b) &\propto \left[\prod_{i=1}^{n} R_j\right] \cdot R_j^{a-1} e^{-bR_j} \\
&\propto R_j^n \cdot R_j^{a-1} e^{-bR_j} \\
&\propto R_j^{n+a-1} e^{-bR_j} \qquad (92)
\end{aligned}
$$

Therefore,

$$R_j|\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b \backsim \text{Gamma}\ (n+a, b) \quad \text{for} \quad j=1,\cdots,J=11;\ a,b>0 \tag{93}$$

.

$(\boldsymbol{iii})$.**Sampling** (ξ_j, ω_j^2) **and** α_j::

Like what has been done in the previous chapter, here again it is straight forward to sample (ξ_j, ω_j^2) and α_j for $j = 1, \cdots, J = 11$. The sampling can be done from their conditional distributions in Equation 94 and Equation 95 respectively.

$$\begin{aligned} p\left(\xi_j, \omega_j^2|\alpha_j, R_j, \boldsymbol{z}, \boldsymbol{w}, a, b\right) &\propto \left\{\prod_{i=1}^{n} \frac{1}{\omega_j}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j\frac{z_{ij}-\xi_j}{\omega_j}\right)\right\} \times p\left(\xi_j\right) \times p\left(\omega_j^2\right) \\ &\propto \left\{\prod_{i=1}^{n} \text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)\right\} \times \text{N}\left(\xi_j;\xi_o,\kappa\omega^2\right) \times \text{Inv.Gamma}\left(\omega_j^2;c,d\right) \end{aligned} \tag{94}$$

and

$$\begin{aligned} p\left(\alpha_j|\textit{the rest}\right) &\propto \begin{cases} \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j\frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \times p_1\left(\alpha_j\right), & \text{where } \ p_1\left(\alpha_j\right) \text{ is} \\ & \text{a Normal prior} \\ \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j\frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \times p_2\left(\alpha_j\right), & \text{where } \ p_2\left(\alpha_j\right) \text{ is} \\ & \text{a SN prior} \end{cases} \\ &\propto \begin{cases} \left\{\prod_{i=1}^{n} \text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)\right\} \times \text{N}\left(\alpha_j;\alpha_o,\psi_o^2\right) \\ \left\{\prod_{i=1}^{n} \text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)\right\} \times \text{SN}\left(\alpha_j;\alpha_o,\psi_o^2,\lambda_o\right) \end{cases} \end{aligned} \tag{95}$$

Therefore,

$$\alpha_j|\xi_j, \omega_j^2, R_j, \boldsymbol{z}, \boldsymbol{w}, a, b \backsim \begin{cases} \left\{\prod_{i=1}^{n} \text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)\right\} \times \text{N}\left(\alpha_j;\alpha_o,\psi_o^2\right) \\ \left\{\prod_{i=1}^{n} \text{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)\right\} \times \text{SN}\left(\alpha_j;\alpha_o,\psi_o^2,\lambda_o\right) \end{cases} \tag{96}$$

$(\boldsymbol{a})$.**Sampling** (ξ_j, ω_j^2)::

From Equation 84, it is trivial that

$$z_{ij} \backsim \mathrm{SN}(\xi_j, \omega_j^2, \alpha_j), \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, J \tag{97}$$

As in the previous section, it is, therefore, admissible to describe this a skew normal distributed auxiliary random variable z_{ij} using the stochastical representation of the model (Henze, 1986), which is of the form :-

$$z_{ij} = \xi_j + \omega_j \delta \eta_{ij} + \omega_j \sqrt{1 - \delta_j^2} \epsilon_{ij} \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, J \tag{98}$$

where, $\epsilon_{ij} \backsim N(0, 1)$, the standard normal distribution and $\eta_{ij} \backsim TN_{[0,\infty)}(0, 1)$, the half normal distribution. Results from Equation 98 follow that the conditional distribution $z_{ij}|\eta_{ij}, \omega_j^{-2}, \xi_j, \delta_j$ is a normal distribution with mean

$$\begin{aligned} E(z_{ij}) =& E\left\{\xi_j + \omega_j \delta \eta_{ij} + \omega_j \sqrt{1 - \delta_j^2} \epsilon_{ij}\right\} \\ &= \xi_j + \omega_j \delta \eta_{ij} + \omega_j \sqrt{1 - \delta_j^2} E\left\{\epsilon_{ij}\right\} \\ &= \xi_j + \omega_j \delta_j \eta_{ij} \ \text{ since } E\left\{\epsilon_{ij}\right\} = 0 \end{aligned} \tag{99}$$

and variance

$$\begin{aligned} E(z_{ij}) =& Var\left\{\xi_j + \omega_j \delta \eta_{ij} + \omega_j \sqrt{1 - \delta_j^2} \epsilon_{ij}\right\} \\ &= Var\left\{\xi_j + \omega_j \delta \eta_{ij}\right\} + \omega_j^2 \left(1 - \delta_j^2\right) Var\left\{\epsilon_{ij}\right\} \\ &= 0 + \omega_j^2 \left(1 - \delta_j^2\right) Var\left\{\epsilon_{ij}\right\} \\ &= \omega_j^2 \left(1 - \delta_j^2\right) \ \text{ since } Var\left\{\epsilon_{ij}\right\} = 1 \end{aligned} \tag{100}$$

where $\delta_j = \alpha_j / \sqrt{1 + \alpha_j^2}$. Having taken these results into consideration, a hierarchical formulation for $z_{ij} \backsim \mathrm{SN}(\xi_j, \omega_j^2, \alpha_j)$ in Equation 85 is given as a new model in what follows:

$$\begin{aligned} &\bigstar z_{ij}|\eta_{ij}, \omega_j^{-2}, \xi_j, \delta_j \backsim N\left(\xi_j + \omega_j \delta_j \eta_{ij}, \omega_j^2\left(1 - \delta_j^2\right)\right) \\ &=> p(\boldsymbol{z}|\eta_{ij}, \omega_j^{-2}, \xi_j, \delta_j) \propto \prod_{j=1}^{J} \prod_{i=1}^{n_j} \frac{1}{\omega_j \sqrt{1 - \delta_j^2}} e^{-\frac{1}{2\omega_j\left(1-\delta_j^2\right)}(z_{ij} - \xi_j - \omega_j \delta_j \eta_{ij})^2} \end{aligned} \tag{101}$$

$$\begin{aligned} &\bigstar \eta_{ij} \backsim TN_{[0,\infty)}(0, 1) \\ &=> p(\eta_{ij}) \propto e^{-\frac{1}{2}\eta_{ij}^2} \mathbb{I}_{\eta_{ij} > 0} \quad i = 1, 2, \cdots, n_j; j = 1, 2, \cdots, J \end{aligned} \tag{102}$$

Hence, sampling (ξ_j, ω_j^2) from Equation 85 becomes analogous with taking the sample of these parameters from the following complete posterior conditional distribution:-

$$\begin{aligned} p(\eta_{ij}, \omega_j^{-2}, \xi_j, \delta_{ij}|\boldsymbol{z}) \propto& \left\{\left[\prod_{j=1}^{J} \prod_{i=1}^{n_j} \frac{1}{\omega_j \sqrt{1 - \delta_{ij}^2}} e^{-\frac{1}{2\omega_j \sqrt{1-\delta_{ij}^2}}(z_{ij} - \xi_j - \omega_j \delta \eta_{ij})^2}\right]\right. \\ &\left. \times e^{-\frac{1}{2}\eta_{ij}^2} \mathbb{I}_{\eta_{ij} > 0} \times \frac{1}{\sqrt{\kappa \omega_j^2}} e^{-\frac{1}{2\kappa\omega_j^2}(\xi_j - \xi_o)^2} \times (\omega_j^{-2})^{c-1} e^{-d/\omega_j^{-2}}\right\} \end{aligned} \tag{103}$$

Clearly, Equation 103 tells us the full conditional distributions for Gibbs sampling are all in standard form and straightforward to derive. Having simplified, the resulting conditional distributions the parameters involved are:-

$$\blacktriangle\ \eta_{ij}|\omega_{ij}^{-2}, \xi_j, \delta_{ij}, \boldsymbol{z} \backsim \text{TN}\left(\frac{\delta(z_{ij}-\xi_j)}{\omega_j^2}, 1-\delta_{ij}^2\right),\ i=1,2,\cdots,n_j; j=1,2,\cdots,J \tag{104}$$

$$\blacktriangle\quad \xi_j|\eta_{ij}, \omega_j^{-2}, \delta_{ij}, \boldsymbol{z} \backsim \text{N}\left(\widehat{\mu}, \widehat{\kappa}\omega_j^2\right),\ j=1,2,\cdots,J \tag{105}$$

where,

$$\widehat{\mu} = \frac{\kappa\sum_{i=1}^{n_j}(z_{ij}-\delta_{ij}\eta_{ij}) + \xi_o(1-\delta_{ij})^2}{n\kappa + (1-\delta_{ij}^2)} \quad \text{and} \quad \widehat{\kappa} = \frac{\kappa(1-\delta_{ij}^2)}{n\kappa+(1-\delta_{ij}^2)}$$

$$\blacktriangle\quad \omega_j^{-2}|\eta_{ij}, \xi_j, \delta_{ij}, \boldsymbol{z} \backsim \text{InvGamma}\left(c+\frac{1}{2}(n+1), d+\widehat{d}\right), j=1,2,\cdots,J;\ c,d>0 \tag{106}$$

$$\begin{aligned}\widehat{d} &= \frac{1}{2\omega_j^2(1-\delta_{ij}^2)}\sum_{i=1}^{n_j}(z_{ij}-\xi_j-\omega_j\delta_{ij}\eta_{ij})^2 + \frac{1}{2\kappa\omega_j^2}(\xi_j-\xi_o)^2 \\ &= \frac{1}{2(1-\delta_{ij}^2)}\left\{\delta_{ij}^2\sum_{i=1}^{n_j}\eta_{ij}^2 - 2\delta_{ij}\sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j) + \sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j)^2 + \frac{(1-\delta_{ij}^2)}{\kappa}(\xi_j-\xi_o)^2\right\}\end{aligned}$$

(***b***).**Sampling** α::

The trick utilized by (Canale and Scarpa, 2013) is implemented to draw samples for the skewed parameter, α. The notion behind is that for simplicity, we assume the distribution of latent variables z_{ij} associated with the i^{th} woman in j^{th} region, as deemed in Equation 85, is standard skew normal distribution. Hence, the scale and the location parameters, ξ_j and ω_j respectively in Equation 88 are vanished because $\xi_j = 0, \forall j = 1, \cdots, J$ and $\omega_j = 1, j =$

$1, \cdots, J$. Eventually, the expression in Equation 88, then, reduces to :-

$$
\begin{aligned}
p(\alpha_j|\boldsymbol{z}) &\propto \begin{cases} \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j \frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \times p_1(\alpha_j), & \text{where } p_1(\alpha_j) \text{ is} \\ & \text{a Normal prior} \\ \left[\prod_{i=1}^{n} \frac{1}{\omega}\phi\left(\frac{z_{ij}-\xi_j}{\omega_j}\right)\Phi\left(\alpha_j \frac{z_{ij}-\xi_j}{\omega_j}\right)\right] \times p_2(\alpha_j), & \text{where } p_2(\alpha_j) \text{ is} \\ & \text{a S-Normal prior} \end{cases} \\
&\propto \prod_{i=1}^{n_j} \phi(z_{ij})\,\Phi(\alpha_j z_{ij}) \times p_k(\alpha_j), \quad k = 1, 2 \qquad (107)
\end{aligned}
$$

where, $\alpha_j \backsim \mathrm{N}\left(\alpha_j; \alpha_o, \psi_o^2\right)$ or $\alpha_j \backsim \mathrm{SN}\left(\alpha_j; \alpha_o, \psi_o^2, \lambda_o\right)$, which is equivalent to ,

$$p_1(\alpha_j) = \frac{1}{\psi_o}\phi\left(\frac{\alpha_j - \alpha_o}{\psi_o}\right) \text{ and } p_2(\alpha_j) = \frac{2}{\psi_o}\phi\left(\frac{\alpha_j - \alpha_o}{\psi_o}\right)\Phi\left(\lambda_o \frac{\alpha_j - \alpha_o}{\psi_o}\right)$$

, which are two informative prior distributions of α_j(Canale and Scarpa, 2013).

The full conditional posterior distribution, $\alpha_j|\xi_j, \omega_j^2, \boldsymbol{z}, \boldsymbol{w}$,under the two prior cases is given as follows:

When using Normal Prior, $p_1(\alpha_j) = \frac{1}{\psi_o}\phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)$

$$\begin{aligned}
p\left(\alpha_j|\boldsymbol{z}\right) &\propto \prod_{i=1}^{n}\phi\left(z_{ij}\right)\Phi\left(\alpha_j z_{ij}\right)\frac{1}{\psi_o}\phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\\
&\propto \prod_{i=1}^{n}\Phi\left(\alpha_j z_{ij}\right)\phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\\
&\propto \Phi_n\left(\alpha_j z;\mathbb{I}_n\right)\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)\\
&\propto \Phi_n\left(\alpha_o z-\alpha_o z+\alpha_j z;\mathbb{I}_n\right)\phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\\
&\propto \Phi_n\left(\alpha_o z+z\left(\alpha_j-\alpha_o\right);\mathbb{I}_n\right)\phi\left(\frac{\alpha-\alpha_o}{\psi_o}\right)\\
&\propto \phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\Phi_n\left(\alpha_o z+z\left(\alpha_j-\alpha_o\right);\mathbb{I}_n\right)\\
&\propto \phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\Phi_n\left(\frac{\Delta_1\alpha_o}{\psi_o}+\frac{\Delta_1}{\psi_o}\left(\alpha_j-\alpha_o\right);\mathbb{I}_n\right)\text{, in which case, } z=\frac{\Delta_1}{\psi_o}.\\
&\propto \phi\left(\frac{\alpha_j-\alpha_o}{\psi_o}\right)\frac{\Phi_n\left(\frac{\Delta_1\alpha_o}{\psi_o}+\frac{\Delta_1}{\psi_o}\left(\alpha_j-\alpha_o\right);\mathbb{I}_n\right)}{\Phi_n\left(\frac{\Delta_1\alpha_o}{\psi_o};\mathbb{I}_n\right)}\text{, the denominator is a constant w.r.t. } \alpha_j
\end{aligned}$$

This in general is of the form ,

$\mathbf{y} \backsim \mathrm{SUN}_{d,m}\left(\boldsymbol{\mu},\boldsymbol{\gamma},\overline{\boldsymbol{\omega}},\boldsymbol{\Omega}^*\right)$

having the pdf:

$$f\left(\mathbf{y}\right) = \phi_d\left(\boldsymbol{y}-\boldsymbol{\mu};\boldsymbol{\Omega}\right)\frac{\Phi_m\left(\boldsymbol{\gamma}+\boldsymbol{\Delta^T}\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega^{-1}}\left[\boldsymbol{y}-\boldsymbol{\mu}\right];\boldsymbol{\Gamma}-\boldsymbol{\Delta^T}\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\right)}{\Phi_m\left(\boldsymbol{\gamma};\boldsymbol{\Gamma}\right)}$$

where, $\overline{\boldsymbol{\omega}} = \boldsymbol{\omega}\mathbf{1_d}, \boldsymbol{\Omega} = \boldsymbol{\omega}\overline{\Omega}\boldsymbol{\omega}$, a covariance matrix with a $d \times d$ diagonal matrix $\boldsymbol{\omega}$, and $\boldsymbol{\Omega}^* = \begin{pmatrix}\boldsymbol{\Gamma} & \boldsymbol{\Delta}^T\\ \boldsymbol{\Delta} & \overline{\Omega}\end{pmatrix}$, a correlation matrix.Particularly, in our case,we have:

$\mu = \alpha_o,\ \gamma = \dfrac{\Delta_o\alpha_o}{\psi_o},\ \Delta^T = \Delta_1,\ \overline{\Omega} = 1,\ \omega = \psi_o,\ \Omega = \omega\overline{\Omega}\omega = \psi_o^2$

$\Omega^* = \begin{bmatrix}\Gamma & \Delta\\ \Delta^T & \overline{\Omega}\end{bmatrix} = \begin{bmatrix}\Gamma_1 & \Delta_1^T\\ \Delta_1 & 1\end{bmatrix}$ and $\mathbb{I}_n = \Gamma - \Delta^T\overline{\Omega}\Delta = \Gamma_1 - \Delta_1^T\Delta_1$

Thus,

$$\begin{aligned}
&\alpha_j|z \backsim \mathrm{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,\Omega^*\right),\quad \text{or}\\
&\alpha_j|z \backsim \mathrm{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,1,\Delta_1,\Gamma_1\right) \qquad (108)
\end{aligned}$$

$(\boldsymbol{v})$.**Sampling the hyperparameter** a**:**: As for the hyperparameter a, it is apparent from

(3.1.4) that this parameter's full conditional distribution is such that

$$
\begin{aligned}
p\left(a|\xi_j,\omega_{ij}^{-2},\alpha_j,b,\boldsymbol{z},\boldsymbol{w}\right) &\propto \frac{b^a}{\Gamma(a)}R_j^{a-1}\times \text{Gamma}(m_o,n_o)\\
&\propto \frac{b^a}{\Gamma(a)}R_j^{a-1}\times a^{m_o-1}e^{-(n_o)\cdot a}\\
&\propto \frac{b^a}{\Gamma(a)}R_j^{a}\times a^{m_o-1}e^{-(n_o)\cdot a}\\
&\propto \frac{1}{\Gamma(a)}a^{m_o-1}\cdot\left[bR_je^{-n_o}\right]^a,\ a,b,m_o,n_o>0\\
&\propto a^{m_o-1}\cdot\left[bR_je^{-n_o}\right]^a\left[\Gamma(a)\right]^{-1}\mathbb{I}\left(a\in(0,\infty)\right)
\end{aligned}
\tag{109}
$$

Generating a random draw from this distribution is a slightly more complicated task, but it may be accomplished by making use of the procedure set out in the paragraph below. As seen, the conditional does not have a standard form. Therefore, the Gibbs sampler can not be implemented.

($\boldsymbol{v}$).**Sampling the hyperparameter b:**:

Finally, the full conditional distribution of the other hyperparameter is

$$
\begin{aligned}
p\left(b|\xi_j,\omega_{ij}^{-2},\alpha_j,R_j,a,\boldsymbol{z},\boldsymbol{w}\right) &\propto b^ae^{-bR_j}\times b^{r_o-1}e^{-(s_o)\cdot b}\\
&\propto b^{a+r_o-1}e^{-(R_j+s_o)b},\ a,b,r_o,s_o>0
\end{aligned}
\tag{110}
$$

Therefore, the gamma hyperprior on b allows for direct sampling from the conjugate conditional posterior

$$
b|\xi_j,\omega_j^2,\alpha_j,R_j,a,\boldsymbol{z},\boldsymbol{w}\backsim\text{Gamma}\ \left(a+r_o,R_j+s_o\right) \tag{111}
$$

In this thesis, a Metropolis-Hastings within Gibbs sampling algorithm is used. In particular, we use Metropolis-Hastings step to generate samples from the full conditional density of the hyperparameter, a, in Equation 110 owning to the fact that this conditional posterior is not of any known closed forms (does't appear to be any of the standard distributions, such as normal, gamma, etc), and that Gibbs sampling is not straightforward under this condition. On the other hand, we develop a Gibbs sampling algorithm to sample from the conditional posterior distribution of all other parameters of interest, $\xi_j,\omega_j^2,\alpha_j,R_j$, and b, upon acceptance of a.

In what follows, we shall discuss on how to approximate these conditional posteriors using Gibbs Sampling and Metropolis-Hastings Algorithm, using a proposal *standard Gaussian distribution.*

4.5 Parameter Estimation

4.5.1 Bayesian Computation: Markov Chain Monte Carlo Implementation

In complicated Bayesian models, it is often not easy to obtain the posterior distribution analysitically. This analytic bottleneck has been eliminated by the emergence of Markove

Chain Monte Carlo (MCMC). We use a combination of the Accept-Reject and MCMC (the Gibbs sampler and single-component Metropolis-Hastings algorithm) Sampling method to estimate the fertility rate parameters for each region. The former, the Accept-Reject Sampling (aka: rejection sampling) method, has been conducted to sample from the truncated Skew-Normal distribution. In this case, upper bound (global maximum) is determined with the aid of the `R`-optimization command, `nlminb`. Where as, all the other parameters are sampled using MCMC (the Gibbs sampler and Metropolis-Hastings algorithm). This requires the full conditional distributions for each parameter, which are summarized as follows: Having used the valuable approaches of the paper by Canale and Scarpa (2013) and `R` simulation codes obtained from as a clue, we developed an `R` program/script for simulation a truncated Skew Normal Model and unified skew normal distribution. We use a Rejection sampling to draw samples from Truncated skew normal given above. In this case, the simulation is carried out using uniform distribution as proposal density, and upper bound computed by `nlminb`, R optimization function.

We set, $z_{ij}^{(t+1)}$ equal to a random draw using accept-reject algorithm from a $\mathtt{TSN}(\xi_j^{(t)}, \omega_j^{2(t)}, \alpha_j^{(t)})$ distribution based on $\mathtt{U}(j-1/2, j+1/2)$. Then, for $(t+1)^{st}$ iteration of MCMC, for $j = 1, \cdots, 11$, we set $R_j^{(t+1)}$ equal to a random draw from $\mathtt{Gamma}(n + a^{(t)}, b^{(t)})$ distribution given in *(ii)* above, with the conditioning arguments , $\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b$, respectively. Following this, we set we set $\eta_{ij}^{(t+1)}$ equal to a random draw from TN $\left(\frac{\delta_j^{(t)}(z_{ij}^{(t)} - \xi_j^{(t)})}{\omega_j^{2((t))}}, 1 - \delta_j^{2(t)}\right)$, where, $\delta_j^{(t)} = \alpha_j^{(t)} / \sqrt{1 - \alpha_j^{(t)}}$ distribution given in *(ii)* above, with the conditioning arguments , $\boldsymbol{z}, \xi_j, \omega_j^2, \alpha_j, \boldsymbol{w}, a, b$, respectively. Similarly, to sample $\xi_j^{(t+1)}$, we set we set $\xi_j^{(t+1)}$ equal to a random draw from N $\left(\widehat{\mu}, \widehat{\kappa}\omega_j^{2(t)}\right)$ distribution, where

$$\widehat{\mu} = \frac{\kappa \sum_{i=1}^{n_j} (z_{ij}^{(t+1)} - \delta_j^{(t)} \eta_{ij}^{(t+1)}) + \xi_o (1 - \delta_{ij})^2}{n\kappa + (1 - \delta_j^{2(t)})} \quad \text{and} \quad \widehat{\kappa} = \frac{\kappa (1 - \delta_j^{2(t)})}{n\kappa + (1 - \delta_j^{2(t)})}$$

We also set $\omega_j^{-2(t+1)}$ equal to a random sample from InvGamma $\left(c + \frac{1}{2}(n+1), d + \widehat{d}\right)$ distribution given in *(v)* above, where,

$$\widehat{d} = \frac{1}{2(1 - \delta_j^{2(t)})} \left\{ \left[\delta_j^{2(t)} \sum_{i=1}^{n_j} \eta_{ij}^{2(t+1)} - 2\delta_j^{(t)} \sum_{i=1}^{n_j} \eta_j^{(t+1)} (z_{ij}^{(t+1)} - \xi_j^{(t+1)}) \right] \right.$$
$$\left. + \sum_{i=1}^{n_j} \eta_{ij}^{(t+1)} (z_{ij}^{(t+1)} - \xi_j^{(t+1)})^2 + \frac{(1 - \delta_j^2(t))}{\kappa} (\xi_j^{(t+1)} - \xi_o)^2 \right\}$$

To sample $a^{(t+1)}$, we set a candidate point a^* equal to a random draw from a proposal distribution $\mathtt{N}(0,1)\mathbf{I}(\mathbf{x} > \mathbf{0})$. As a general heuristic, we choose the standard deviation of the proposal distribution so that the candidate acceptance probability is between 0.25

Summary of full conditionals for Posterior simulation

$$(\boldsymbol{i}).\quad z_{ij}|R_j,\xi_j,\omega_j^2,\alpha_j,\boldsymbol{w},a,b \backsim \begin{cases} C_1\mathrm{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right) & \text{if } z_{ij}>0 \ \&\ \mathrm{w}_{ij}=1 \\ C_2\mathrm{SN}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right) & \text{if } z_{ij}\leq 0 \ \&\ \mathrm{w}_{ij}=0 \end{cases}$$

where, C_1 and C_2 are normalization constants given by

$$C_1=\frac{1}{cdf_{\mathrm{SN}}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)} \quad \text{and} \quad C_2=\frac{1}{1-cdf_{\mathrm{SN}}\left(z_{ij};\xi_j,\omega_j^2,\alpha_j\right)}$$

$$(\boldsymbol{ii}).\quad R_j|\boldsymbol{z},\xi_j,\omega_j^2,\alpha_j,\boldsymbol{w},a,b \backsim \text{Gamma }(n+a,b) \quad \text{for} \quad j=1,\cdots,J=11$$

$$(\boldsymbol{iii}).\quad \eta_{ij}|\omega_{ij}^{-2},\xi_j,\delta_{ij},\boldsymbol{z} \backsim \mathrm{TN}\left(\frac{\delta(z_{ij}-\xi_j)}{\omega_j^2},1-\delta_{ij}^2\right),\ i=1,2,\cdots,n_j; j=1,2,\cdots,J$$

$$(\boldsymbol{iv}).\quad \xi_j|\eta_{ij},\omega_j^{-2},\delta_{ij},\boldsymbol{z} \backsim \mathrm{N}\left(\widehat{\mu},\widehat{\kappa}\omega_j^2\right),\ j=1,2,\cdots,J$$

where,

$$\widehat{\mu}=\frac{\kappa\sum_{i=1}^{n_j}(z_{ij}-\delta_{ij}\eta_{ij})+\xi_o(1-\delta_{ij})^2}{n\kappa+(1-\delta_{ij}^2)} \quad \text{and} \quad \widehat{\kappa}=\frac{\kappa(1-\delta_{ij}^2)}{n\kappa+(1-\delta_{ij}^2)}$$

$$(\boldsymbol{v}).\quad \omega_j^{-2}|\eta_{ij},\xi_j,\delta_{ij},\boldsymbol{z} \backsim \text{InvGamma}\left(c+\frac{1}{2}(n+1),d+\widehat{d}\right),j=1,2,\cdots,J$$

$$\widehat{d}=\frac{1}{2(1-\delta_{ij}^2)}\left\{\delta_{ij}^2\sum_{i=1}^{n_j}\eta_{ij}^2-2\delta_{ij}\sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j)+\sum_{i=1}^{n_j}\eta_{ij}(z_{ij}-\xi_j)^2+\frac{(1-\delta_{ij}^2)}{\kappa}(\xi_j-\xi_o)^2\right\}$$

$$(\boldsymbol{vi}).\quad \alpha_j|z \backsim \begin{cases} \mathrm{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,\Omega^*\right) \\ \\ \mathrm{SUN}_{1,n}\left(\alpha_o,\Delta_1\alpha_o/\psi_o,\psi_o,1,\Delta_1,\Gamma_1\right) \end{cases}$$

$$(\boldsymbol{vii}).\quad p\left(a|\xi_j,\omega_j^2,\alpha_j,b,\boldsymbol{z},\boldsymbol{w}\right) \propto a^{m_o-1}\cdot\left[bR_je^{-n_o}\right]^a\left[\Gamma(a)\right]^{-1}\mathbb{I}\left(a\in(0,\infty)\right)$$

$$(\boldsymbol{viii}).\quad b|\xi_j,\omega_j^2,\alpha_j,R_j,a,\boldsymbol{z},\boldsymbol{w} \backsim \text{Gamma }(a+r_o,R_j+s_o),\ b\geq 0 \tag{112}$$

and 0.45 (Gelman et al., 1995). For this proposal distribution, the candidate acceptance probability is

$$A=\min\{1,R\}$$

in which case,

$$R=\frac{(a^*)^{m_o-1}\left[b^{(t)}R_j^{(t+1)}exp(-n_o)\right]^{(a^*)}\left[\Gamma(a^*)\right]^{-1}}{(a)^{m_o-1}\left[b^{(t)}R_j^{(t+1)}exp(-n_o)\right]^{(a)}\left[\Gamma(a)\right]^{-1}} \tag{113}$$

Finally, we set $b^{(t+1)}$ equal to a random draw from a Gamma$(a^{(t+1)} + r_o, R_j^{(t+1)} + s_o)$ distribution with the conditioning arguments , $\xi_j, \omega_j^2, \alpha_j, R_j, a, \boldsymbol{z}, \boldsymbol{w}$, respectively.

4.5.2 Maximum Likelihood Estimation(MLE)

There are two different approaches for parameter estimation, which is either Bayesian or maximum likelihood. Bayesian parameter estimation requires the assignment of a prior distribution for the unknown parameter θ^* . The objective is then to calculate the posterior distribution of θ^* given the observed data. When a point estimate of $\widehat{\theta}^*$ is required, some feature of this posterior distribution can be provided. The common Bayesian estimators are the posterior mean, posterior median, and the posterior mode, or the maximum a posteriori probability (MAP) estimate. There are several Monte Carlo based methods for Bayesian parameter estimation when exact calculation of the posterior distribution is not available. Alternatively, the maximum likelihood approach regards the likelihood of the observed data, which is a function of θ^*, to contain all relevant information for estimating $\widehat{\theta}^*$. The point estimate of $\widehat{\theta}^*$ is the maximising argument of the likelihood. When maximum likelihood estimation (MLE) cannot be done analytically, iterative searchbased algorithms such as expectation-maximisation (EM) and gradient ascent guarantee maximising the likelihood locally given certain regularity conditions on densities of the random variables involved.

Whether one should in principle use the Bayesian or maximum likelihood approach for estimating $\widehat{\theta}^*$ is a fundamental debate which we will not go into. There are indeed cases when these two approaches do produce dramatically different suggestions on what $\widehat{\theta}^*$ might be, especially when the observed data is of small size and a highly informative prior for Bayesian estimation is used. However, as data size tends to infinity, the likelihood of the data sweeps away the effect of the prior in the posterior distribution and the difference between the estimates of the two approaches vanishes (say when the MAP estimate is used for Bayesian estimation), provided that the prior is well-behaved (*i.e.,* it does not assign zero density to any "feasible" parameter value). Note that the maximum likelihood (ML) estimation methods used commonly in multilevel or hierarchical analysis are asymptotic, which translates to the assumption that the sample size must be sufficiently large. This raises questions about the acceptable lower limit to the sample size, and the accuracy of the estimates and the associated standard errors with relatively small sample sizes.

Inference is a central problem in many studies of fertility. Like Bayesian inference, there are various parameter estimation techniques , which are widely used in frequentist framework as well. In the paper, the maximum likelihood approach has been adopted in order to estimate model parameters and compare the result obtained with the Bayesian settings as it is done with the previous chapter, Chapter 3. It is a commonly used procedure (Harter and Moore, 1965), and (Cohen, 1965) with many desirable properties[18]. The idea is: let $x_1, x_2, x_3, \cdots, x_n$, be a random variables of size n drawn from a probability density function $f(x;\theta)$, where θ is an unknown parameter. The likelihood function of this random sample is the joint density of the n random variables and is a function of the unknown parameter. Thus,

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{114}$$

is the Likelihood function. The Maximum Likelihood Estimator (MLE) of θ, say $\widehat{\theta}$, is the value of θ, that maximizes L or, equivalently, the logarithm of L, called log-likelihood. As a practical matter, when computing the maximum likelihood estimate it is often easier to work with the log-likelihood, $l(\theta) := log\ f(x;\theta)$. Because the logarithm is monotonic, it does not affect the argmax:

$$\widehat{\theta} = \arg\max_{\theta} l(\theta)$$

Often, but not always, the MLE of $\widehat{\theta}$ is a solution of

$$\frac{\partial LogL}{\partial\theta} = 0 \tag{115}$$

As described categorically in Equation 36, the likelihood function of the model under study is given as

$$\begin{aligned}
L\left(R,\xi,\omega^2,\alpha|\mathbf{d}\right) &= \prod_{x=0}^{n}\left\{\binom{n_x}{d_x}\left[R\cdot\ \ g(x;\xi,\omega^2,\alpha)\right]^{d_x}\left[1-R\cdot\ \ g(x;\xi,\omega^2,\alpha)\right]^{n_x-d_x}\right\} \\
&\propto \prod_{x=0}^{n}\left\{\left[R\cdot\frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]^{d_x}\right. \\
&\left.\left[1-R\cdot\frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]\right\}^{n_x-d_x}
\end{aligned} \tag{116}$$

Thus, the corresponding log-likelihood function is

$$\begin{aligned}
l\left(R,\xi,\omega^2,\alpha|\mathbf{d}\right) &= log\left\{\prod_{x=0}^{n}\binom{n_x}{d_x}\left[R\cdot\ \ g(x;\xi,\omega^2,\alpha)\right]^{d_x}\right. \\
&\left.\left[1-R\cdot\ \ g(x;\xi,\omega^2,\alpha)\right]^{n_x-d_x}\right\} \\
&= log\left\{\prod_{x=0}^{n}\left[R\cdot\frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]^{d_x}\right. \\
&\left.\left[1-R\cdot\frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)\right]\right\}^{n_x-d_x}
\end{aligned} \tag{117}$$

[18] By far the best justification for the use of the maximum likelihood method of estimation is its ability to satisfy the following crucial properties: *sufficiency,invariance, consistency, efficiency,and asymptotic normality*

which is simply the natural logarithm of (55).The maximum likelihood estimate of $\theta^* = (\xi, \omega, \alpha, R)$, that is, $\widehat{\theta}^* = (\widehat{\xi}^*, \widehat{\omega}^*, \widehat{\alpha}^*, \widehat{R}^*)$,is a value of θ^* that maximizes the likelihood Equation 116, or equivalently the log likelihood Equation 117. However, such estimates are extremely complicated nonlinear functions of the observed data. As a result, closed form expressions for the MLEs doesn't generally exist for the models that we are working with. and therefore a numerical or analytical approximation is required. However, analytical approximation approaches often fail to give entirely satisfactory results. Therefore,instead of computing analytically, the maximation of the log-likelihood function is made using numerical approach. In this regard, we utilized the nlminb[19](*non-linear minimization subject to box constraints*)function in `R software` to compute the desired MLEs for the model parameters[20]. Results for two regions (Addis Ababa and Somali region) are shown as Table 4.2 and Table 4.3 where as the results of the remaining regions are given in Appendix C.

4.6 Empirical Results

4.6.1 Application to ASFR data

We implemented a Bayesian hierarchical estimation procedure for our ASFR data, as follows. First, recall that ASFR model has four parameters: `R` is the **TFR parameter** of the model, which reflects the fertility circumstance of childbearing age mothers in each of the regions, and assumes values in the interval $(0, 16)$. As paraphrased in subsection 3.2, ξ and α are the **location parameter** and the **shape parameter** of the model and they assume values in the interval $(-\infty, \infty) = \mathbb{R}$ where as ω is the **scale parameter** of the model and it takes values in the interval $(0, \infty)$. ξ indicates the function of the average age of mother at childbearing in each region, and ω stands for a value proportional to the variance or variability in fertility rate of mothers in each region during childbearing age. Apparently, α is interpreted as how the shape of ASFR resembles for mothers in the study across each of the 11 regions, as described by Mazzuco and Scarpa (2015).

[19] `nlminb` finds the parameter estimates that minimize a function. Thus, in order to perform maximum likelihood estimation, the user must provide nlminb with the negative of the log likelihood function.

[20] Another standard algorithm to find ML point estimates or posterior modes is the expectation maximization (EM) algorithm. Mathematically finding ML estimates or posterior modes is done by taking the partial derivatives of the likelihood or posterior distribution and solve a set of equations. Usually the set of equations cannot be solved directly. Latent variables are added to the model which formulate the model more easily, notably they make it easy to estimate the parameters given the latent variables. Latent variables and parameters cannot be estimated simultaneously. The EM algorithm estimates the latent variables and parameters alternatingly until convergence. It can be proven that each iteration of the EM algorithm increases the log likelihood or log posterior density. The latent variable representation of the binomial model makes EM widely applicable.

So far, three major approaches have been discussed that may affect the estimates: *Complete Pooling, Complete Independence, and Hierarchical.* In addition, the estimation methods, Bayesian or not, may be important. Multilevel or Hierarchical modeling mostly uses ML estimation. After our ASFR data sets had been generated, the parameters of ASFR model were estimated using both MLE and our hierarchical Bayesian method for comparison purpose. For the latter method, posterior distributions were approximated by a total of 6,000 MCMC samples, after a burn-in of 1000 samples. To examine how robust our result is, we conducted *sensitivity analysis*, *convergence diagnosis* and *model checking* as stated in what follows.

4.6.2 Sensitivity Analysis: Sensitivity analysis on hyperprior distribution

Posterior inferences, namely point estimation (posterior mean,posterior median, etc), interval estimation (posterior credible interval), and hypothesis testing on the classification probabilities are often highly dependent on the choice or specification of priors, especially when the sample sizes are small.
In presence of model uncertainty, the priors on the parameters need to be specified with care, and a sensitivity analysis should always be performed and discussed for a number of sensible priors (King et al., 2009).
We, therefore, performed a sensitivity analysis to examine the effect of noise on parameter by changing the values for hyper-parameters to investigate whether results in the analysis remained unaffected/unchanged in the presence of different prior information (Nilsson et al., 2011). Vague[21] but proper prior distributions, namely:-

$$\begin{aligned}
&Gamma\left(m_o, n_o\right), \;\; where \;\; m_o = 0.01, \quad m_o = 0.01\\
&Gamma\left(r_o, s_o\right), \;\; where \;\; r_o = 0.01, \quad s_o = 0.01\\
&\mathrm{N}\left(\xi_o, k\omega_j^2\right), \;\; where \;\; \xi_o = 20, \quad k = 0.6\\
&\mathrm{SN}\left(\alpha_o, \psi_o^2, \lambda_o\right), \;\; where \;\; \alpha_o = 1, \quad \psi_o^2 = 4, \quad \lambda_o = 2
\end{aligned}$$

were first specified for the country-level shape parameter,a, the country-level scale parameter, b, region-specific average age parameter, ξ_j and region-specific fertility shape parameter, α_j respectively. Moreover, we use inverse $Gamma(0.01, 0.01)$ prior distributions, as they provide little prior information. As the same time, we use a $\mathrm{N}(20, 0.06)$ prior for ξ_o, as this support covers the factors natural range of variability. Then we made some

[21] Vague Priors: are also called *non-informative priors, automated priors, default priors*, or *priors of ignorance*

modifications to the hyperprior distributions as in what follows,

$$\begin{aligned}
&Gamma\,(m_o, n_o)\,, \quad where \;\; m_o = 0.001, \quad\;\; m_o = 0.001\\
&Gamma\,(r_o, s_o)\,, \quad where \;\; r_o = 0.001, \quad\;\; s_o = 0.001\\
&\mathrm{N}\left(\xi_o, k\omega_j^2\right), \quad where \;\; \xi_o = 3, \quad\;\; k = 0.9\\
&\mathrm{SN}\left(\alpha_o, \psi_o^2, \lambda_o\right), \quad where \;\; \alpha_o = 2, \quad\;\; \psi_o^2 = 5, \quad\;\; \lambda_o = 3
\end{aligned}$$

, recomputed the posterior quantities of interest and checked whether they imposed a practical impact on interpretations or decisions. At last, we adapted the following informative prior distribution

$$\begin{aligned}
&Gamma\,(m_o, n_o)\,, \quad where \;\; m_o = 0.1, \quad\;\; m_o = 0.1\\
&Gamma\,(r_o, s_o)\,, \quad where \;\; r_o = 0.1, \quad\;\; s_o = 0.1\\
&\mathrm{N}\left(\xi_o, k\omega_j^2\right), \quad where \;\; \xi_o = 4, \quad\;\; k = 2\\
&\mathrm{SN}\left(\alpha_o, \psi_o^2, \lambda_o\right), \quad where \;\; \alpha_o = 0.5, \quad\;\; \psi_o^2 = 5, \quad\;\; \lambda_o = -2
\end{aligned}$$

for this purpose. The three distributions gave almost identical results for all the parameters, indicating the results are robust to changes in prior information. All the analyses were performed using the R programming language/environment, (R programming language/environment,R version 3.0.2 Development Core Team, 2005), which is a free downloadable software.

4.6.3 Diagnosing Convergence

Convergence of the MCMC chains was confirmed by visual inspection using trace plot and by computing the `R statistic` developed by Heidelberger and Welch (1983); Geweke (1992); Raftery et al. (1992), which are available in the package Bayesian Output Analysis Program (`BOA` and `coda`) (Smith, 2005) within `R` (Development Core Team, 2005). We used the default values of `BOA` and `coda` to define the length of the burn-in stage, thin the chain check stationarity and define the adequate sample size to achieve the precision required, when sampling from the posterior distribution. Both, the traceplots and convergence diagnostic tests of the parameters show excellent mixing and rapid convergence. The procedure is not sensitive to the initial choices of the parameters. Results are shown in subsubsection 4.6.5.

4.6.4 Goodness of Fit or Model Checking

Bayesian prior-to-posterior analysis conditions on the whole structure (likelihood and prior distribution) of a probability model and can yield very misleading inferences when the model is far from reality. A good Bayesian analysis, therefore, should at least check to see if the fit of the model to the data is so poor that the model should be rejected without other evidence. In the classical setting, this checking is often facilitated by a *goodness-of-fit test*, which quantifies the extremeness of the observed value of a selected measure of discrepancy (e.g., differences between *observations* and *predictions*) by calculating a tail-area probability given that the model under consideration is true. Assessment of any statistical model can be aided by computing appropriate diagnostic measures that characterize the model's fit to the observed data. When a Bayesian analysis is employed, posterior predictive distributions can be used for model assessment (Gelman et al., 1996). The posterior predictive distribution is the distribution of an unknown observation from the same process given the observed data. In a general sense, the goal of posterior predictive assessment is to evaluate the "closeness" of data generated from the fitted model to the actual observed data. In what follows, we discussed how to employ a Bayesian test of model fit using the posterior predictive distribution (Gelman and Meng, 1996).

Posterior predictive checking uses a replicated data set generated by the model in question to compare with the observed data. For y the observed data and θ the vector of all parameters, y^{rep} is defined as the replicated data that could have been observed. The distribution of y^{rep} or the posterior predictive distribution is

$$p\left(y^{rep}|y\right) = \int p\left(y^{rep}|\theta\right) p\left(\theta|y\right) d\theta \tag{118}$$

4.6.5 Results:Hierarchical Bayesian Estimation(HBE) vs Maximum Likelihood Estimation (MLE)

Posterior Findings:- In the same fashion as the previous **Chapter 3**, the Gibbs sampling along with Metropolis-Hastings algorithm in `R software` was applied to the data set of ASFR from 11 regions. In each case 6000 iterations of the algorithm were carried out, of which the first 1000 were discarded as the so-called "*burn-in*" period[22] of the process. Thus, the remaining 5000 simulations can then be regarded as realizations of the marginal distributions of the posterior, and were used to calculate the posterior moments (*mean, mode, median, sd*) and 95% *credible interval* of for model parameters in each region.

As a posterior summary, we, therefore, determined the posterior point and interval estimates. This is to say, as posterior point estimate, we computed the posterior moments such as *means, median* and *standard deviations*; and as posterior interval estimate, we were also able to figure out the the 95% *highest density region* (also termed the 95% *credible interval*) for the ASFR model parameters for each regions. On top of this, we also manage to

[22] "Burn-in": *is a colloquial term that describes the practice of throwing away some iterations at the beginning of an MCMC run. The idea is that you start somewhere, say at x, then you run the Markov chain for n steps, from which you throw away all the data (no output). This is the burn-in period. After the burn-in you run normally, using each iterate in your MCMC calculations.*

incorporate findings from the maximum likelihood methods as it enables us to comprehend better on whether or not our hierarchical model outperforms and is flexible enough for the traditional formulas. Table 3-4, given in what follows, wraps up these aforementioned posterior summaries and the the finding from maximum likelihood method for two regions, namely: *Somali* and *Addis Ababa*. The result for all other regions is also scrutinized in the same manner as those regions. The detail is, however, provided in Appendix C.

Table 4.2: **Hierarchical Bayesian and ML estimates for fertility data set of Somali Region**

Parameters	Bayesian Estimates					ML Estimate		
	Mean	Median	SD	95%[Conf	Int]	MLE	95%[Conf	Int]
ξ	20.530	20.528	0.0734	20.400	20.648	19.460	17.311	21.609
ω^2	8.653	8.657	0.1858	8.371	8.924	8.207	7.301	9.113
α	3.789	3.790	0.0461	3.711	3.883	3.002	2.853	3.561
R	7.099	7.102	0.0518	7.047	7.166	6.962	6.193	7.731

Table 4.2 offers a lot more information: the posterior mean of the shape parameter for Somali Region, α, was more or less 3.79 with 95% credible interval of (3.71, 3.88), revealing strong evidence of skewness (positive skewness). Analogously, we also observe the posterior mean of the location parameter ξ was 20.5 with 95% credible interval of (20.40, 20.65) (the probability that (20.40, 20.65) contains ξ is 95%), which implies the average age, calculated as

$$E[Y] = \xi + \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}} \omega$$

is more or less about 27.20. As the distribution is positively skewed, it is intuitive the modal age will be far less than the average age for those mothers in this region incorporated up on the study. Thus, we can conclude that there is a strong evidence which supports the vast majority (*i.e.*, about the modal value) of women age 15-49 in the study became fertile (deliver their first birth) in early 20 or younger, a proof for early fertility. This could be have various justifications and also be associated with different problems prevailing to date in the region. One deriving factor behind in early fertility is that the region is one of the remotes and the four least developing regions [23] with a mobile nomadic population, which significantly suffers from social and physical infrastructure developments. It is also a region where *Cultural and traditional barriers to effectively utilize modern birth control methods,*

[23] *Least developing regions in Ethiopia*: are regions, such as *Gambella, Afar, Somali, and Benshangul-Gumuz*

low status of women and gender inequality, as well as *poor health service coverage* are quite visible. The other point we observe in our result for this region is that the posterior mean of the scale parameter ω^2 was 8.65 with 95% credible interval of (8.37,8.92) indicating the presence of nearly 5.50 uncertainty (variability) in the ASFR data in the region. The result also reveals the posterior mean total fertility rate in the region was about 7.10 with 95% credible interval of (7.05,7.17), testifying strong evidence the total fertility rate was quite high.

As outlined previous, maximum likelihood estimators of the parameters of our proposed model were also determined for all regions. The rationale behind the need of the estimates obtained from this underlining procedure is to simply compare with estimates of its Bayesian counterparts. As recapped in the last column of Table 4.2, the maximum likelihood estimates of the location, ξ, the scale, ω^2, the shape, α and the fertility, R parameters was respectively 19.46 years, 8.21, 3.00 and 6.96 with their corresponding 95% confidence interval of (17.31 21.61), (7.301 9.11), (2.85 3.56), and (6.19 7.73). Our result showed the 95% confidence interval of those parameters was wider than the 95% credible interval, which suggested the maximum likelihood estimates were less precise compared to its Hierarchical Bayesian analogues. This, in turn, implied the fact that we had a lot of uncertainty about the true parameter values with ML method. On the other hand, the Hierarchical Bayesian estimates seemed more precise. They had narrower credible interval for all model parameters: location, ξ, scale, ω^2, shape, α and fertility, R. The discrepancy between credible and confidence interval was due to sample size ($n = 346$). As the sample size gets large, the approximate 95% CI found from maximum likelihood estimates will have very nearly probability of 0.95 of covering the true value of our model parameters. Thus, the Hierarchical Bayesian method performed better in this small-sample setting (Royle et al., 2009; Brown et al., 2004).

The other issue of concern is convergence. Trace plots of MCMC samples are very useful in assessing convergence. In general, the trace tells whether or not the chain has converged to its stationary distribution-that is, whether the chain needs a longer burn-in period, and is also mixing well. As to this study is pertained, the MCMC sampling appears to be convergent 5000 post burn-in samples were generated from the posterior distribution. Figure 4.7 through Figure 4.10 display the trace plots for parameters the location, ξ, the scale, ω^2, the shape, α and the fertility, R parameters of this region, Somali region. All our figures showed a "perfect" trace plot, with the center of the chain appears to have very small fluctuations. This indicates that the chain was mixing well; the mixing was quite good here. Besides, The MCMC convergence diagnostics using CODA (Cowles and Carlin, 1996) were applied to test the chains of parameters for convergence. The estimates

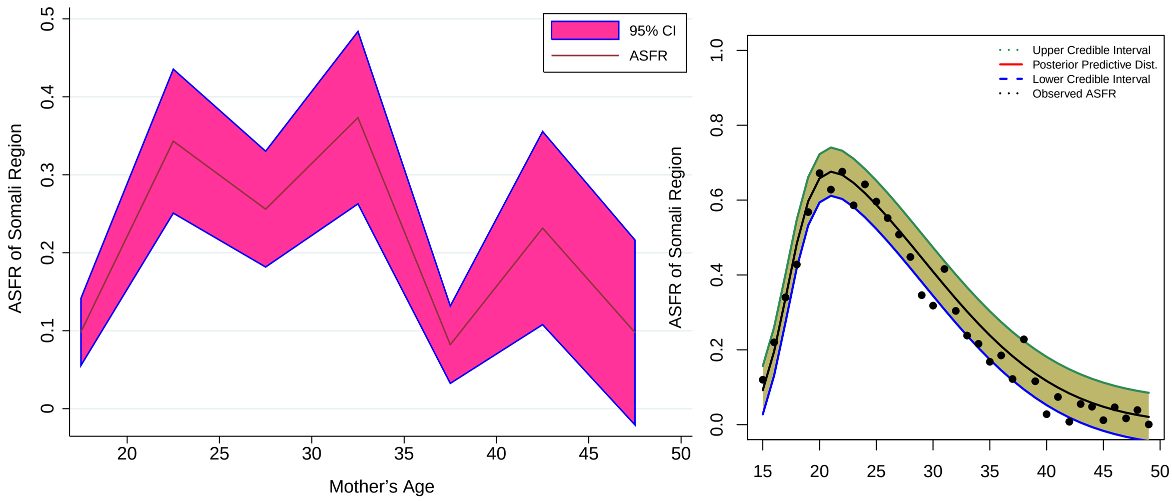


Figure 4.6: ***Left:*** **95% Confidence interval for the MLE fitted ASFR model** ; ***Right:*95% credible interval for the Posterior Predictive distribution** of **Somali Region**

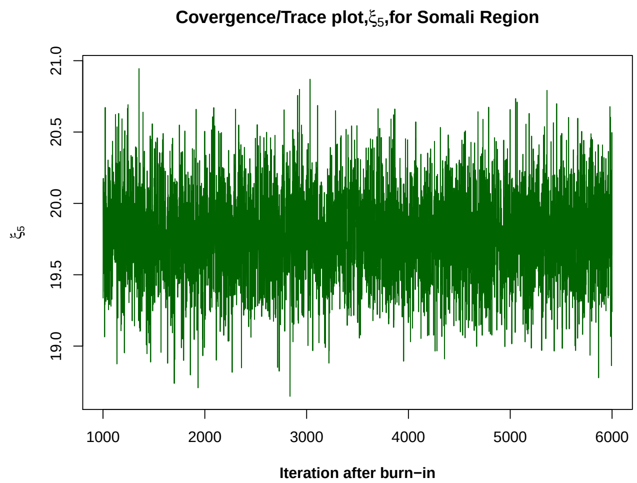


Figure 4.7: MCMC trace plots after burn-in for the fertility parameter of Somali, ξ_{Somali}

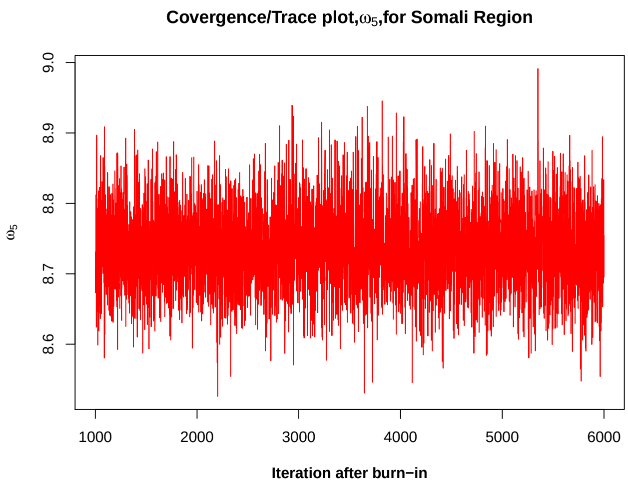


Figure 4.8: MCMC trace plots after burn-in for the fertility parameter of Somali, ω_{Somali}

using MCMC were then summarised for statistical inference in Table 4.3. The z-scores for parameters the location, ξ, the scale, ω^2, the shape, α and the fertility, R were all between -2 and 2 for the Geweke diagnostic test. The stationarity tests for parameters ξ, ω^2, α and R were passed with *p-values* greater than *0.05*, suggesting that the null hypothesis of being stationary was not rejected for each of parameter. The half-width tests of all parameters were passed as their values were less than the product of eps (0.1) with the corresponding sample mean for each parameter. The dependence factors (I) for the R-L diagnostic test were all below 5.0, which suggested that the sample was less correlated. All of these results

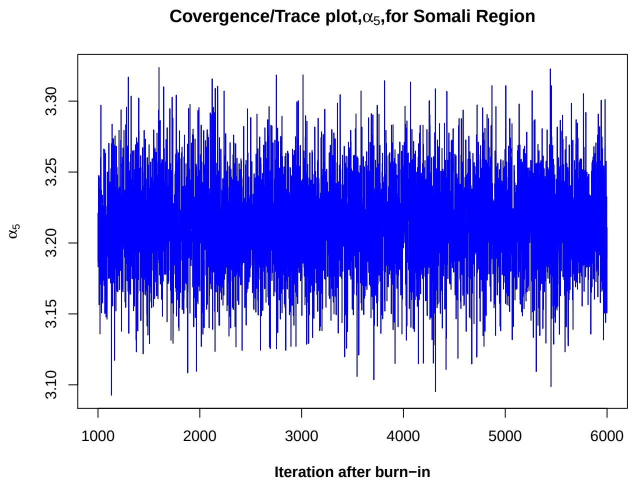


Figure 4.9: MCMC trace plots after burn-in for the fertility parameter of Somali, α_{Somali}

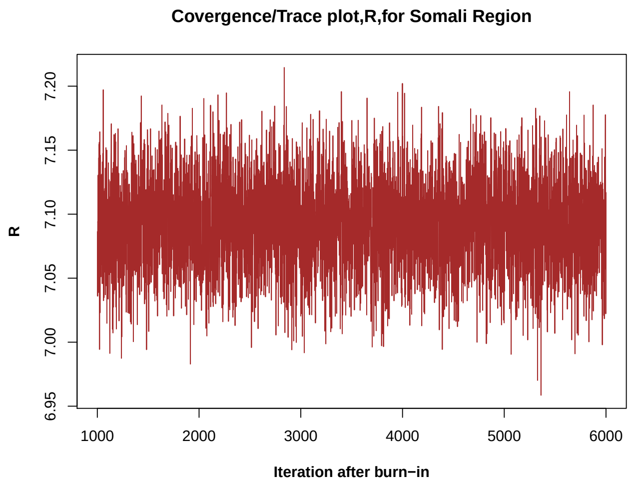


Figure 4.10: MCMC trace plots after burn-in for the fertility parameter of Somali , R_{Somali}

together suggested that the chains of parameters had converged.

Table 4.3: **MCMC convergence diagnostic test for Somali Region fertility model parameters ξ_{Somali},ω^2_{Somali},α_{Somali},R_{Somali} using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.8433	*passed*	0.0021	-0.3624	1.53
ω^2	*passed*	0.7270	*passed*	0.0641	-0.2561	1.62
α	*passed*	0.3692	*passed*	0.0452	-1.243	1.86
R	*passed*	0.1521	*passed*	0.0825	-1.0451	1.17

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

Figure 4.6 visualizes the 95% confidence bound of the fitted MLE of ASFR model (*left*), and 95% credible bound for posterior predictive distribution of the model (*right*). The shaded area between the blue line 2.5% lower bound and 97.5% upper bound represents 95% confidence bound for the maximum likelihood estimated ASFR (*left*), and the latter represents 95% credible bound for Posterior Predictive distribution while the solid line in the center represents the maximum likelihood estimated ASFR (*left*), and posterior predictive mean (*right*). By inspecting the two figures, we could also reaffirm that the confidence interval for MLEs was larger than the credible interval for HBEs (the distance between the two blue lines in the left of the figure was larger than that between the two lines on the right). Our result tells us the Bayesian Credible Interval is shorter or narrower than the confidence interval from Maximum likelihood estimation suggesting more uncertainties with MLEs. In this empirical work , we also observed that Hierarchical procedure allowed for borrowing of information across each of the regions. In similar manner, results

Table 4.4: **Hierarchical Bayesian and ML estimates for fertility data set of Addis Ababa City Administration**

Parameters	Bayesian Estimates					ML Estimate		
	Mean	Median	SD	95%[Conf	Int]	MLE	95%[Conf	Int]
ξ	33.928	33.630	0.0734	33.498	33.748	32.654	29.049	36.262
ω^2	8.343	8.347	0.1862	8.056	8.613	8.654	7.698	9.610
α	-3.289	-3.292	0.0506	-3.390	-3.195	-3.117	-3.461	-2.773
R	1.883	1.880	0.0578	1.822	1.956	1.985	1.766	2.204

for Addis Ababa, as shown in Table 4.3, clearly indicate that the posterior moments and their credible intervals were different from respective the maximum likelihood estimates of the model parameters. For instance, the posterior mean of the location parameter, ξ, was 33.928 with 95% credible interval of (33.498, 33.748) while the corresponding values from maximum likelihood method was 32.654 with 95% credible interval of (29.049,36.262), certain disparity both on the value of the estimate and its interval.These analysis shows that the use of reasonable prior information in our Hirarchical Bayesian Analysis led to a result different from that of the frequentist analysis. The Hirarchical Bayesian estimate was not only different but also more efficient than the maximum likelihood counterpart.

The result also reports the fertility rate of Addis Ababa was estimated to be more or less 10% larger under MLEs than under HBEs, and also its uncertainty was larger in the former procedure. In contrast, both approaches gave about the same estimated mean age of delivery the first child.

Furthermore, the MLEs of ξ, ω, α and R were once again less precise than their Hierarchical Bayesian analogy for Addis Ababa as well. But, unlike to Somali region, the posterior mean of the shape parameter, *i.e.*, α, which was -3.289 with 95% credible interval of (-3.381,-3.183), indicates strong evidence of negative skewness for Addis Ababa. Thus, the majority of mothers age 15-49 became fertile at later ages,*i.e.*, in early 30 and on[24] in the Capital, Addis Ababa. By the same token, both estimates, MLEs and HBEs, show a strong evidence that the fertility rate of most of the women age 15-49 were quite low, specially compared to Somali region. This provides strong evidence the fertility rate in the Capital was declining and even was below replacement level, which is 2.1 chirdren per woman. In a nut shell, result shows most mothers in childbearing age in Addis Ababa gave birth their first birth very late when became older than 30 years and the fertility rate was more or less 2 births(children)per woman during the study period, which is the exact opposite of Somali region. The rise in costs of living & in school-related costs, the increase demand for family planning, violating the traditionally dominant fertility norms and intentions, unprecedented unemployment and under employment crises, rationalizing on the quality of children rather than the number of children they would have could be some of the reasons to result in increased age at first marriage and shortened the exposure to childbearing before menopause in Addis Ababa. Hence, The result seems persuasive.

The MCMC trace plots for the ASFR model parameters are given in Figure 4.11 through

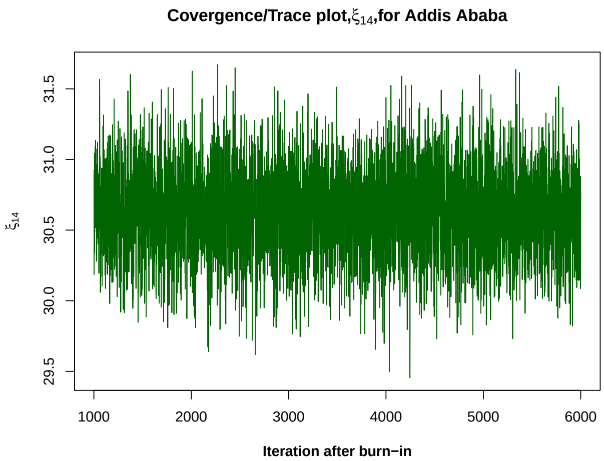


Figure 4.11: MCMC trace plots after burn-in for the location parameter of Addis Ababa , ξ_{AA}

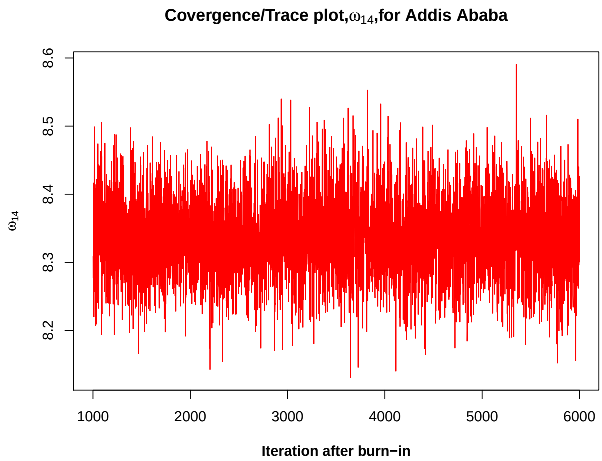


Figure 4.12: MCMC trace plots after burn-in for the scale parameter of Addis Ababa, ω_{AA}

Figure 4.14. To check that the chains had converged to the correct place, the same diagnostic test was carried out using Geweke, Heidelberger & Welch (H-W), and Raftery & Lewis (R-L) tests. The results of the MCMC convergence diagnostics using `CODA` are presented in the following table,Table 4.5. Here again, the over all result indicated the chains for the all parameters converged very well. For instance, The z-score for ξ_{AA} 2 was -1.481 for the

[24] This is the average age calculated using the formula for SN. As the resulting distribution was negatively skewed. Thus, we have:$modalage > medianage > 30$

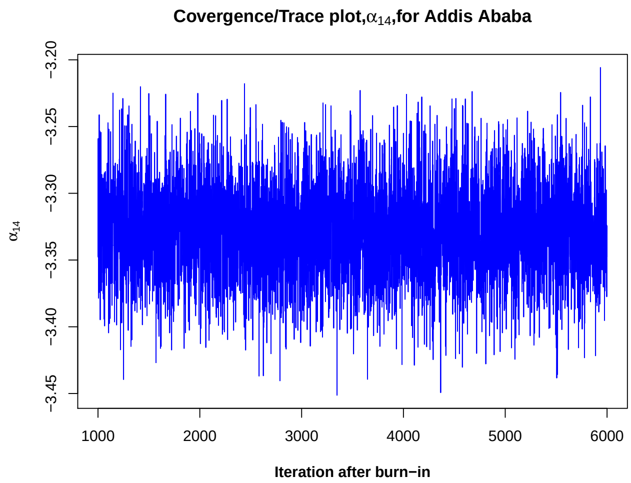


Figure 4.13: MCMC trace plots after burn-in for the shape parameter of Addis Ababa, α_{AA}

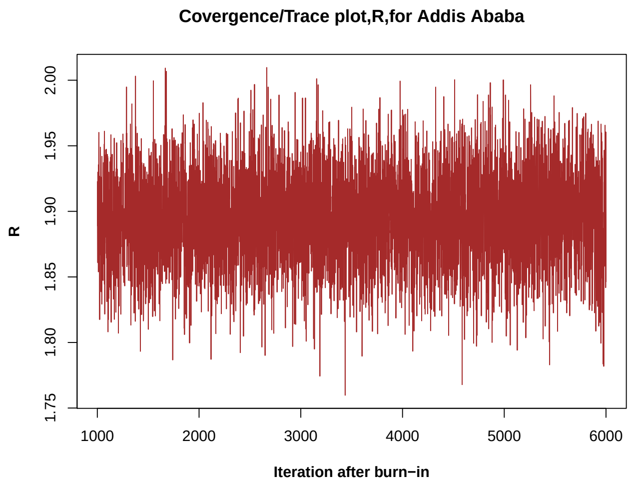


Figure 4.14: MCMC trace plots after burn-in for the fertility parameter of Addis Ababa , R_{AA}

Geweke test. As this value was between -2 and 2, it could be concluded that the chains of parameters had reached convergence at a 5% significance level. The stationarity test forξ_{AA} was passed with a p-value of 0.3567 for the H-W diagnostic test, under the null hypothesis that the MCMC chain was stationary. Furthermore, the half-width test was passed as the ratio between the half-width and the mean was lower than $eps = 0.1$ for the HW test. This also suggested that the chains of parameters had reached convergence. The R-L test showed that the dependence factor (I) forξ_{AA} was also2.63, which is lower than 5.0, indicating that the sample was less correlated confirming the convergence.

Table 4.5: **MCMC convergence diagnostic test for Addis Ababa fertility model parameters $\xi_{\mathbf{AA}}$,$\omega^2_{\mathbf{AA}}$,$\alpha_{\mathbf{AA}}$,$R_{\mathbf{AA}}$ using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.3567	*passed*	0.03450	-1.481	2.63
ω^2	*passed*	0.2710	*passed*	0.0883	0.0029	2.01
α	*passed*	0.4211	*passed*	0.0782	0.1012	1.77
R	*passed*	0.1453	*passed*	0.0628	-0.3748	1.64

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

Figure 4.6 illustrates the agreement between the model predictions and the observations, after updating the input distributions. The posterior predictive distributions of the fitted data along with the observation. The fact that the posterior predictive distributions agree well with the observed data suggests a successful calibration. Our proposed methods pro-

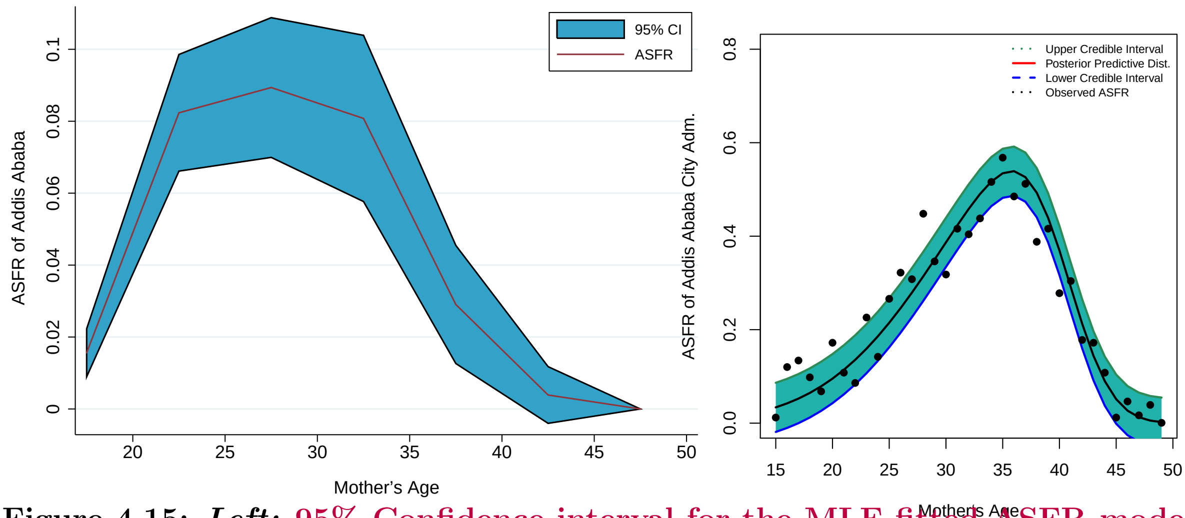


Figure 4.15: ***Left:*** **95% Confidence interval for the MLE fitted ASFR model** ; ***Right:*** **95% credible interval for the Posterior Predictive distribution** of **Addis Ababa**

duce more precise estimates of the model parameters, in particular conferring statistical significance to the regions from which smaller number of childbearing age mothers were sampled as compared to the other procedures, which is mainly due to the ability of the Hierarchical procedure to allow for borrowing of information/strength across each of the other regions.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Fertility rate is one of the most important determinants of overall population growth and demographic transitions in a given country. It has important consequences for economic growth, poverty reduction, and improved health and nutrition outcomes. Ethiopia currently has one of the highest fertility rates in the world, with marked differences among its regions. In 1993, government launched an explicit national population policy for the first time which aimed at reducing total fertility rate from the then 7.7 children per woman to 4.0 by 2015, and since then, several efforts have been made to reduce this high fertility levels (NOP, 1993). But, despite the effort, the rate still remains high, especially at regional-level.

Needless to say, local-level fertility analysis in Ethiopia is paramount important for it will help to effectively design flexible region-specific strategies or programs, which might be worthwhile in implementation of family planning programs, and other socio-economic policies down at regional levels. The modeling of fertility patterns is one of the essential methods/ approaches researchers often use to understand the demographic nature of a given population, and thereby, make budgeting, planning, and monitoring policy decisions at different levels, national and regional-levels.

In this work, we first took advantage of real data set from 2011 Demographic and Health Survey (DHS) of Ethiopia, from which we extracted one-year age specific fertility rate (ASFR) and examined its pattern at both national and regional levels. All in all, the plot under scrutiny revealed that the pattern of ASFRs at regional level was similar to that of at national level, except for some regions sampling units of which is much smaller compared to the others. On top of this, considerable variation in the pattern of ASFRs of women was also reflected across regions. To wrap up, the figure reveals the presence of huge disparity not only fertility intensity but also ASFR shapes across regions, and, this variation, in turn, calls for a flexible model, which can capture all the available information in and provide a good fit for the pattern in each region.

A large number of parametric and non-parametric models have been proposed in demographic litratures for modeling the one-year age specific fertility pattern of different countries, yet no model has been put forward to fit local-level curves(age specific fertility pattern)of developing countries including Ethiopian to date.

In this study, we, therefore, proposed a four parametric Skew Normal model (Mazzuco and Scarpa, 2011) to fit the fertility schedules shown at both country and regional levels of Ethiopia. This model has many similar properties to normal distribution and include extra parameters which regulates (represents) its skewness (Lin, 2009; Azzalini, 1985; Azzalini and Dalla Valle, 1996; Flecher et al., 2009; Gupta and Chen, 2004; Gupta et al., 2004; Lin et al., 2007) and fertility intensity. The main capability of this class of distributions in applications is its ability in capturing and modeling departures from symmetry, whilst retaining tractability and closeness (Azzalini, 1985; Liseo and Loperfido, 2003; Minozzo and Ferracuti, 2012).

In order to determine the performance of this proposed model, we further conducted some preliminary analysis of fitting the model along with ten other commonly used parametric and non-parametric models in demographical litratures, namely: the Quadratic Spline function (Schmertmann, 2003), Cubic Splines (Hoem and Rennermalm, 1977) ,Coale-Trussell function (Coale and Trussell, 1978), the Beta and Gamma (Hoem et al., 1981), the Hadwiger distribution (Hadwiger, 1940),the Polynomial models (Brass, 1960), the Adjusted Error Model (Gayawan et al., 2010),Gompertez curve (Pasupuleti and Pathak, 2010),Skew Normal(Mazzuco and Scarpa, 2011), and Model 1 and Model 2 (Peristera and Kostaki, 2007). The purpose of fitting various models was to compare the performance of the proposed model, Skew Normal Distribution, with those other models mentioned. The criterion followed in fitting these models was *Nonlinear Regression with nonlinear least squares* (`nls`) estimation. We used **Akaike Information Criterion (AIC)** as model selecction criterion and the results obtained from this preliminary analysis testified that the values of the AIC for the proposed model, Skew Normal (SN), is lowest: *in the capital, Addis Ababa, Dire Dawa, Harari, Affar, Gambela, Benshangul-Gumuz*, and *country level data* as well. On the contrary, its value was also higher on some of the models and lower the rest on the remain regions, namely: *Tigray, Oromiya, Amhara, Somali* and *SNNP*. This tells us that the proposed model was better able to capturing the pattern of fertility at the empirical fertility data of Ethiopia and its regions than the other existing models considered since still the values of AIC for SN model was lowest not only at country level but also for majority regions, *i.e.*, in 6 of the 11 regions.

For many demographers, however, estimating regional-specific ASFRs and the associated uncertainty introduced due those factors can be difficult, especially in a situation where we have extremely varying sample size among different regions. Recently, it has been proposed that Hierarchical procedures might provide more reliable parameter estimates than Non-Hierarchical procedures, such as complete pooling and independence to make local/regional-level analyses. In this study, a Hierarchical Bayesian model was formulated to explore the posterior distribution of model parameters (for generation of *region-specific ASFR point estimates* and *uncertainty bound*). Beside to this procedure, in this thesis, maximum likelihood and Bayesian procedures had also been implemented to estimate the parameters of the proposed ASFR model and compare the result obtained from those procedures with our Hierarchical Bayesian counterparts.

Hierarchical Bayesian models offer many advantages, including *the ability to borrow strength/ information for smaller regions from other regions to obtain more accurate estimates* and the ability to specify complex models that reflect physical realities, to estimate the uncertainties in parameter estimates due to other approaches and factors. Because model fitting takes place in a single step, estimation uncertainty is properly propagated at all levels.

One stumbling block encounter in using this methodology was computational intractability. That is, the joint posterior distributions was non-linear, and too complex & intractable to easily drive the full conditional in standard/closed form. Data *Augmentation strategy* (*latent variable method*) has, hence, been instrumented as possible remedy in this respect. Although Griggio (2013/2014) incorporated the skew-normal latent variables in inference from a multimonial model, no studies were found considering skew-normal model proposed by Mazzuco and Scarpa (2015) as latent variables in analysis of binary data modeled via Bayesian Analysis. Therefore, as another new contribution, in this study a skew-normal (Mazzuco and Scarpa, 2015) latent variable methodology has been implemented in our Hierarchical Bayesian model developed from Binomially distributed fertility (birth) data so as to overcome this computational plight.

Eventually, the results from the all analyses are then combined, and inferences are made from these results. Thus, the overall result indicates that the Hierarchical Bayesian model had offered results that are superior to, more precise and reliable than those of other methods employed here, the mainstream Bayesian model as well as the maximum likelihood method. Therefore, the proposed methodology based on the augmenting skew-normal latent variables in analyzing binary fertility data is an adequate and appropriate model, and a novel way for fitting or inferring the ASFR model parameters.
The other point evident from the result is that almost all the regions had varied fertility rate distribution or shape patterns, *for example:-*

- ★ Somali, Affar, Gambela, and Benshangul-gumuz regions; Addis Ababa, Dire Dawa, and Harari city administrations had asymmetric or skewed distributions. Of those,the distribution of fertility in Somali, Affar, Gambela, and Benshangul-gumuz region were characterized by negative skewness. The rest, however, had negatively skewed fertility pattern or shape.

- ★ In the same token, relatively mild symmetric distribution had also observed on on the remaining regions, namely Tigray, Amhara, SNNP and Oromiya .

The posterior predictive distribution also clearly tells us:

- ★ a delay in childbirth for Addis Ababa, Dire Dawa, and Harari city administrations;
- ★ an early in childbirth or fertilty (delivery of the first child) for regions, such as Somali, Affar, Gambela, and Benshangul-gumuz;
- ★ the result also reports delivery of the first child at about a normal age (*i.e.*, from age 25 to 28) for the rest of the regions, *viz.*, Tigray, Amhara, SNNP and Oromiya regions.

Our fining also shows that the credible interval for the Hierarchical Bayesian Estimates was much narrower than that of the Bayesian Estimates or the Maximum likelihood Estimates. Similarly, the Bayesian credible intervals were also shown to be smaller than most of the confidence interval of the the Maximum likelihood Estimates. This suggests the fact that the Hierarchical Bayesian Estimates provided a more precise result in almost all the regions compared its Non-Hierarchical Bayesian counterparts, which are Bayesian and Maximum Likelihood Estimates. In contrary, Maximum Likelihood Estimate gave as similar result as Bayesian Estimates though the former seemed prone to a larger uncertainty.
Last but not the least, our Hierarchical procedure produces more precise and reliable estimates of the model parameters, in particular conferring statistical significance to the regions from which smaller number of childbearing age mothers were sampled as compared to the other procedures, which is mainly due to the ability of the Hierarchical procedure to allow for borrowing of information/strength

5.2 RECOMMENDATIONS:

As matter of policy implications, we need to draw several proposals at both the regional and national levels:

- ♣ ***At Regional-Level:-*** the lowering of high fertility rate need to be addressed through:
  - ★ integrating family planning and safe motherhood programmes into primary health care systems,
  - ★ providing access to reproductive health services,
  - ★ promoting the responsibility of men in sexual and reproductive health,
  - ★ raising the minimum legal age at marriage,
  - ★ improving female education and employment opportunities,
  - ★ discouraging son preference, and
  - ★ providing low cost, safe and effective contraception
  - ★ creating awareness on early and universal marriage, the high social and economic value attached to children,

- ★ fighting the cultural and traditional barriers to effectively utilize modern birth control methods
- ★ fighting the attitude of the desire for more children and extremely low contraceptive/practice
- ★ reducing the high child mortality,and womens limited achievements in the sphere of educational

♣ ***At National-Level:-*** the government need to :

- ★ design viable policies in relation family planning program , population and fertility policies that can address all the above sticking points.

As a caveat, this study is not aimed at tackling the issues,"which approach of parameter estimation is the most appropriate to address the study objectives?". This issue is left for future research

Appendix A Demography Part:Results (Tables & Plots)

Table A1: **Empirical and fitted values for *Tigray Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0817279	0.08630799	0.065074286	0.03295330	0.06794583	0.03558571	0.080467145
20-24	0.1588390	0.17439386	0.195627557	0.16938082	0.16094149	0.16854050	0.160139302
25-29	0.2383333	0.21971996	0.240953657	0.26335743	0.25420035	0.26421572	0.238886082
30-34	0.2675810	0.22228629	0.222286286	0.23218789	0.25435024	0.23254050	0.266190951
35-39	0.1587785	0.18209284	0.160859143	0.14679462	0.15369295	0.14606639	0.159586579
40-44	0.0406856	0.09913963	0.077905929	0.07456827	0.05399869	0.07414573	0.041499636
45-49	0.0114219	-0.02657336	-0.005339657	0.03244061	0.01073420	0.03264048	0.004680983
AIC		-19.49772	-20.35797	-19.44974	-52.76935	-21.7426	-33.45919
RSS		0.00806507	0.00535985	0.006102407	5.2274e-05	0.005852363	0.0008247563

Table A2: **Empirical and fitted values for *Affar Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	9.467930e-02	0.13672459	0.1061441190	0.10087416	0.09474379	0.10622310	0.10985166
20-24	2.481186e-01	0.18339025	0.2139707167	0.23050349	0.24200980	0.22447948	0.24712998
25-29	2.083633e-01	0.20027239	0.2308528595	0.23331030	0.22335031	0.23463698	0.22223935
30-34	1.651468e-01	0.18737101	0.1873710143	0.16792891	0.16919914	0.17043986	0.16218115
35-39	1.462091e-01	0.1446861	0.1141056476	0.10009311	0.10557445	0.10031325	0.10303961
40-44	3.211070e-02	0.07221769	0.0416372262	0.05291610	0.05425861	0.05176682	0.05701502
45-49	2.310000e-10	-0.03003425	0.0005462169	0.02574503	0.02296823	0.02449042	0.02747613
AIC		-18.70674	-23.50529	-22.06307	-23.45833	-23.57737	-24.54359
RSS		0.00902989	0.0034189	0.00420112	0.003441911	0.00450296	0.00294761

Table A3: **Empirical and fitted values for *Amhara Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0668713	0.107001648	0.08032641	0.08009104	0.06687130	0.08717417	0.10307644
20-24	0.2181686	0.152341879	0.17901711	0.18890896	0.21001767	0.18390267	0.16417801
25-29	0.1729865	0.172169993	0.19884523	0.20241993	0.18965640	0.20103901	0.19342808
30-34	0.1475083	0.166485990	0.16648599	0.15526166	0.14974716	0.15655781	0.16873908
35-39	0.1310119	0.135289871	0.10861464	0.09888596	0.10337824	0.10024709	0.10911515
40-44	0.0524531	0.078581636	0.05190640	0.05593741	0.06239921	0.05681760	0.05236563
45-49	0.0192326	-0.003638717	0.02303652	0.02914290	0.03293127	0.02971272	0.01867461
AIC		-19.97963	-23.84034	-24.19125	-29.75755	-25.28034	-20.08349
RSS		0.00752852	0.003259112	0.00309976	0.001399534	0.003530553	0.00557419

Table A4: **Empirical and fitted values for *Oromiya Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0941743	0.13610875	0. 09296405	0.09254963	0.09221398	0.09744416	0.09426439
20-24	0.2173480	0.18302150	0.22616620	0.22546710	0.22521853	0.22062705	0.21402413
25-29	0.2661646	0.20352759	0. 24667229	0.24825021	0.24249969	0.24983312	0.26147107
30-34	0.1836348	0.19762701	0. 19762701	0.19539632	0.19833963	0.19776984	0.20304436
35-39	0.1172261	0.16531977	0. 12217507	0.12763248	0.13348039	0.12729374	0.11652082
40-44	0.0741960	0.10660587	0. 06346117	0.07402548	0.07391543	0.07200126	0.05327251
45-49	0.0609520	0.02148531	0. 06463001	0.03953608	0.03367925	0.03739313	0.01976478
AIC		-16.7296	-33.60102	-31.47664	-27.80372	-33.13733	-25.573
RSS		0.01197701	0.0008082	0.0010948	0.001850131	0.0011492	0.002545

Table A5: **Empirical and fitted values for *Somali Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0804650	0.114496355	0.08315714	0.08067907	0.08166555	0.08330195	0.08104655
20-24	0.2201506	0.184324971	0.21566419	0.21958816	0.21685691	0.21495268	0.21334690
25-29	0.2547025	0.218202936	0.24954215	0.25401872	0.25780156	0.25876103	0.26123423
30-34	0.2044675	0.216130248	0.21613025	0.20992757	0.20931267	0.20985499	0.20916301
35-39	0.1463118	0.178106907	0.14676769	0.14045959	0.13590498	0.13537261	0.13615367
40-44	0.0817731	0.104132914	0.07279370	0.07608717	0.07617382	0.07566442	0.07597935
45-49	0.0217321	-0.005791731	0.02554749	0.03336950	0.03847125	0.03845859	0.03650429
AIC		-21.36322	-40.8871	-37.4672	-42.31808	-38.72417	-37.45082
RSS		0.006178298	0.000285419	0.0004652224	0.0002326493	0.0005173205	0.0004663124

Table A6: **Empirical and fitted values for *Benshangul Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.1238722	0.1460335048	0.08315714	0.11351211	12608168	0.11392909	0.12666666
20-24	0.2151792	0.1926315786	0.21566419	0.22381204	0.20893038	0.22245636	0.20682294
25-29	0.2248832	0.2108340643	0.24954215	0.23897705	0.23900296	0.24021093	0.23997957
30-34	0.2257467	0.2006409619	0.21613025	0.19074084	0.20330154	0.19127062	0.20468576
35-39	0.1093836	0.1620522714	0.14676769	0.12836562	0.13438618	0.12782876	0.13327614
40-44	0.0891213	0.0950679929	0.07279370	0.07721639	0.06903116	0.07662291	0.06862643
45-49	0.0187623	-0.0003118738	0.02554749	0.04289867	0.02755575	0.04276045	0.02868261
AIC		-22.84393	-27.84823	-25.1824	-27.79231	-27.29466	-40.8871
RSS		0.005000371	0.001838405	0.002690513	0.00185315	0.002647707	0.000285419

Table A7: **Empirical and fitted values for *Gambela Region* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0766938	07592061	0.06787381	0.04918594	0.06811252	0.05986901	0.06861532
20-24	0.0871907	0.10469356	0.11274036	0.11327848	0.10578113	0.11372587	0.10562596
25-29	0.1497611	0.11867410	0.12672090	0.13634118	0.12887706	0.13100025	0.12852152
30-34	0.1047982	0.11786221	0.11786221	0.12061097	0.12408076	0.11685628	0.12441163
35-39	0.1183587	0.10225791	0.09421111	0.08907685	0.09652425	0.09007362	0.09652952
40-44	0.0373788	0.07186118	0.06381438	0.05842111	0.06076234	0.06344817	0.06052297
45-49	0.0437603	0.02667203	0.03471883	0.03522449	0.03095271	0.04214314	0.03092546
AIC		-26.0029	-24.91371	-23.88051	-25.93929	-25.93044	-26.49806
RSS		0.003184297	0.002795791	0.003240459	0.00241477	0.002417826	0.002966829

Table A8: **Empirical and fitted values for *Dire Dawa* with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	5.010350e-02	0.08835550	0.0510356190	0.048460904	0.051253734	0.050085685	0.05412441
20-24	1.638853e-01	0.12252853	0.1598484119	0.161477650	0.160490112	0.159773533	0.15650967
25-29	1.654942e-01	0.13462922	0.1719491048	0.176166682	0.175080656	0.177321238	0.17886808
30-34	1.287507e-01	0.12465758	0.1246575809	0.113150784	0.117376355	0.113448283	0.11463077
35-39	5.535670e-02	0.09261361	0.0552937238	0.053275581	0.052961367	0.052623791	0.05203413
40-44	2.190000e-10	0.03849730	0.0011774169	0.020448552	0.016083252	0.019927123	0.01775147
45-49	5.340000e-10	-0.03769134	-0.0003714566	0.006791152	0.003287189	0.006589209	0.00457384
AIC		-19.18482	-33.70871	-33.37825	-36.83299	-35.33994	-50.0479
RSS		0.008433755	0.0007958749	0.0008343482	0.0005093404	0.0008389264	7.711299e-05

Table A9: **Empirical and fitted values for SNNP with values of minimization and model selection criterion**

AgeGroup	DHS	Qud Spline	Cub.Spline	Gamma Dist	PeristeraK	Hadwiger	Skew Normal
15-19	0.0804650	0.114496355	0.08315714	0.08166556	0.08067907	0.08330196	0.08104655
20-24	0.2201506	0.184324971	0.21566419	0.21685691	0.21958816	0.21495268	0.21334690
25-29	0.2547025	0.218202936	0.24954215	0.25780156	0.25401872	0.25876103	0.26123423
30-34	0.2044675	0.216130248	0.21613025	0.20931267	0.20992757	0.20985499	0.20916301
35-39	0.1463118	0.178106907	0.14676769	0.13590498	0.14045959	0.13537261	0.13615367
40-44	0.0817731	0.104132914	0.07279370	0.07617383	0.07608717	0.07566443	0.07597935
45-49	0.0217321	-0.005791731	0.02554749	0.03847125	0.03336949	0.03845859	0.03650429
AIC		-21.36322	-40.8871	-37.4672	-42.31808	- 38.72417	-37.45082
RSS		0.006178298	0.000285419	0.0004652	0.0005173	0.0005173205	0.0004663124

Appendix B Bayesian Part Part:Results of Bayesian Analysis & Maximum likelihood estimates

B.0.1 Posterior moments of Simple Bayesian Analysis & Maximum likelihood estimates

Table B1: **Bayesian and ML estimates for fertilty data set of Tigray Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	27.0255	27.0293	0.2870	26.4496	27.5856	28.601
ω^2	8.4996	8.4988	0.1573	8.1947	8.8165	8.6145
α	0.0109	0.0107	0.0736	-0.1129	0.1362	0.0918
R	4.3912	4.3616	0.2459	3.9913	4.8791	4.301

Table B2: **Bayesian and ML estimates for fertility data set of Afar Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	23.5317	20.5320	2.2910	19.0484	28.0269	21.699
ω^2	9.2260	8.4564	0.8952	7.4710	10.9863	10.4781
α	1.6782	1.2815	0.6032	0.4905	2.8633	1.118
R	5.332	5.012	1.653	4.1269	6.5370	5.0168

Table B3: **Bayesian and ML estimates for fertility data set of Amhara Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	27.0393	27.0357	0.2976	26.4502	28.6227	26.981
ω^2	8.9972	8.9936	0.1135	8.7909	9.2288	8.8698
α	-0.3475	-0.3472	0.0359	-0.4171	-0.2788	-0.2179
R	4.0369	4.4301	0.2510	3.5485	4.5297	4.3346

Table B4: **Bayesian and ML estimates for fertility data set of Oromiya Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	26.4417	26.4383	0.2867	25.8880	27.0117	26.0718
ω^2	8.5072	8.5067	0.0586	8.3947	8.6246	8.5882
α	0.0441	0.0443	0.0331	-0.0209	0.1092	0.0864
R	4.9892	4.9818	0.1952	4.6059	5.3792	5.2115

Table B5: **Bayesian and ML estimates for fertility data set of Somali Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	20.3679	19.0645	0.8835	18.6361	22.0996	19.4602
ω^2	8.5087	8.00831	0.6588	7.2176	9.7966	8.2069
α	3.1491	2.8492	1.0316	1.1286	5.1704	3.0015
R	7.3932	7.0161	1.1521	5.1304	9.6516	6.962387

Table B6: **Bayesian and ML estimates for fertility data set of Benishangul Gumuz Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	23.7684	23.7688	0.2913	23.2003	24.3370	24.0867
ω^2	8.5172	8.5161	0.06031	8.4011	8.6388	8.4859
α	0.26507	0.26569	0.1340	0.00243	0.52771	0.1184
R	5.1360	5.1369	0.1562	4.8248	5.4452	5.2202

Table B7: **Bayesian and ML estimates for fertility data set of SNNP Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	28.7132	28.7149	0.2887	28.1503	29.2865	28.6078
ω^2	8.5051	8.5042	0.0582	8.3929	8.6210	8.5400
α	-0.0378	-0.03762	0.0334	-0.1041	0.02813	0.01664
R	4.8120	4.8109	0.11671	4.5848	5.0401	4.7813

Table B8: **Bayesian and ML estimates for fertility data set of Gambela Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	24.7697	24.7683	0.2822	24.2223	25.3278	24.6030
ω^2	8.5001	8.4997	0.05745	8.3896	8.6138	8.26878
α	0.1166	0.1159	0.0328	0.0522	0.1810	0.0664
R	4.2184	4.2349	0.1230	3.9401	4.5184	4.3410

Table B9: **Bayesian and ML estimates for fertility data set of Harari**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	30.79457	31.20562	0.9852944	28.6553	32.03131	28.601
ω^2	8.499285	8.898356	0.56299	7.9488744	9.413138	9.340
α	-1.07714018	-1.29857669	0.3329421	-1.893271	-0.745683	-1.118
R	3.791	4.393	2.651	4.137	4.850	3.948

Table B10: **Bayesian and ML estimates for fertility data set of Addis Ababa**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	32.0939	34.0472	1.2800	31.5851	36.6028	30.6539
ω^2	8.1325	8.5017	0.2578	7.6272	9.0079	8.6538
α	-3.3299	-3.0360	0.6327	-4.5699	-1.7954	-3.1168
R	1.9427	1.9452	0.3141	1.3271	2.5608	1.9848

Table B11: **Bayesian and ML estimates for fertility data set of Dire Dawa**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	30.231	31.012	0.4521	30.6732	32.0943	27.432
ω^2	8.7331	8.7336	0.05931	8.6182	8.8491	9.3407
α	-1.0192	-1.0195	0.03531	-0.08823	0.04932	-1.1189
R	3.197	3.197	0.0131	3.102	3.2260	2.8389

B.0.2 Convergence Diagnosis: Trace Plots for some selected regions

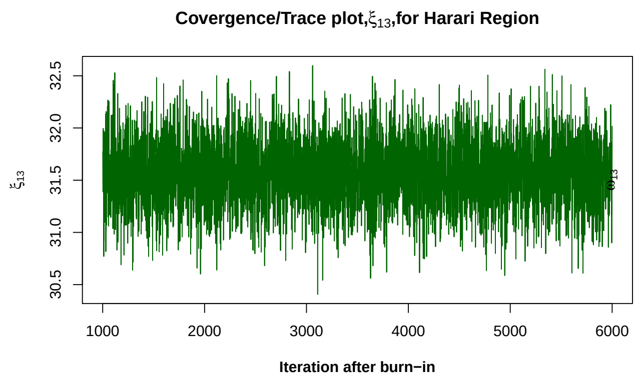


Figure B1: MCMC trace plots after burn-in for the location parameter,ξ_{Harari} ,of Harari

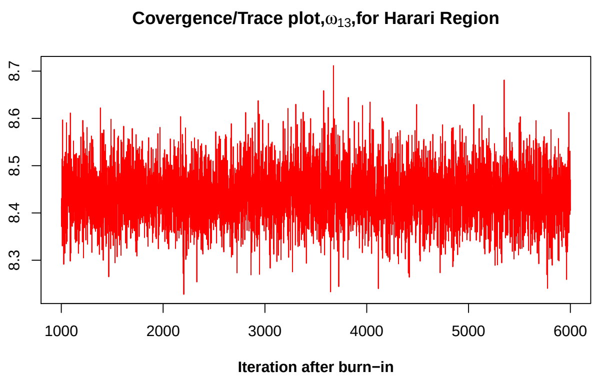


Figure B2: MCMC trace plots after burn-in for the scale parameter,ω_{Harari} ,of Harari

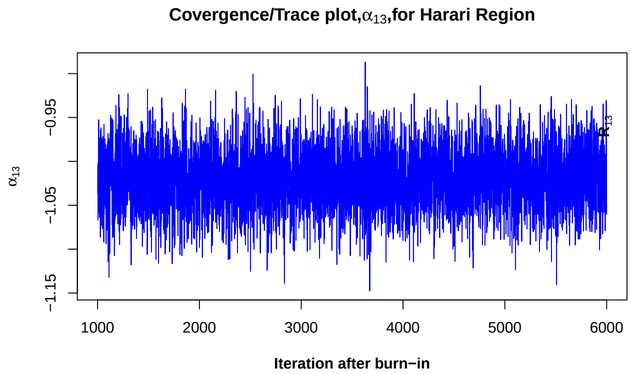


Figure B3: MCMC trace plots after burn-in for the shape parameter,α_{Harari} ,of Harari

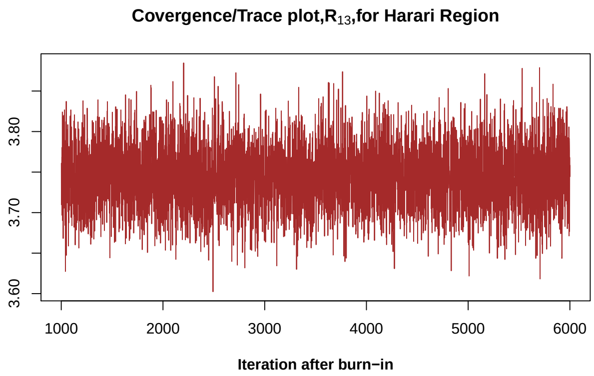


Figure B4: MCMC trace plots after burn-in for the fertility parameter,R_{Harari} ,of Harari

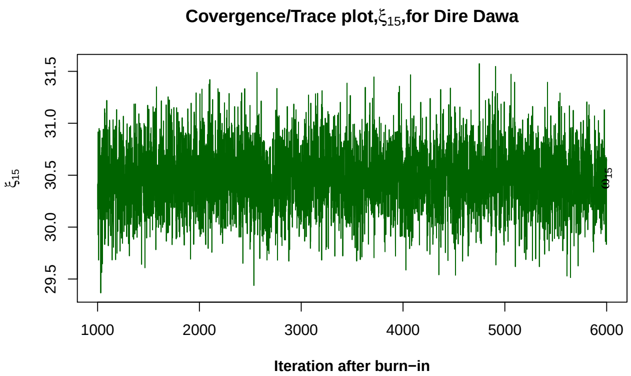


Figure B5: MCMC trace plots after burn-in for the location parameter,ξ_{Harari} ,of Dire Dawa

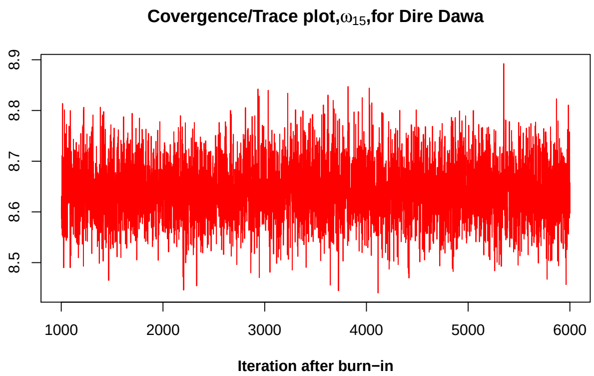


Figure B6: MCMC trace plots after burn-in for the scale parameter,ω_{Harari} ,of Dire Dawa

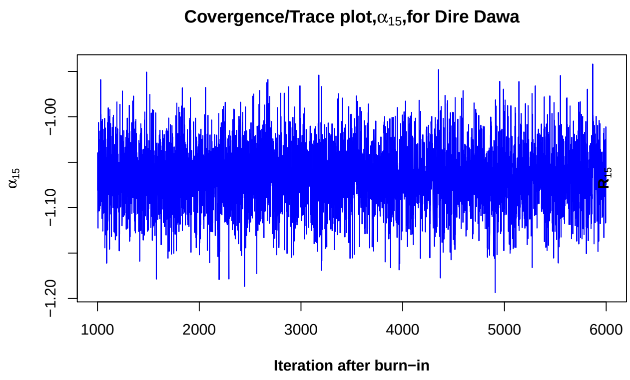


Figure B7: MCMC trace plots after burn-in for the shape parameter,α_{Harari} ,of Dire Dawa

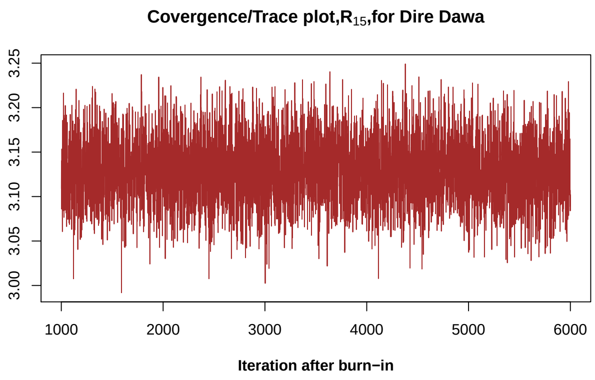


Figure B8: MCMC trace plots after burn-in for the fertility parameter,R_{Harari} ,of Dire Dawa

Appendix C Hierarchical Bayesian Part Part:Results of Hierarchical Bayesian Analysis & Maximum likelihood estimates

C.0.3 Posterior findings

Table C1: **Hierarchical Bayesian and ML estimates for fertility data set of Tigray**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	28.72822	28.73009	0.07351024	28.59748	28.84806	30.6539
ω^2	7.049991	7.055248	0.187495	6.768042	7.323103	8.6538
α	0.7716849	0.7694604	0.04478035	0.6896781	0.8612153	-3.1168
R	4.883442	4.880719	0.05767206	4.824510	4.956106	1.9848

Table C2: **Hierarchical Bayesian and ML estimates for fertility data set of Afar**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	22.72819	22.72966	0.07343009	22.59941	22.84841	30.6539
ω^2	7.051629	7.057748	0.1872722	6.768042	7.323103	8.6538
α	1.77646	1.776453	0.04487827	6.767261	7.314874	-3.1168
R	6.883853	6.880495	0.05783069	6.821847	6.953273	1.9848

Table C3: **Hierarchical Bayesian and ML estimates for fertility data set of Amhara Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	26.02817	26.02997	0.07340693	25.89888	26.14898	26.981
ω^2	8.55211	8.557161	0.1879005	8.256476	8.823803	8.8698
α	-0.2193244	-0.2193754	0.0451623	-0.2989179	-0.1305847	-0.2179
R	4.399365	4.396553	0.05205303	4.344941	4.461359	4.3346

Table C4: **Hierarchical Bayesian and ML estimates for fertility data set of Oromiya Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	26.52822	26.52922	0.07339988	26.40038	26.64873	26.0718
ω^2	8.553992	8.558146	0.1872623	8.264525	8.826085	8.5882
α	0.485459	0.4852391	0.04587836	0.4055632	0.5753489	0.0864
R	5.300462	5.297776	0.05258933	5.245917	5.364655	5.2115

Table C5: **Hierarchical Bayesian and ML estimates for fertility data set of Somali Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	20.52834	20.53031	0.07339265	20.39961	20.64772	19.4602
ω^2	8.653065	8.657814	0.1858623	8.370826	8.924242	8.2069
α	3.790433	3.789708	0.04608167	3.710736	3.883051	3.0015
R	7.102082	7.098769	0.05184025	7.047321	7.165859	6.962387

Table C6: **Hierarchical Bayesian and ML estimates for fertility data set of Benishangul Gumuz Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	23.52824	23.52985	0.07346258	23.39906	23.64838	24.0867
ω^2	8.653002	8.653458	0.1865011	8.365662	8.930263	8.4859
α	0.3945781	0.3923852	0.04712302	0.3124801	0.4903891	0.1184
R	5.300252	5.297851	0.05242365	5.245986	5.363053	5.2202

Table C7: **Hierarchical Bayesian and ML estimates for fertility data set of SNNP Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	28.52826	28.53004	0.07344476	28.39891	28.64850	28.6078
ω^2	8.856343	8.86319	0.1874363	8.568206	9.140805	8.5400
α	-0.2014707	-0.2028191	0.04773997	-0.2842188	-0.1085910	0.01664
R	5.12044	5.118172	0.05090992	5.067654	5.183093	4.7813

Table C8: **Hierarchical Bayesian and ML estimates for fertility data set of Gambela Region**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	22.32815	22.33011	0.07338661	22.19918	22.44827	24.6030
ω^2	5.25861	5.262248	0.1853738	4.973846	5.527638	8.26878
α	2.202881	2.201616	0.04859733	2.115641	2.298706	0.0664
R	2.883317	2.880293	0.05844014	2.821510	2.954272	4.3410

Table C9: **Hierarchical Bayesian and ML estimates for fertility data set of Harari**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	30.32818	30.32953	0.07347812	30.19912	30.44869	28.601
ω^2	7.340395	7.344207	0.1852949	7.054930	7.607353	9.340
α	-2.093463	-2.095367	0.04928861	-2.181773	-1.992148	-1.118
R	3.383518	3.3808	3.324127	3.454066	3.948	

Table C10: **Hierarchical Bayesian and ML estimates for fertility data set of Addis Ababa**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	33.92821	33.6300	0.0733869	33.49898	33.74815	32.6539
ω^2	8.343236	8.347118	0.1862105	8.055792	8.612995	8.6538
α	-3.289408	-3.292281	0.05056788	-3.380609	-3.183485	-3.1168
R	1.383188	1.37989	0.05781846	1.322299	1.455633	1.9848

Table C11: **Hierarchical Bayesian and ML estimates for fertility data set of Dire Dawa**

Parameters	Bayesian Estimates					ML Estimate
	Mean	Median	SD	95%[Conf	Interval]	MLE
ξ	29.72826	29.72939	0.07344202	29.59898	29.84742	27.432
ω^2	8.172812	8.180471	0.1860283	7.893764	8.441436	9.3407
α	-1.286179	-1.28903	0.05139713	-1.38127	-1.18204	-1.1189
R	2.883449	2.880347	0.05751831	2.822901	2.952694	2.8389

C.0.4 Convergence Diagnosis for Hierarchical Bayesian model:MCMC Trace Plots for some selected regions

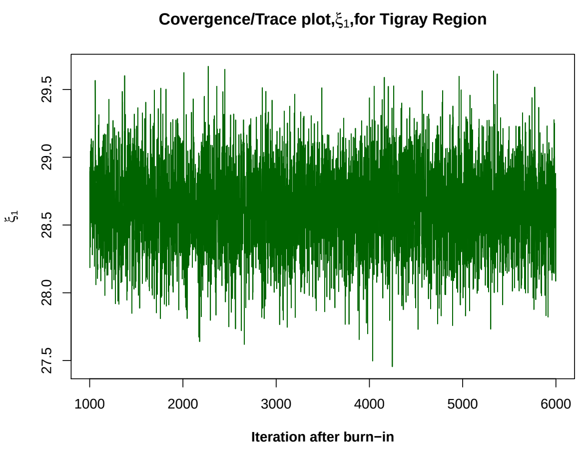


Figure C1: MCMC trace plots after burn-in for the location parameter of Tigray, ξ_{Tigray}

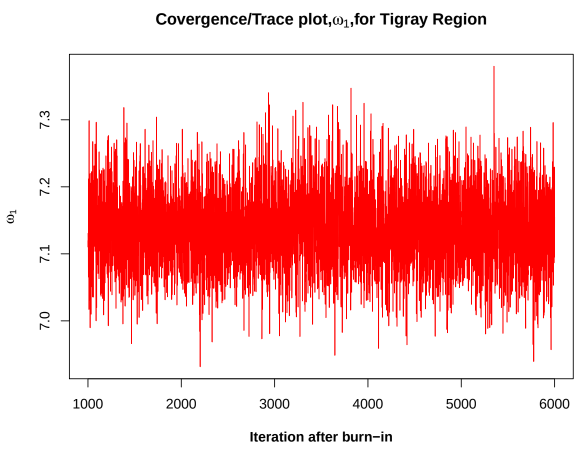
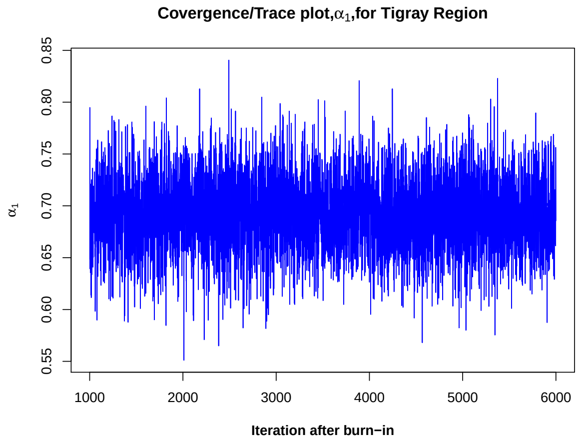


Figure C2: MCMC trace plots after burn-in for the scale parameter of Tigray, ξ_{Tigray}



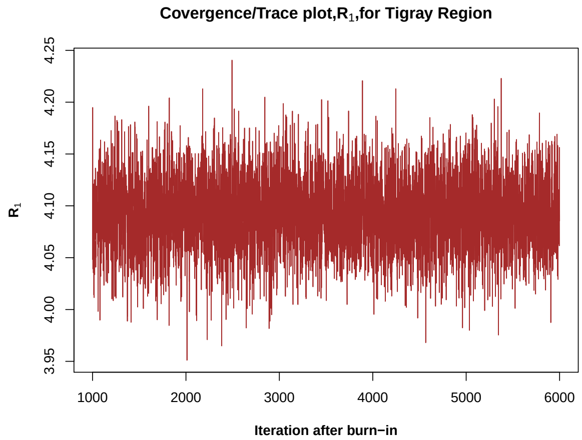
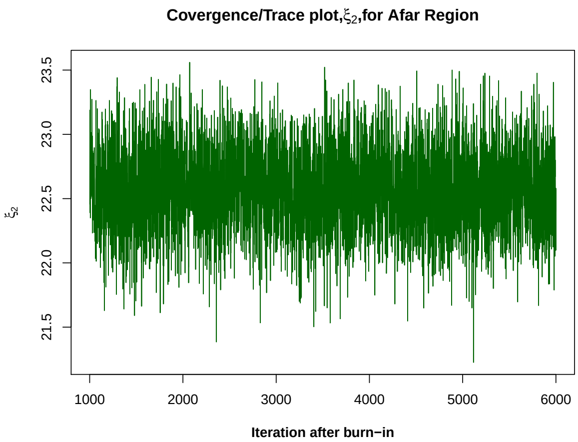


Figure C3: MCMC trace plots after burn-in for the shape parameter of Tigray,α_{Tigray}

Figure C4: MCMC trace plots after burn-in for the fertility parameter of Tigray, R_{Tigray}



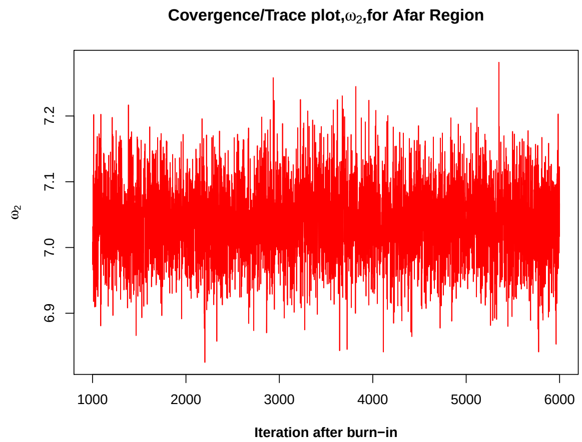


Figure C5: MCMC trace plots after burn-in for the location parameter of Afar, ξ_{Afar}

Figure C6: MCMC trace plots after burn-in for the scale parameter of Afar, ω_{Afar}

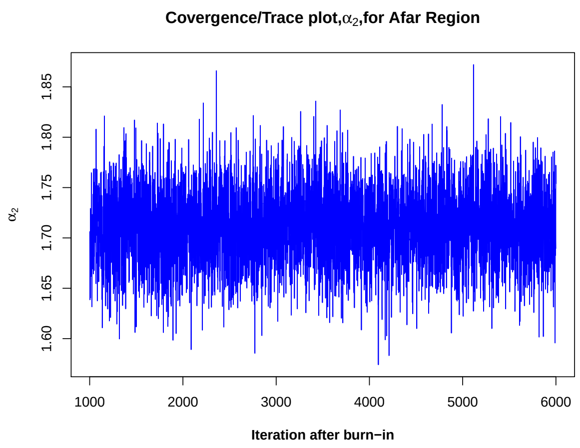


Figure C7: MCMC trace plots after burn-in for the shape parameter of Afar,α_{Afar}

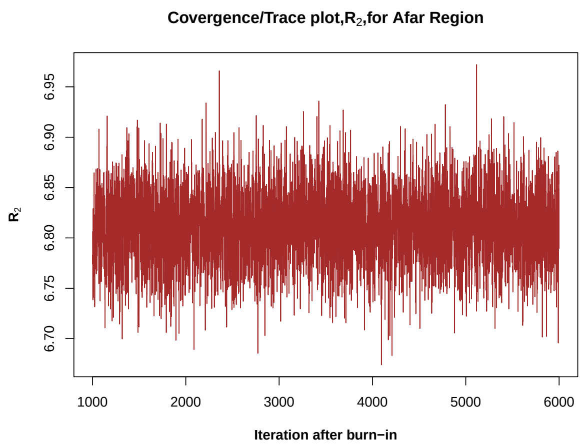


Figure C8: MCMC trace plots after burn-in for the fertility parameter of Afar, R_{Afar}

Table C12: **MCMC convergence diagnostic test for Hierarchical Bayesian Model of Afar Region fertility model parameters using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.431*passed*	0.2351	0.3421	0.39	
ω^2	*passed*	0.289	*passed*	0.0837	0.0093	1.13
α	*passed*	0.197	*passed*	0.1129	-0.0934	1.42
R	*passed*	0.152	*passed*	1.65	0.0671	3.01

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

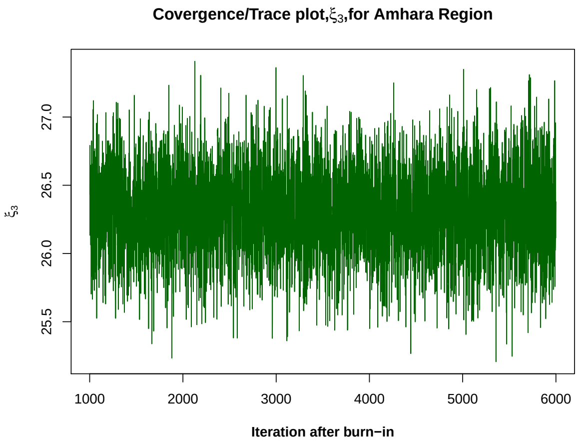


Figure C9: MCMC trace plots after burn-in for the location parameter of Amhara, ξ_{Amhara}

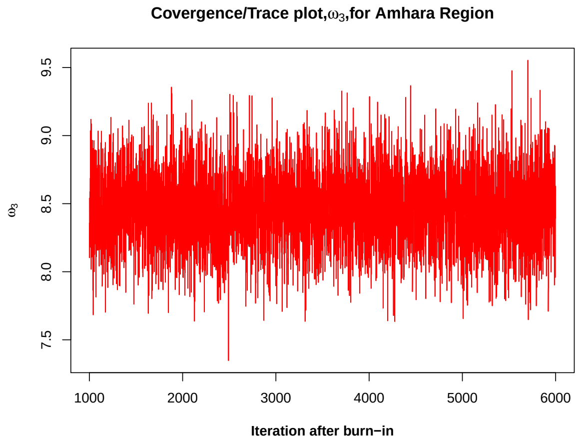


Figure C10: MCMC trace plots after burn-in for the s caleparameter of Amhara, ω_{Amhara}

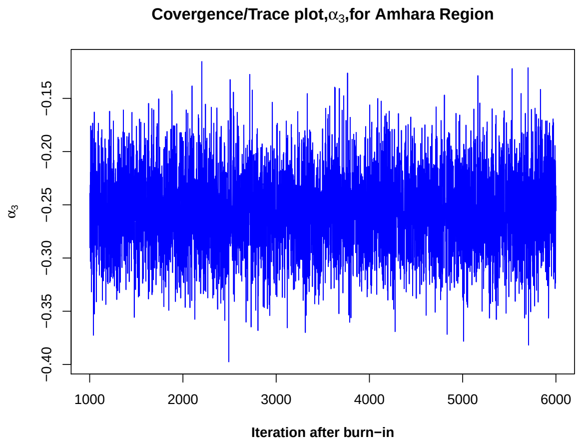


Figure C11: MCMC trace plots after burn-in for the shape parameter of Amhara, α_{Amhara}

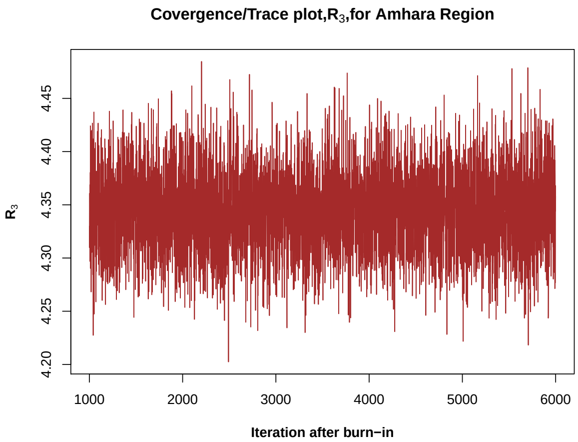


Figure C12: MCMC trace plots after burn-in for the fertility parameter of Amhara, R_{Amhara}

Table C13: **MCMC convergence diagnostic test for Hierarchical Bayesian Model of Amhara Region fertility model parameters using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.823*passed*	0.2351	1.43	2.31	
ω^2	*passed*	0.397	*passed*	0.0837	0.1840	2.49
α	*passed*	0.287	*passed*	0.1129	0.0451	1.81
R	*passed*	0.185	*passed*	1.65	0.9871	2.08

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

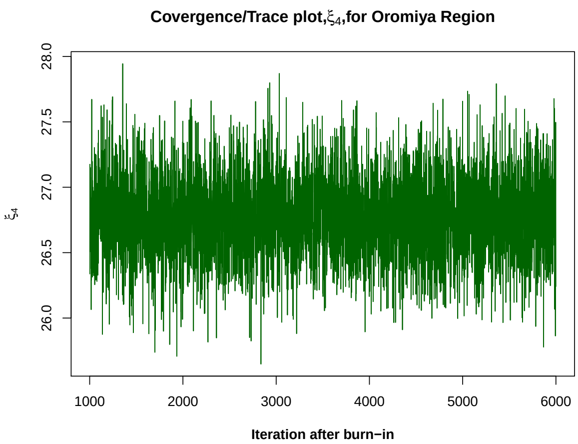


Figure C13: MCMC trace plots after burn-in for the location parameter of Oromiya, ξ_{Oromiya}

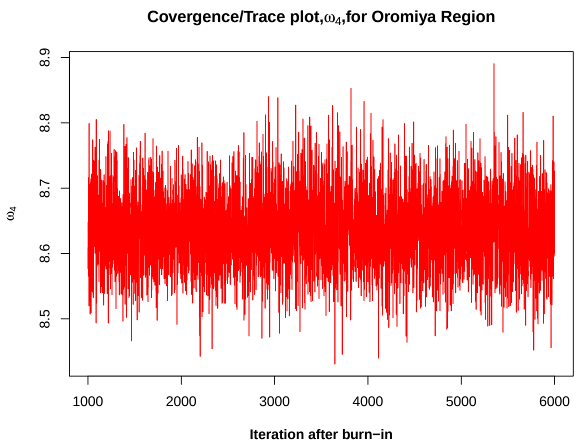


Figure C14: MCMC trace plots after burn-in for the scale parameter of Oromiya, ω_{Oromiya}

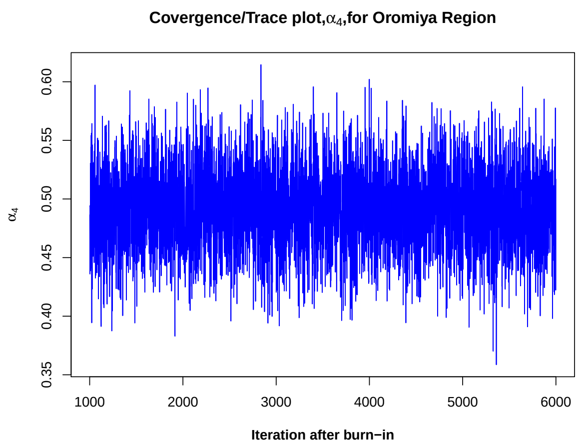


Figure C15: MCMC trace plots after burn-in for the shape parameter of Oromiya,α_{Oromiya}

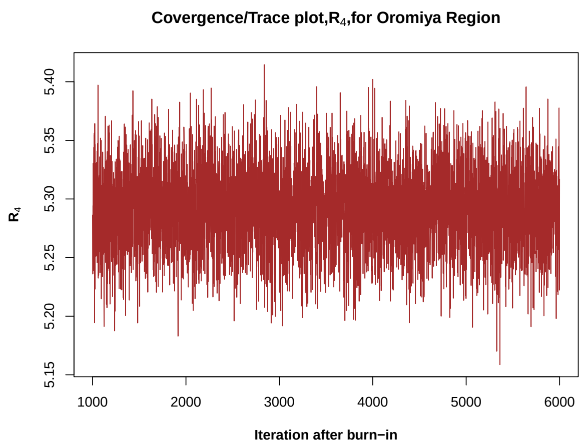


Figure C16: MCMC trace plots after burn-in for the fertility parameter of Oromiya, r_{Oromiya}

Table C14: **MCMC convergence diagnostic test for Hierarchical Bayesian Model of Oromiya Region fertility model parameters using Geweke, H-W and R-L**

Parameters	Heidelberger & Welch (H-W)				Geweke	R-L test
	St.test[a]	P-value	HW test[b]	HW test[c]	Z-score	Dep. factor(I)[d]
ξ	*passed*	0.517	*passed*	0.00893	-1.3	0.982
ω^2	*passed*	0.345	*passed*	0.00162	0.5092	1.02
α	*passed*	0.371	*passed*	0.00103	0.005691	1.02
R	*passed*	0.133	*passed*	1.65	0.0889	1.64

[a] Stationary Test ; [b] Half-width test; [c] Half-width test; [d] Dependence factor(I)

C.0.5 Posterior Predictive Distribution

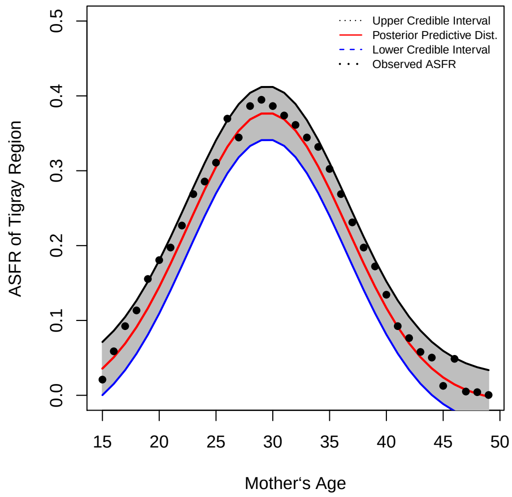


Figure C17: **95% Credible interval** of the **Posterior Predictive distribution** for **Tigray Region**

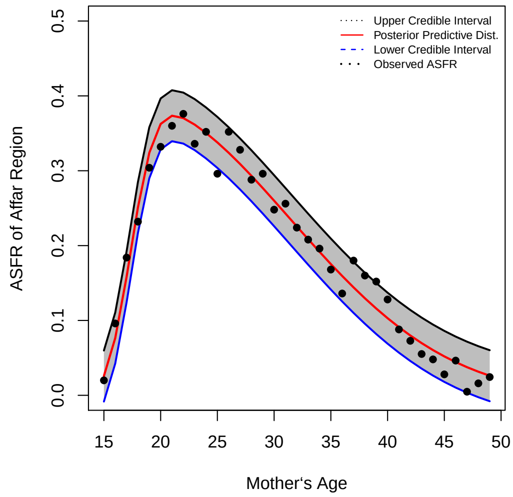


Figure C18: **95% Credible interval** of the **Posterior Predictive distribution** for **Affar Region**

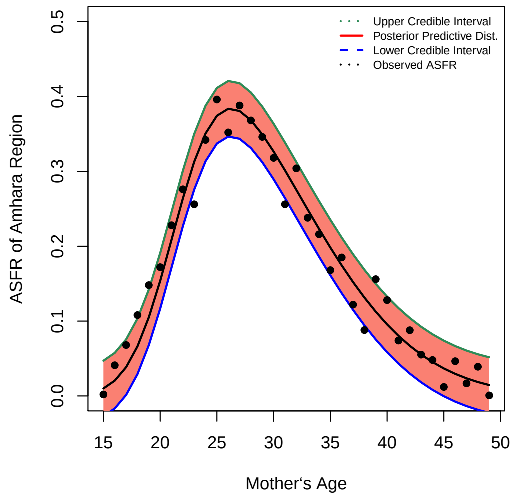


Figure C19: 95% Credible interval of the **Posterior Predictive distribution** for Amhara Region

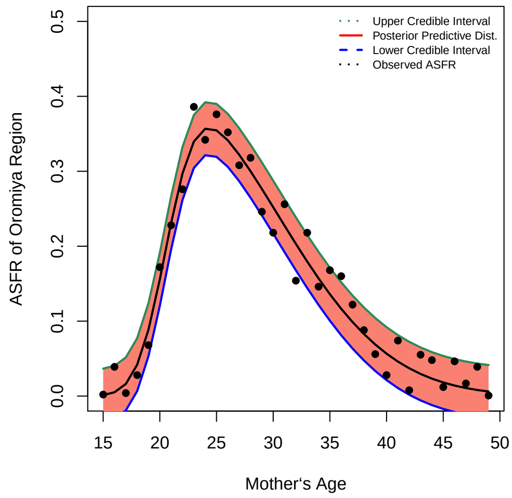


Figure C20: 95% Credible interval of the **Posterior Predictive distribution** for Oromiya Region

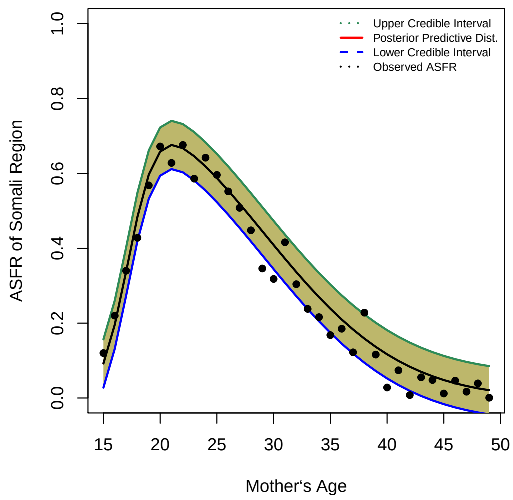


Figure C21: 95% Credible interval of the **Posterior Predictive distribution** for Somali Region

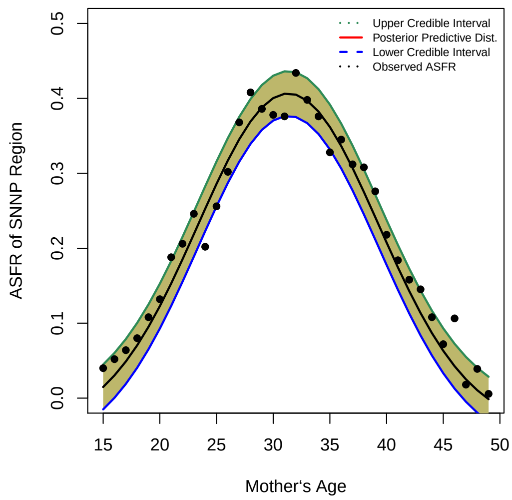


Figure C22: 95% Credible interval of the **Posterior Predictive distribution** for SNNP Region

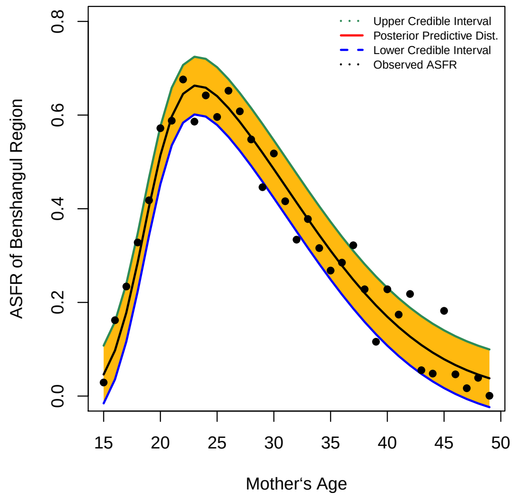


Figure C23: **95% Credible interval** of the **Posterior Predictive distribution** for **Benshangul Region**

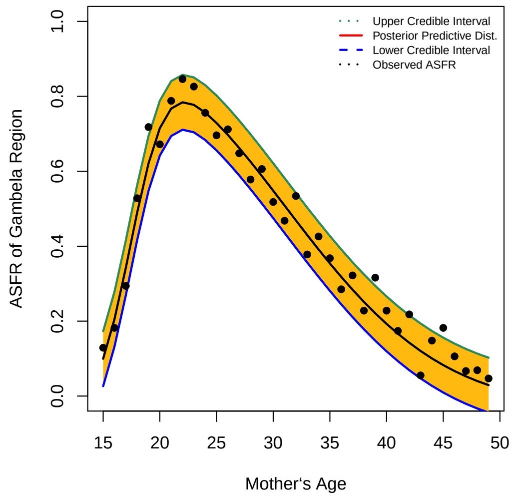


Figure C24: **95% Credible interval** of the **Posterior Predictive distribution** for **Gambela Region**

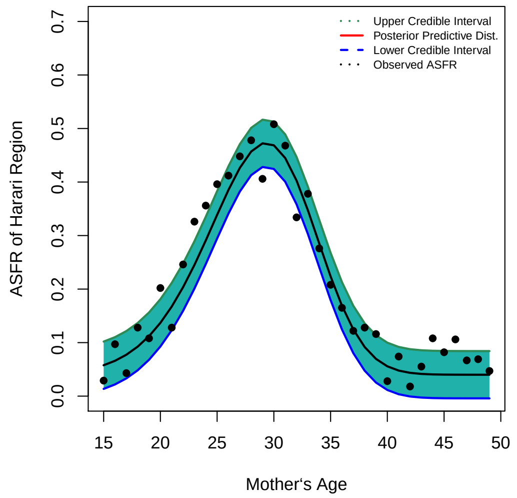


Figure C25: **95% Credible interval** of the **Posterior Predictive distribution** for **Harari Region**

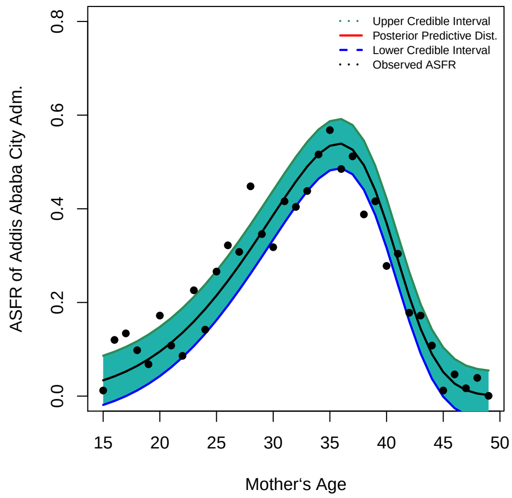


Figure C26: **95% Credible interval** of the **Posterior Predictive distribution** for **A.A. City Admin.**

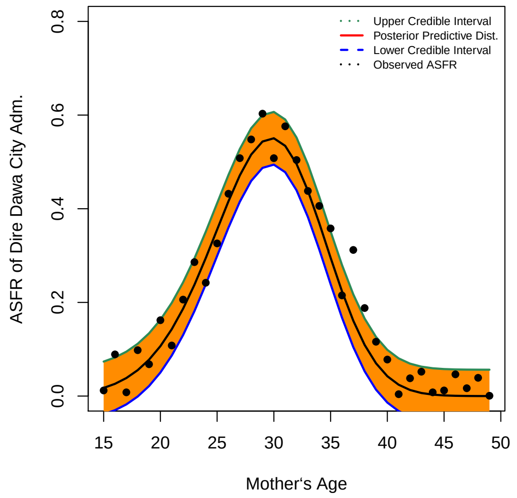


Figure C27: **95% Credible interval of the Posterior Predictive distribution** for **D.D. City Admin.**

Appendix D Maximum Likelihood Estimation

♣ ★ ♠ ♣ ★ ♠ **RESULT FOR TIGRAY REGION** ♣ ★ ♠ ♣ ★ ♠

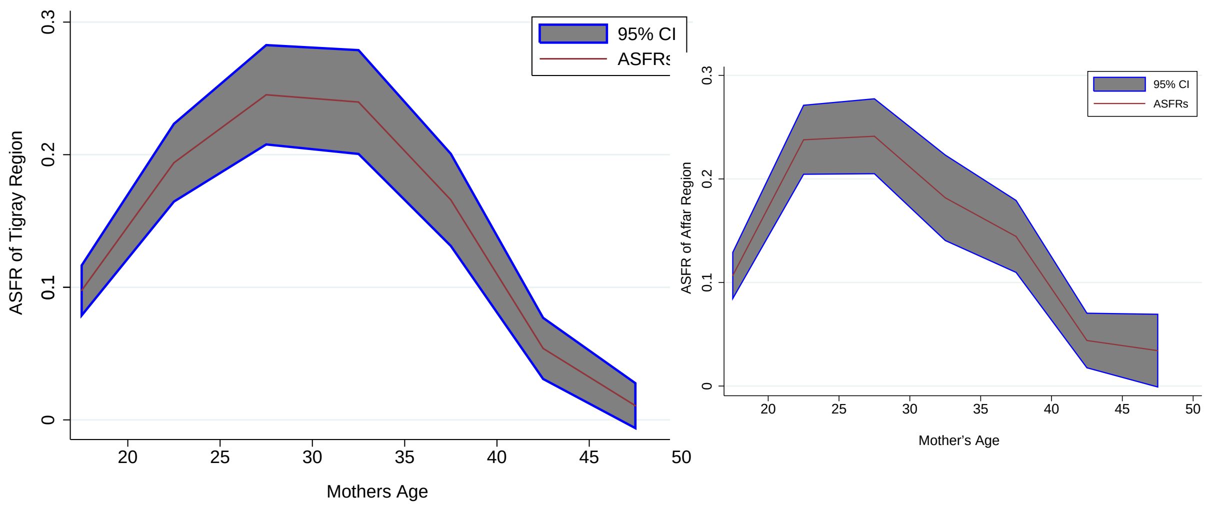


Figure D1: **95% Confidence interval for the MLE fitted ASFR model of Tigray & Afar Regions**

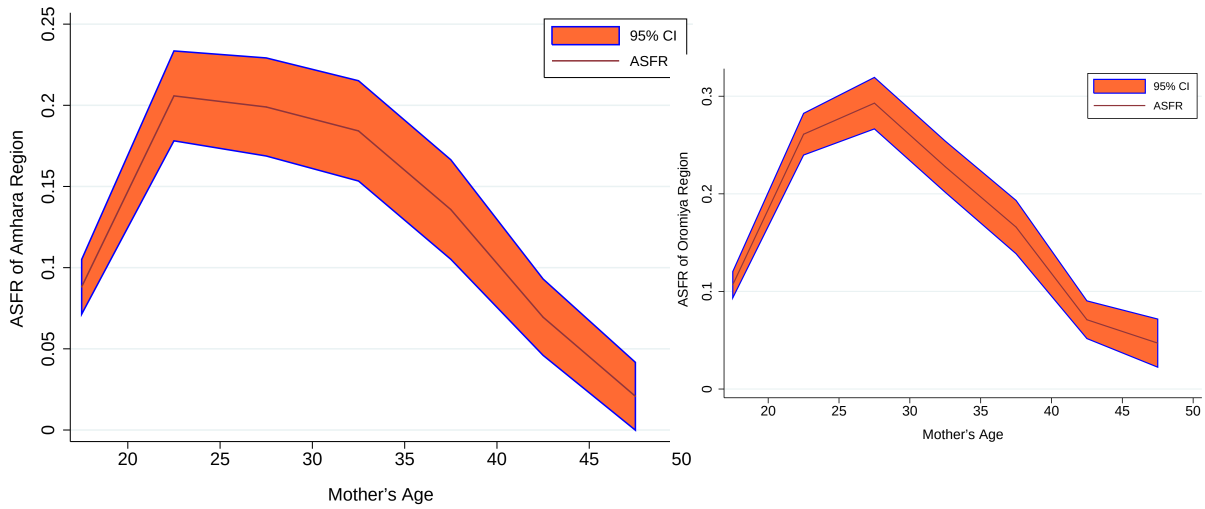


Figure D2: 95% Confidence interval for the MLE fitted ASFR model of Amhara & Oromiya Regions

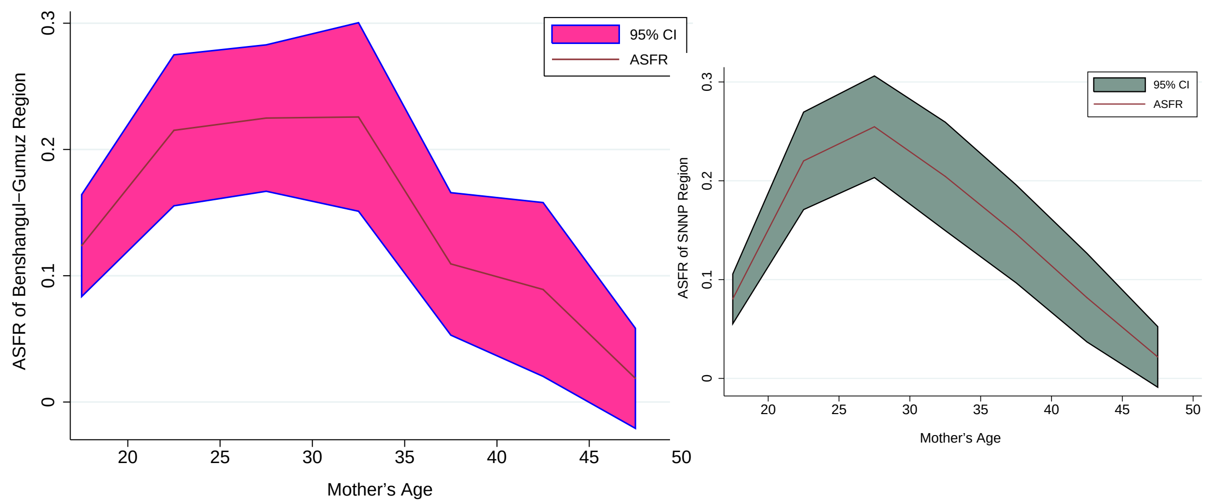


Figure D3: 95% Confidence interval for the MLE fitted ASFR model of SNNP & Benshangul Regions

Bibliography

Hirotugu Akaike. Canonical correlation analysis of time series and the use of an information criterion. *Mathematics in Science and Engineering*, 126:27–96, 1976. (page 21)

James H Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:160–170, 1993. (page 8, 55, 65, 67)

Tewodros Alemayehu, Jemal Haider, and Dereje Habte. Determinants of adolescent fertility in ethiopia. *Ethiopian Journal of Health Development*, 24(1), 2010. (page 17)

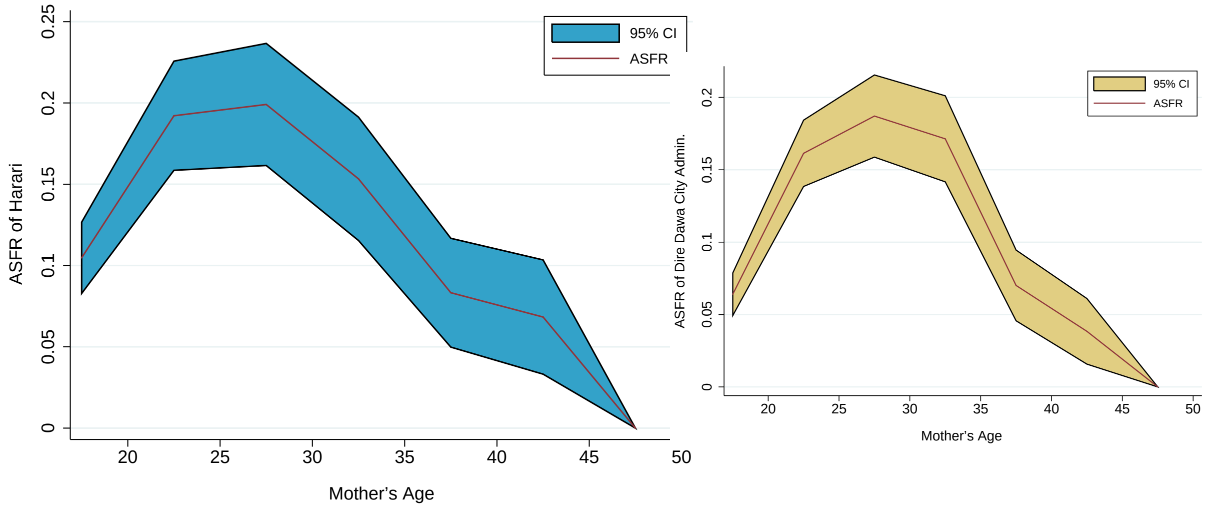


Figure D4: 95% Confidence interval for the MLE fitted ASFR model of SNNP & Dire Dawa Regions

Getu D Alene and Alemayehu Worku. Differentials of fertility in north and south gondar zones, northwest ethiopia: A comparative cross-sectional study. *BMC Public Health*, 8 (1):397, 2008. (page 17)

Reinaldo B Arellano-Valle and Adelchi Azzalini. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, pages 561–574, 2006. (page 30)

Reinaldo B Arellano-Valle and Adelchi Azzalini. The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7):1362–1382, 2008. (page 29)

V GOPALAKRISHNAN Asari and C John. Determinants of desired family size in kerala. *Demography India*, 27(2):369–381, 1998. (page 17)

Greg Atkinson and Alan M Nevill. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine*, 26(4):217–238, 1998. (page 68)

Azzalini. Further results on a class of distributions which includes the normal ones. *Biometrika*, 46(2):199–208, 1986. (page 29)

A Azzalini and A Capitanio. The skew-normal and related families. *Cambridge University Press, IMS Monographs series*, 2014. (page 27)

Adelchi Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985. (page 15, 27, 100, 101)

Adelchi Azzalini and Alessandra Dalla Valle. The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996. (page 100)

Radheshyam Bairagi. Effects of sex preference on contraceptive use, abortion and fertility in matlab, bangladesh. *International Family Planning Perspectives*, pages 137–143, 2001. (page 3)

James O Berger. Statistical decision theory and bayesian analysis, 1985. (page 40)

Jane T Bertrand, Janet Rice, Tara M Sullivan, and James Shelton. *Skewed method mix: a measure of quality in family planning programs*. Citeseer, 2000. (page 3)

Julian Besag. Markov chain monte carlo for statistical inference. *Center for Statistics and the Social Sciences*, 2001. (page 35)

Julian Besag, Peter Green, David Higdon, and Kerrie Mengersen. Bayesian computation and stochastic systems. *Statistical science*, pages 3–41, 1995. (page 35)

MK Bhasin and Shampa Nag. A demographic profile of the people of jammu and kashmir: Estimates, trends and differentials in fertility. *Journal of Human Ecology*, 13(1-2):57–112, 2002. (page 17)

Veena Bhasin and MK Bhasin. *Habitat, habitation, and health in the Himalayas*. Kamla-Raj Enterprises, 1990. (page 17)

JC Bhatia. Prevalent knowledge and attitude of males towards family planning in a punjab village. *Journal of Family Welfare*, 16(3):3–14, 1970. (page 17)

William Brass. The graduation of fertility distributions by polynomial functions. *Population Studies*, 14(2):148–162, 1960. (page 101)

William Brass. The use of the gompertz relational model to estimate fertility. 1981. (page 15)

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998. (page)

Emery N Brown, Robert E Kass, and Partha P Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature neuroscience*, 7(5):456–461, 2004. (page 93)

Scott Brown and Andrew Heathcote. Averaging learning curves across and within participants. *Behavior Research Methods, Instruments, & Computers*, 35(1):11–21, 2003. (page 64)

John C. Caldwell. Theory of Fertility Decline. *New York: Academic Press*, 1982. (page 3)

John C Caldwell, Israel O Orubuloye, and Pat Caldwell. Fertility decline in africa: A new type of transition? *Population and development review*, pages 211–242, 1992. (page 2)

Antonio Canale and Bruno Scarpa. Informative bayesian inference for the skew-normal distribution. *arXiv preprint arXiv:1305.3080*, 2013. (page 41, 48, 49, 50, 80, 81, 84)

Antonio Canale and Bruno Scarpa. Age-specific probability of childbirth. smoothing via bayesian nonparametric mixture of rounded kernels. *Statistica*, 75(1):101–110, 2015. (page 69)

John B Casterline and RT Lazarus. Determinants and consequences of high fertility: a synopsis of the evidence. *Addressing the Neglected MDG: World Bank Review of Population and High Fertility, World Bank publications*, 2010. (page 4)

Eduardo A Castro, Zhen Zhang, Arnab Bhattacharjee, José M Martins, and Tapabrata Maiti. Regional fertility data analysis: A small area bayesian approach. *Current Trends in Bayesian Methodology with Applications*, page 203, 2015. (page 25)

S Paul Chachra and MK Bhasin. Anthropo-demographic study among the caste and tribal groups of central himalayas: fertility differentials and determinants. *Journal of Human Ecology*, 9:417–429, 1998. (page 17)

T Chandola, DA Coleman, and RW Hiorns. Recent european fertility patterns: Fitting curves to distorteddistributions. *Population Studies*, 53(3):317–329, 1999. (page 13, 14)

Y Chen and D Fournier. Impacts of atypical data on bayesian inference and robust bayesian approach in fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(9):1525–1533, 1999. (page 26)

James S Clark. Why environmental scientists are becoming bayesians. *Ecology letters*, 8 (1):2–14, 2005. (page 69)

James S Clark and Alan E Gelfand. *Hierarchical Modelling for the Environmental Sciences: Statistical methods and applications: Statistical methods and applications*. Oxford University Press, 2006. (page 69)

James Samuel Clark. *Models for ecological data: an introduction*, volume 11. Princeton university press Princeton, New Jersey, USA, 2007. (page 69)

Ansley J Coale and T James Trussell. Technical note: Finding the two parameters that specify a model schedule of marital fertility. *Population Index*, pages 203–213, 1978. (page 101)

A Clifford Cohen. Maximum likelihood estimation in the weibull distribution based on complete and on censored samples. *Technometrics*, 7(4):579–588, 1965. (page 87)

Peter Congdon. *Applied bayesian modelling*. John Wiley & Sons, 2003. (page 31)

Mary Kathryn Cowles and Bradley P Carlin. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91 (434):883–904, 1996. (page 55, 57, 93)

Noel Cressie, Catherine A Calder, James S Clark, Jay M Ver Hoef, and Christopher K Wikle. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3):553–570, 2009. (page 69)

CSA and G ICF International. Central Statistical Agency, ORC Macro. Ethiopia 2011 Demographic and Heath Survey. Addis Ababa, Ethiopia ; Calverton, Maryland 2011. *Preliminary report*, 2012. (page 4)

JAA de Beer. A new relational method for smoothing and projecting age specific fertility rates: Topals. *Demographic Research*, 24, 2011. (page 16)

K Desta and G Seyoum. Family system in ethiopia. *Hand Book on Population and for Secondary School Teachers in Ethiopia. Edited by Seyoum Gebreselassie and Markos Ezra DTRC and ICDR, Addis Ababa. Family line Education*, 1998. (page 4)

R Development Core Team. Development core team. 2005. r: A language and environment for statistical computing. the r foundation for statistical computing, vienna, austria, 2005. (page iii, 90)

Debarshi Dey. Estimation of the parameters of skew normal distribution by approximating the ratio of the normal density and distribution functions. 2010. (page 27)

Dipak K Dey, Sujit K Ghosh, and Bani K Mallick. *Generalized linear models: A Bayesian perspective*. CRC Press, 2000. (page 68)

J Dominguez-Molina, G González-Farías, and AK Gupta. The multivariate closed skew normal distribution. Technical report, Technical report, 2003. (page 30)

Jean Drèze and Mamta Murthi. Fertility, education and development: Further evidence from india. *Vol*, 2000. (page 5)

W Eshetu and A Habtamu. The influence of selected social and demographic factors on fertility: the case of bahirdar town. *Ethiopian Journal of Development Research*, 20(1): 1–21, 1998. (page 17)

Ethiopian Society of Population Studies[ESPS] ESPS. Levels, Trends and Determinants of Lifetime and Desired Fertility in Ethiopia:Findings from EDHS 2005. *In-depth Analysis of the Ethiopian Demographic and Health Survey 2005*, 2008. (page 4)

William K Estes and W Todd Maddox. Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, 12(3):403–408, 2005. (page 64)

WK Estes. Traps in the route to models of memory and decision. *Psychonomic bulletin & review*, 9(1):3–25, 2002. (page 64)

Eleanor R Fapohunda and Michael P Todaro. Family structure, implicit contracts, and the demand for children in southern nigeria. *Population and Development Review*, pages 571–594, 1988. (page 3)

SM Farid. On the pattern of cohort fertility. *Population Studies*, 27(1):159–168, 1973. (page 15)

Patrizia Farina, Eshetu Gurmu, Abdulahi Hasen, and Dionisia Maffioli. *Fertility and Family Change in Ethiopia*. Central Statistical Authority, 2001. (page 2, 9)

C Fernández, S Cerviño, N Pérez, and E Jardim. Stock assessment and projections incorporating discard estimates in some years: an application to the hake stock in ices divisions viiic and ixa. *ICES Journal of Marine Science: Journal du Conseil*, 67(6): 1185–1197, 2010. (page 26)

Carmen Fernández, Eduardo Ley, and Mark FJ Steel. Bayesian modelling of catch in a north-west atlantic fishery. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(3):257–280, 2002. (page 26)

Yohannes Fitaw, Yemane Berhane, and Alemayehu Worku. Impact of child mortality and fertility preferences on fertility status in rural ethiopia. *East African medical journal*, 81 (6):300–306, 2004. (page 17)

Yohannis Fitaw, Yemane Berhane, and Alemayehu Worku. Differentials of fertility in rural butajira. *Ethiopian Journal of health development*, 17(1):17–25, 2003. (page 17)

Cedric Flecher, Ph Naveau, and Denis Allard. Estimating the closed skew-normal distribution parameters using weighted moments. *Statistics & Probability Letters*, 79(19): 1977–1984, 2009. (page 101)

John Fox. *Nonparametric simple regression: smoothing scatterplots*. Number 130. Sage, 2000. (page 16)

Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006. (page 41)

Dani Gamerman. Markov chain monte carlo: stochastic simulation for bayesian inference. 1997. *London: Chapman Hall*. (page 35)

Ezra Gayawan, Samson B Adebayo, Reuben A Ipinyomi, Benjamin A Oyejola, et al. Modeling fertility curves in africa. *Demographic Research*, 22(10):211–236, 2010. (page 101)

Samson Gebremedhin. *Level and Differentials of Fertility in Awassa Town*. PhD thesis, Addis Ababa University, 2006. (page 17)

Samson Gebremedhin and Mulugeta Betre. Level and differentials of fertility in awassa town, southern ethiopia: original research article. *African journal of reproductive health*, 13(1):93–112, 2009. (page 17)

Alan E Gelfand. Bayesian methods and modeling. *Encyclopedia of Environmetrics*, 2014. (page 43)

Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990. (page 35)

A Gelman and J Hill. Missing-data imputation. *Behavior research methods*, 43(2):310–30, 2007. (page 65, 66, 68, 69)

A Gelman, JB Carlin, HS Stern, and D Rubin. Bayesian data analysis/'chapman & hall, new york. 1995. (page 85)

Andrew Gelman and Xiao-Li Meng. Model checking and model improvement. In *Markov chain Monte Carlo in practice*, pages 189–201. Springer, 1996. (page 91)

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992. (page)

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. Bayesian data analysis, (chapman & hall/crc texts in statistical science). 2003. (page 35, 75)

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014. (page 26, 33, 34, 54, 67)

J Geweke. Evaluating the accuracy of sampling0based approaches to the calculation of posterior moments, bayesian statistics 4, ed. *JM Bernardo, JO Berger, AP David, and AFM Smith*, 1690193, 1992. (page 55, 90)

John Geweke, Michael Keane, and David Runkle. Alternative computational approaches to inference in the multinomial probit model. *The review of economics and statistics*, pages 609–632, 1994. (page 44)

Tewodros Tsegaye Gezahegn. Cross sectional study of women employment and fertility in ethiopia. 2011. (page 3)

E Gilje. Fitting curves to age-specific fertility rates: Some examples. *Statistical Review of the Swedish National Central Bureau of Statistics III*, 7:118–134, 1972. (page 13)

Walter R Gilks. The relationship between birth history and current fertility in developing countries. *Population Studies*, 40(3):437–455, 1986. (page 16, 71)

Andreas Glöckner and Thorsten Pachur. Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1):21–32, 2012. (page)

Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, pages 513–583, 1825. (page 15)

William E Griffiths, R Carter Hill, and Peter J Pope. Small sample properties of probit model estimators. *Journal of the American Statistical Association*, 82(399):929–937, 1987. (page 6, 18, 25)

Riccardo Griggio. Tassi di feconditá per etá del costa rica.unanalisi bayesiana. 2013/2014. (page 48, 69, 102)

Arjun K Gupta and John T Chen. A class of multivariate skew-normal models. *Annals of the Institute of Statistical Mathematics*, 56(2):305–315, 2004. (page 101)

Arjun K Gupta, Truc T Nguyen, and Jose Almer T Sanqui. Characterization of the skew-normal distribution. *Annals of the Institute of Statistical Mathematics*, 56(2):351–360, 2004. (page 101)

Eshetu Gurmu. *Fertility transition driven by poverty: the case of Addis Ababa (Ethiopia)*. PhD thesis, University of London, 2005. (page 17)

Eshetu Gurmu and Ruth Mace. Fertility decline driven by poverty: the case of Addis Ababa, Ethiopia. *Journal of biosocial science*, 40(03):339–358, 2008. (page 17)

Hugo Hadwiger. Eine analytische reproduktionsfunktion für biologische gesamtheiten. *Scandinavian Actuarial Journal*, 1940(3-4):101–113, 1940. (page 13, 101)

Hilde Haider and Peter A Frensch. Why aggregated learning follows the power law of practice when individual learning does not: Comment on rickard (1997, 1999), delaney et al.(1998), and palmeri (1999). 2002. (page 64)

H Leon Harter and Albert H Moore. Point and interval estimators, based on m order statistics, for the scale parameter of a weibull population with known shape parameter. *Technometrics*, 7(3):405–422, 1965. (page 87)

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. (page 36, 44)

Andrew Heathcote, Scott Brown, and DJK Mewhort. The power law repealed: The case for an exponential law of practice. *Psychonomic bulletin & review*, 7(2):185–207, 2000. (page 64)

Philip Heidelberger and Peter D Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983. (page 55, 90)

Norbert Henze. A probabilistic representation of the'skew-normal'distribution. *Scandinavian journal of statistics*, pages 271–275, 1986. (page 29, 48, 79)

Jan M Hoem and Bo Rennermalm. *On the statistical theory of graduation by Splines*. Laboratory of Actuarial Mathematics, University of Copenhagen, 1977. (page 101)

Jan M Hoem, Dan Madien, Jørgen Løgreen Nielsen, Else-Marie Ohlsen, Hans Oluf Hansen, and Bo Rennermalm. Experiments in modelling recent danish fertility curves. *Demography*, 18(2):231–244, 1981. (page 13, 14, 16, 101)

C Jennison. Contribution to the discussion of paper by richardson and green (1997). *Journal of the Royal Statistical Society, series B*, 59:778–779, 1997. (page 41)

Robert E Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996. (page 40)

Y Kinfu. *The quiet revolution: An analysis of change toward below-replacement level fertility in Addis Ababa*. PhD thesis, PhD Thesis, Australian National University, 2001. (page 3, 17)

Ruth King, Byron Morgan, Olivier Gimenez, and Steve Brooks. *Bayesian analysis for population ecology*. CRC Press, 2009. (page 89)

Anne B Koehler and Emily S Murphree. A comparison of results from state space forecasting with forecasts from the makridakis competition. *International Journal of Forecasting*, 4(1):45–55, 1988. (page 21)

Heemun Kwok and Roger J Lewis. Bayesian hierarchical modeling and the integration of heterogeneous information on the effectiveness of cardiovascular therapies. *Circulation: Cardiovascular Quality and Outcomes*, 4(6):657–666, 2011. (page 69)

Wallace E Larimore and Raman K Mehra. Problem of overfitting data. *Byte*, 10(10): 167–178, 1985. (page 21)

AM Latimer, S Banerjee, H Sang Jr, ES Mosher, and JA Silander Jr. Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern united states. *Ecology Letters*, 12(2):144–154, 2009. (page 26)

Michael D Lee and Michael R Webb. Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, 12(4):605–621, 2005. (page 69)

Sik-Yum Lee. *Structural equation modeling: A Bayesian approach*, volume 711. John Wiley & Sons, 2007. (page 19)

Tsung I Lin. Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, 100(2):257–265, 2009. (page 100)

Tsung I Lin, Jack C Lee, Shu Y Yen, et al. Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, 17(3):909, 2007. (page 101)

David P Lindstrom and Zewdu Woubalem. The demographic components of fertility decline in addis ababa, ethiopia: a decomposition analysis. *Genus*, pages 147–158, 2003. (page 17)

Brunero Liseo and Nicola Loperfido. A bayesian interpretation of the multivariate skew-normal distribution. *Statistics & probability letters*, 61(4):395–401, 2003. (page 30, 101)

David J Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337, 2000. (page 26)

Scott M Lynch. *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer Science & Business Media, 2007. (page 32)

E Mumbi Machera. Social, economic and cultural barriers to family planning among rural women in kenya: A comparative case study of the divisions in embu district. 1997. (page 3)

Kuttan Mahadevan. *Sociology of fertility: Determinants of fertility differentials in South India*. Sterling, 1979. (page 17)

Peter Martin. Une application des fonctions de gompertz a l'etude de la fecondite d'une cohorte. *Population*, 22(6):1085–1096, 1967. (page 15)

Stefano Mazzuco and Bruno Scarpa. Fitting age-specific fertility rates by a skew-symmetric probability density function. 2011. (page 6, 100, 101)

Stefano Mazzuco and Bruno Scarpa. Fitting age-specific fertility rates by a flexible generalized skew normal probability density function. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):187–203, 2015. (page 6, 15, 16, 18, 25, 26, 29, 44, 88, 102)

Michael A McCarthy. *Bayesian methods for ecology*. Cambridge University Press, 2007. (page 26)

Donald R McNeil, T Jamel Trullell, and John C Turner. Spline interpolation of demographic oata. *Demography*, 14(2):245–252, 1977. (page 16)

Demena Melake. *Population and Development: Lecture note*. HaramayaUniversity in collaboration with the Ethiopia Public Health Training Initiative, The Carter Center, the Ethiopia Ministry of Health, and the Ethiopia Ministry of Education, 2005. (page 1)

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953. (page 36, 44)

Russell B Millar. Reference priors for bayesian fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(9):1492–1502, 2002. (page 26)

Marco Minozzo and Laura Ferracuti. On the existence of some skew-normal stationary processes. *Chilean Journal of Statistics*, 3:157–170, 2012. (page 101)

Anteneh Mulugeta Eyasu. Multilevel Modeling of Determinants of Fertility Status of Married Women in Ethiopia. *American Journal of Theoretical and Applied Statistics*, 4 (1):19–25, 2015. (page 4)

Edmund M Murphy and Dhruva N Nagnur. A gompertz fit that fits: applications to canadian fertility patterns. *Demography*, 9(1):35–50, 1972. (page 15)

Håkan Nilsson, Jörg Rieskamp, and Eric-Jan Wagenmakers. Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55 (1):84–93, 2011. (page 8, 63, 66, 67, 68, 69, 89)

NOP. National Population Office. National Population Policy of Ethiopia, Addis Ababa. *Office of the Prime Minister*, 1993. (page 100)

T Ozaki. On the order determination of arima models. *Applied Statistics*, pages 290–301, 1977. (page 21)

Samba SR Pasupuleti and Prasanta Pathak. Special form of gompertz model and its application. *Genus*, 66(2):95–125, 2010. (page 13, 101)

Paraskevi Peristera and Anastasia Kostaki. Modeling fertility in modern populations. *Demographic Research*, 16(6):141–194, 2007. (page 14, 16, 101)

Adrian E Raftery, Steven Lewis, et al. How many iterations in the gibbs sampler. *Bayesian statistics*, 4(2):763–773, 1992. (page 55, 90)

Christian P Robert. The bayesian choice: From decision-theoretic foundations to computational implementation (springer texts in statistics) by. 2001. (page 69)

Jeffrey N Rouder and Jun Lu. An introduction to bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4): 573–604, 2005. (page 68)

Jeffrey N Rouder, Jun Lu, Richard D Morey, Dongchu Sun, and Paul L Speckman. A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, 137(2):370, 2008. (page 68)

J Andrew Royle, K Ullas Karanth, Arjun M Gopalaswamy, and N Samba Kumar. Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, 90(11):3233–3244, 2009. (page 93)

M Schaub and M Kéry. Combining information in hierarchical models improves inferences in population ecology and demographic population analyses. *Animal Conservation*, 15 (2):125–126, 2012. (page 67)

Benjamin Scheibehenne and Thorsten Pachur. Hierarchical bayesian modeling: Does it improve parameter stability? In *CogSci 2013: 35th Annual Conference of the Cognitive Science Society*, pages 1277–1282. Cognitive Science Society, 2013. (page 65, 68)

Benjamin Scheibehenne and Bettina Studer. A hierarchical bayesian model of the influence of run length on sequential predictions. *Psychonomic bulletin & review*, 21(1):211–217, 2014. (page 68)

Carl P Schmertmann. A system of model fertility schedules with graphically intuitive parameters. *Demographic Research*, 9(5):81–110, 2003. (page 16, 101)

Rajendra K Sharma. *Demography and population problems*. Atlantic Publishers & Dist, 2004. (page 1)

Richard M Shiffrin, Michael D Lee, Woojae Kim, and Eric-Jan Wagenmakers. A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive Science*, 32(8):1248–1284, 2008. (page 66, 68)

Susan E Short and Gebre-Egziabher Kiros. Husbands, wives, sons, and daughters: Fertility preferences and the demand for contraception in ethiopia. *Population Research and Policy Review*, 21(5):377–402, 2002. (page 17)

Amson Sibanda, Zewdu Woubalem, Dennis P Hogan, and David P Lindstrom. The proximate determinants of the decline to below-replacement fertility in addis ababa, ethiopia. *Studies in family planning*, 34(1):1–7, 2003. (page 17)

Brijesh P Singh, Kushagra Gupta, and KK Singh. Analysis of fertility pattern through mathematical curves. *American Journal of Theoretical and Applied Statistics*, 4(2):64–70, 2015. (page 17)

Brian J Smith. Bayesian output analysis program (boa) for mcmc. *R package version*, 1 (5), 2005. (page 90)

Tom AB Snijders and Roel Bosker. J. 1999: Multilevel analysis. an introduction to basic and advanced multilevel modeling. (page 26)

David J Spiegelhalter, Keith R Abrams, and Jonathan P Myles. *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons, 2004. (page 26, 75)

Martin A Tanner and Wing Hung Wong. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540, 1987. (page 44)

Hailemariam Teklu, Alula Sebhatu and Tesfayi Gebreselassie. Components of Fertility Change in Ethiopia. Further Analysis of the 2000, 2005 and 2011 Demographic and Health Surveys. *DHS Further Analysis Reports.No. 80. Calverton, Maryland, USA: ICF International*, 2013. (page 2)

Charles Teller and Assefa Hailemariam. *The Demographic Transition and Development in Africa*. Springer, 2011. (page 4)

y The World Factbook. https://www.cia.gov/Library/publications/the-world-factbook/geos/print/country/countrypdf_et.pdf. 2014. (page 2)

Howell Tong. Some comments on the canadian lynx data. *Journal of the Royal Statistical Society. Series A (General)*, pages 432–436, 1977. (page 21)

Don van Ravenzwaaij, Gilles Dutilh, and Eric-Jan Wagenmakers. Cognitive model decomposition of the bart: assessment and application. *Journal of Mathematical Psychology*, 55(1):94–105, 2011. (page 68, 69)

John Von Neumann. 13. various techniques used in connection with random digits. 1951. (page 33)

Christopher K Wikle. Hierarchical models in environmental science. *International Statistical Review*, 71(2):181–199, 2003. (page 69)

Guillaume Wunsch. Courbes de gompertz et perspectives de fécondité. *Recherches Économiques de Louvain/Louvain Economic Review*, pages 457–468, 1966. (page 15)

Arnold Zellner and Peter E Rossi. Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, 25(3):365–393, 1984. (page 18)

Haftu Gebrehiwot

Curriculum Vitae:

Tiepolo 54
35121 Padova
+39 (345) 513 1871
☎ +39 (345) 513 1871
✉ gebrehiwot.stat@unipd.it
haftug2000@yahoo.com

"Success is the ability to go from failure to failure without losing your enthusiasm." – Winston Churchill

---

Personal Information

Birth Date: 01/10/1979
Birth place: Maichew, Ethiopia
Marital Status: Single
Nationality: Ethiopian
Current Address: University of Padova , Department of Statistics, Italy
E-mail: haftug2000@yahoo.com/ gebrehiwot.stat@unipd.it
Cell Phone: 00393455131871
Skype Address: haftug.gebrehiwot

Languages Known

Languages: Amharic (first language)
: English (fluent)

Educational Background

2013-2016: University of Padova,Department of Statistics, Italy.
Qualification: Ph.D in Statistics **Expected to obtain:** March, 2016
Thesis Topic: Implementing Hierarchical Bayesian Modeling to Fertility Data:The case of Ethiopia
Advisor: Prof Stefano Mazzuco **E-mail:** mazzuco@unipd.it

2008-2009: Addis Ababa University, Department of Statistics, Ethiopia.
Qualification: M.Sc. in Applied Statistics
Thesis Topic: Correlates of profitability of Large and Medium Scale Manufacturing industries in Ethiopia: An evidence from panel data regression analysis.
Advisor: Prof Amanual Gebreyohanus **E-mail:** amanualgy@gmail.com

1999-2003: Addis Ababa University, Department of Mathematics, Ethiopia.
Qualification: B.Sc. in Pure Mathematics (major in Mathematics; minor in statistics)
Thesis Topic: Evedence of Longitudunal Data Analysis
Advisor: Prof Estefano mazzuco **E-mail:** haftug.gebrehiwot

Teaching Experience

2009-2012: As Lecturer.
Institution: Mekelle University, Department of Statistics, Ethiopia
P.O.Box: 231
Fax: +251-0344-4-409304
Telephone: +251-0344-4-409017
E-mail: mekelle.Unversity@telecom.net.et

Instructor: Introduction to Statistics, Basic Statistics, Probability and Statistics, Biometry and Experimental Statistics, Multivariate Analysis, Statistical Computing with SPSS and Minitab, Statistical Computing with SAS and R, Statistical Computing with STATA, Fundamental of Biostatistics, Statistics for Geologists, Design and Analysis of Experiments ,Statistical Consultancy and Research Methods

Teaching Assistant: Research Methods(for Engineering) and Advanced Biostatistics (for Biology)

2003-2007: **As Assistant Lecturer**.

Institution: Maichew Technical College, Department of Mathematics, Ethiopia

P.O.Box: 29

Fax: +251-0344-7-770550

Telephone: +251-0344-7-770268

E-mail: mtc@ethionet.et

Instructor: Introduction to Statistics, Preliminary Mathematics , First course in Calculus, Fundamental Concepts of Algebra

Additional Experience

2009-2012: **As Data Analyst**.

♠: Assessment of Dot implementation in Tigray, Seid Ali Yassin, MSc student, AAU, school of Pharmacy

♠: Land use dynamics and its implication to the livelihoods of rural households in Tigray, MSc student, rural development, FDANR, department of NREM, Mekelle university

♠: **Factors Affecting the Sprinter's Performance, at Tirunesh Dibaba National Athletics Training Center**,Samson W., Woldegebreal M., Prof.Dr.Hasrani, Department of Sport Science, Mekelle University,Ethiopia

2009-2012: **As Consultant**.

♠: Using SPSS in condominium house allocation to resident of Mekelle city, Ethiopia

♠: Detailed study of MSEs(Micro & Small Enterprises) and unemployments in Mekelle city

2009-2012: **Preparation teaching & training material**.

♠: **Introduction to statistics**, *Mekelle University, P.O. Box 231, Fax: 251-0344-4-409304,Tel:+251-0344-4-409017, E-mail: mekelle.Unversity@telecom.net.et.*

♠: **Introductory Probability**, *Published:Mekelle University, P.O. Box 231, Fax: 251-0344-4-409304,Tel:+251-0344-4-409017, E-mail: mekelle.Unversity@telecom.net.et.*

♠: **Introduction to Statistical Computing with SPSS & SAS**, *Published:Mekelle University, P.O. Box 231, Fax: 251-0344-4-409304,Tel:+251-0344-4-409017, E-mail: mekelle.Unversity@telecom.net.et.*

2009-2012: **As a trainee:** giving short training for for both staffs and outside communities engaged directly or indirectly in research activities in:, .

♠: in statistical computing (such as SPSS, SAS,STATA,R, etc)

♠: in statistical analysis, and in the interpretation of findings, etc

2009-2012: **As a researcher and research member**.

As PI: Empirical Analysis On Traffic Accidents in Tigray Region

As Member: Assessment of Quality of Education in High Schools, Tigray Region

As Member: Depredation of Livestock by Large Carnivores in the Northern Ethiopian Highlands

PhD School Courses and Seminars

Courses:
- ♠ Functional Analysis
- ♠ Theory and Methods of Statistical Inference
- ♠ Probability Theory
- ♠ An Overview Objective Bayesian Analysis

- ♠ Bayesian Inference
- ♠ Statistical Models
- ♠ Applied Multivariate Techniques
- ♠ Generalized linear mixed models
- ♠ Panel Data Analysis
- ♠ Measurement Error
- ♠ Spatial Statistics
- ♠ Sampling Theory
- ♠ Compressed Sensing Bayesian Analysis
- ♠ Nonparametric Smoothing Techniques
- ♠ Analysis of Survival Data
- ♠ Time Series Analysis
- ♠ From Frailty to Multiprocess Models
- ♠ Introduction to Robust Statistics
- ♠ Statistical Analysis Social Networks
- ♠ The Interface between R and C

(Note: For more, please see the School web page http://www.stat.unipd.it/phd/courses)

Some of Seminars: ♠ Added Predictive Value of Omics Data;

♠ Estimating Causal Effects in Gene Expression from a Mixture of Observational and Intervention Experiments;

♠ A Semi-Parametric Method for Robust Multivariate Skewed Error ;

♠ Detection in Functional Data with Application to Historical Radiosonde Winds ;

♠ Penalized Model Selection in Graphical Models ;

♠ Knowing the dynamics of employment the employee availability and accessibility of data bases obtained from the disclosures required of businesses ;

♠ Nonparametric Multivariate Regression Smoothers ;

♠ Multidimensional Inequality Measures on Finite Partial Orders ;

♠ Energy markets, and quantitative methods: a bridge between Universities and Enterprises ;

♠ Modelling Complex Associations in Survival Models with Application to Nutrition of Critically ill Patients ;

♠ Introduction to Experimental Designs; Hypothesis Test in Mixture Designs, etc.

(Note: For more, please see the School web page http://www.stat.unipd.it/phd/courses)

Honors AND Awards

SPSS: Certificate on SPSS TRAINING

STATA: Certificate on STATA TRAINING

SAS: Certificate on SAS TRAINING

RPW & QDA: Certificate on RESEARCH PROPOSAL WRITING and QUANTITATIVE DATA ANALYSIS

Basic Computer: Certificate on INTRODUCTION TO COMPUTER(Window, Word, Excel, Access)

VB Programming: Certificate on VISUAL BASIC PROGRAMMING Courses

Trainer: Certificate of Trainer on statistical software (SPSS, SAS and R) offered by Mekelle University, College of Natural and Computational Science

Research Interest

R.Interest:
- ♠ Functional Analysis
- ♠ Econometrics Analysis
- ♠ Hierarchical Analysis
- ♠ GLM
- ♠ Survival Analysis & Frailty model
- ♠ Longitudinal and Multivariate data analysis
- ♠ Bayesian Analysis
- ♠ Nonparametric Inference

Conference Presentation

11-13/06/ 2014: SIS Scientific Meeting 2014, Cagliari, Italy.

Topic: Modeling Age-Specific fertility rate in Ethiopia

Tiepolo 54 – 35121 Padova
+39 (345) 513 1871 • ☎ *+39 (345) 513 1871* • ✉ *gebrehiwot.stat@unipd.it*
haftug2000@yahoo.com

Pres. Type: Oral

17-19/09/ 2014: **2nd Bayesian Young Statisticians Meeting 2014(BAYSM 2014), Vienna, Austria** .

Topic: Implementing Bayesian Modeling to Fertility Data: The case of Ethiopia

Pres. Type: Poster

4-6/02/ 2014: **XI-Population Days 2015 (PopDay2015), Palermo, Italy** .

Topic: Modeling Age-Specific fertility rate in Ethiopia

Pres. Type: Oral

20-22/08/ 2015: **Harnessing the Power of Data in Ethiopia, 24th ESA Conference, Ethiopia** .

Topic: Implementing Hierarchical Bayesian Modeling to Fertility Data: The case of Ethiopia

Pres. Type: Oral

Publications (Peer-Reviewed)

=> **The Ecology of Large Carnivores in the Highlands of Northern Ethiopia** ,*Gidey Yirga, Hans H. de Iongh , Herwig Leirs, Kindeya Gebrehiwot,Gebrehiwot Berhe, Tsehaye Asmelash, Haftu Gebrehiwot and Hans Bauer*, http://hdl.handle.net/1887/22747

=> **Correlates of Profitability of Large and Medium Scale Manufacturing Industries in Ethiopia: An Evidence from Panel Data Analysis** ,*Haftu Gebrehiwot*, http://www.ethstat.org.et/proceedings19ac.pdf

=> **Modeling Age-Specific fertility rate in Ethiopia**,*Haftu Gebrehiwot and Stefano Mazzuco*(Accepted)

=> **Implementing Bayesian Modeling to Fertility Data: The case of Ethiopia**,*Haftu Gebrehiwot and Stefano Mazzuco*(Accepted)

=> **Implementing Hierarchical Bayesian Modeling to Fertility Data: The case of Ethiopia**,*Haftu Gebrehiwot and Stefano Mazzuco*(Accepted)

Professional Memberships

ESA: Ethiopian Statistical Association(ESA)

EAE: Economic Association of Ethiopia (EAE)

SIS: The Italian Statistics Society (SIS)

References

References: Available upon request.