UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXX

Projecting Pakistan Population with a Bayesian Hierarchical approach

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Stefano Mazzuco

Dottorando: Muhammad Adil

31st October, 2017

Abstract

UN has generated projections for demographic components for all countries of the world deterministically before revising the methodology and recently started to generate probabilistic projections at national level , but regional/sub national level trajectories are yet to be generated. In this thesis, probabilistic projections for total fertility rate (TFR) , life expectancy at birth for males and females and population totals at national as well as regional level have been generated based on Bayesian Hierarchical modeling approach. The trajectories were also generated for variable number of countries and it was observed that continent based trajectories were enough to model the future pattern instead of going for entire globe data. This in result saves time and decreases the influence of developed world demographic pattern on developing world and vise versa. TFR results were compared based on different values of μ, (i.e., values less than and more than 2.1, which is the ultimate level of replacement) and with the trajectories generated based on Bayesian Hierarchical modeling approach for phase III. For values greater than 2.0, the trajectories were taking long time to converge to replacement level than for smaller values. On the other hand trajectories of Life Expectancy have revealed significant gap between male and female projections at national as well as regional level. It was also observed that Balochistan region has lower life expectancy at birth than the rest of the regions of Pakistan. The trajectories for Population totals were generated based on the probabilistic TFR and life expectancy trajectories. The results revealed a total population of over 207 million at national level which nearly coincides with the

recently released provisional figures of census. Furthermore, the results are significantly different from previously adopted deterministic results and could be a good substitute over classical deterministic approaches.

Sommario

Le proiezioni demografiche delle Nazioni Unite per tutti i paesi del mondo sono sempre state ottenute attraverso un approccio puramente deterministico, e solo recentemente approcci probabilistici sono stati considerati per i livelli nazionali, mentre le traiettorie regionali o sub nazionali non sono ancora state considerate. In questa tesi, le proiezioni probabilistiche del tasso di fecondità, dell'aspettativa di vita alla nascita per maschi e per femmine e il totale delle popolazioni sia a livello nazionale che a livello regionale sono state generate attraverso l'uso di modelli Bayesiani gerarchici. Le traiettorie sono state sviluppate per diverse variabili numeriche specifiche di ogni paese dalle quali è emerso che le informazioni relative al continente risultano più efficienti rispetto a quelle globali per modellare l'andamento futuro dei fenomeni di interesse. Questo approccio permette sia di accelerare la procedura di stima del modello, sia di ridurre l'impatto legato alla variabilità degli eventi osservati su tutto il globo. I risultati per il tasso di fecondità globale sono stati confrontati rispetto a differenti livelli di μ (e quindi minori o maggiori di 2.1, che è l'ultimo valore considerato) e rispetto a diverse traiettorie generate dal modello gerarchico in fase III. Per valori maggiori di 2.1, le traiettorie hanno richiesto un tempo computazionale molto lungo per convergere al livello prestabilito rispetto a valori più piccoli. D'altro canto, le traiettorie per l'aspettativa di vita hanno rivelato un divario significativo tra i maschi e le femmine sia a livello nazionale che a livello regionale. E' stato osservato inoltre che nella regione del Balochistan l'aspettativa di vita alla nascita è pari almeno a quella del resto del paese. Le traiettorie per il totale

delle popolazioni sono state generate basandosi sul tasso di fecondità globale e sulle aspettative di vita generate dal modello Bayesiano. I risultati hanno mostrato una popolazione totale di 207 milioni di persone per il livello nazionale, livello che coincide approssimativamente con i recenti e provvisori dati del censimento. I risultati ottenuti sono inoltre significativamente differenti da quelli fino ad ora ottenuti, ed evidenziano la qualità del nuovo approccio proposto come alternativa ai metodi deterministici fino ad ora utilizzati.

Dedicated to my family

&

my supervisor

Acknowledgements

First of all I would like to express my greatest thanks to almighty Allah, the one and only supreme power, the most merciful, worthy of all praise, the most beneficent who always helps me in difficulties, guides me to the best possible solutions and gives me strength to complete every task.

I would like to express my greatest gratitude to my supervisor Professor Stefano Mazzuco from the core of my heart for his support, wisdom, encouragement and precious advice and guidance throughout this journey of research. I would specially mention his always welcoming behaviour even without taking prior appointments. He was always there to support me in tough and rough times. I would extend my thanks to Professor Monica Chiogna who as a course coordinator was a big support during first year of PhD. I would like to thank Professor Nicola Sartori who during the last few months of our research journey helped in resolving administrative issues. Special thanks go to all teachers who provided us with enough knowledge to successfully accomplish this task. I would like to specifically acknowledge Patrizia Piacentini who tried her best to solve any problem which we faced from the start till the end of PhD period.

I must thank Research Scientist Hana Sevcikova , University of Washington, who always responded to my emails and guided me in BayesPop R package.

I would like to thank my friends in Pakistan Bureau of Statistics specifically Mr. Syed Jawad Ali Shah, who guided me on dealing with glitches relevant to data and always encourages and motivates. I would like to thank Mr. Ahtasham Gul who at all times provided me assistance on technical and bureaucratic issues. I thank my friend S.O. Najeeb Ullah who was always there to help me in getting data from Pakistan Bureau of Statistics. I also thank Professor Qamruzaman, University of Peshawar, for his support and guidance on numerous occasions.

I particularly thank my friend Ismail Shah, who guided, motivated and encouraged me to apply for this opportunity and introduced me to the department of Statistics, University of Padova. I thank my friends Danish Wasim and Kaleem Ullah Shahzad for their motivation and encouragement.

I thank my colleagues from the 30th PhD cycle and feel very lucky to have shared these three years of ups and downs with such an incredible group of people.

I feel Padova as a home due to a bunch of best people i met during my this journey, they include but not limited to Dr.Ali Raza, Mubashar, Dr.Jalal Uddin, Dr.Saeed Khan, Dr.Jagjit Singh, Dr.Ishtiaq, Dr.Saima Imran, Imran bhai and Zaheer bhai.

I especially express my thanks to my wife whose love and support always provided me with strength and feeling of belief in myself towards finishing this task successfully. I must acknowledge my lovely daughter Rahemeen whose pure love and affection always motivated me to keep moving forward. I thank my sister and brother for their unconditional love. I wish to express my love and great respect for the support and unlimited love of my parents who are always a great source of motivation.

Contents

List of Figures

List of Tables

Introduction

Overview

Population count remained a big challenge for Pakistan, which can be revealed from the fact, that since the birth of country in 1947, no regular fixed scheduled based census have been conducted in the past three decades. The series of census conducted are in the years 1951, 1961, 1972, 1981, 1998 and last one just conducted few months back in April to June, 2017, whose finalized results are under process and may be disclosed in the next year. In this connection attempt has been made to have future estimates of population counts along with essential demographic estimates of fertility and mortality using probabilistic approach. As one of the most important and appealing corner in demographic research is the projection of population and its main characteristics. Many deterministic approaches have been adopted in the past to project populations by extrapolating mortality, fertility and migration rates. But need is felt to incorporate the probabilistic approach due to the uncertainty patterns in the population, as said by Keyfitz (1972) uncertainty is all around population estimates. Lee and Carter (1992) & Lee and Tuljapurkar (1994) introduced projection approach based on extrapolation of time series parameters and projected mortality trends which is useful for evaluating change and deviations of mortality effects. Nevertheless, this approach has some concerns, e.g., it is claimed that the procedure of projection does not report the uncertainty of confidence intervals and of population estimates in an integrated approach (Bijak, 2011). Thus need is felt to have some tool which could work both with uncertainty and subjective approach, as a result Bayesian statistics is introduced which gained a lot of attention due to its nice properties and prediction capabilities (Alkema *et al.*, 2011; Bijak, 2010; Girosi and King, 2008). In this regard, there is intensifying belief in population forecasting experts that the upcoming period belongs to Bayesian probabilistic predictions (Alkema *et al.*, 2011; Bijak, 2010; Girosi and King, 2008). The use of Bayesian demography is gaining importance because it can address three main problems very easily, i.e., demographic forecasts, limited data, and highly structured or complex models (Bijak and Bryant, 2016).

The importance of probabilistic approach could be judged from the fact that UN has started adopting this approach for projection of population and its main characteristics United Nations and Social Affairs (2017). Keeping in mind the importance of probabilistic approach, we tried to develop a Bayesian Hierarchical Model to project the future population along with projections for fertility and mortality rates. The model developed for the said purpose by Raftery *et al.* (2012) have been used for projection of population for all countries of the world. However, there is still need to have results at sub-national/regional level. In this thesis we tried to modify the model developed to cope with the regional level situation. The regional level forecast obtained by this approach would be helpful for policy makers as they could build their plans by considering uncertainty level of estimates, which would definitely be of good assistance for better planning.

Main contributions of the thesis

We obtained sub-national/regional level forecasts for fertility, mortality and population totals which are not obtained previously by these approaches. We made some adjustments in the models developed for projection of fertility rate (Alkema *et al.*, 2011) and life expectancy (Raftery *et al.*, 2013, 2014b) together with population projection model (Raftery *et al.*, 2012). Instead of using the default values of prior distributions in the case of above models, we used values based on the pattern of regional level estimates. The model used for projection of TFR, life expectancy and population totals (Alkema *et al.*, 2011; Raftery *et al.*, 2012, 2013, 2014b) takes into consideration all countries data which helps in providing good estimates, but on the other hand this complicates the process of estimation, since much more time is desired for computational process.

We tested the results for our country and regions by keeping in mind different combinations of variable number of countries by taking into consideration the data from i) Pakistan and its regions, ii) Southern Asian countries, iii) Asian countries and iv) Entire world. We found results for case three were almost similar to what we obtained for case four. This made us believe, instead of using entire globe data, it is better to use the data of that part of globe only where the country geographically exist. Bayesian Hierarchical model (BHM) is used because we don't have enough data to support our analysis. BHM has nice property of borrowing information from neighboring countries to help get better estimates for the desired country, but on the other hand it will influence the behaviour of estimates due to the borrowing information. To overcome this we suggest to limit the borrowing information to the most nearly bordered countries. As a result it will remove the effect of developed worlds pattern on the developing world. This as a result increases the speed of simulation process and saves time. We also tried to see the effect of different replacement levels on the fertility model for phase III by considering classical time series auto-regressive AR(1) approach where we were free to choose replacement level other than what has been generalized for all countries. We compared classical approach results with Bayesian Hierarchical model, where the model was free to choose country specific value based on data pattern, and found the two approaches provide quite different result.

One big challenge we faced during the modeling phase was to have regional level data, but it was not easily available in similar format as being reported by UN in the world population prospects (United Nations and Social Affairs, 2015). This is because they have full access to all legal departments and handful of experts who could manage to fix any problem which they may be facing based on data quality/availability. We tried to take help from the methodologies of UN on fixing issues relevant to regional level estimates (United Nations and Social Affairs, 2006, 2015, 2017). We kept national level data as it is reported in world population prospect (United Nations and Social Affairs, 2015, 2017) and computed regional level estimates to meet our needs. Without reliable data it is not possible to have fruitful results, so to meet these standards we tried to cross validate our estimates with different published results and were able to meet the data pattern as reported in world population prospects (United Nations and Social Affairs, 2015, 2017).

In summary our contributions can be summarized as:

- We produced probabilistic projections for TFR at regional level.
- We suggested the number of countries to be considered in the model in order to reduce the influence of borrowing information in BHM and to save the computation time.
- We produced probabilistic projections for Life Expectancy at birth for males and females at regional level.
- We produced probabilistic projections for Population total at regional level.
- We computed regional level estimates from different data sources which were not previously reported anywhere else.

Chapter 1

Methodology

1.1 Fertility

It is the natural ability to produce offspring. It is also defined as the total number of children a woman could have during her reproductive age span in a specified period and with respect to a specified population (Yusuf *et al.*, 2014). There exist a number of ways to define fertility measures, like Crude birth rate, General Fertility rate, Age specific fertility rates, and Total fertility rate.

Fertility is the main component responsible for the changes in the size and structure of population (Alkema *et al.*, 2011).

1.1.1 Age Specific Fertility Rate

It measures the annual number of births to women of a specified age or age group per 1,000 women in that age group, with respect to a specified reference period (United Nations and Social Affairs, 2009).It is a useful measure in translating the relationship between age and fertility (Yusuf *et al.*, 2014). Mathematically

$$f^i_{t\to t+1} = \frac{B^i_{t\to t+1}}{W^i_{t+0.5}} \tag{1.1}$$

where $f^i_{t\to t+1}$ is the age-specific fertility rate at age or age group i during the year t, $B_{t\to t+1}$ represents the number of live births during the same period and $W_{t+0.5}$ is the mid year female population in the same age or age group i. The reproductive age usually taken to be between 15 to 49, however, it is possible for women to give birth out side this interval but their number is very low (Yusuf *et al.*, 2014).

1.1.2 Total Fertility Rate

it is the average number of children a woman would bear, if she survived through the end of the reproductive age span, experiencing at each age the age-specific fertility rates of that period and assuming no mortality (Alkema *et al.*, 2011) (Yusuf *et al.*, 2014). Mathematically

$$TFR = \Sigma ASFR$$

OR

$$TFR_{t\to t+1} = \sum_{i=15}^{49} f^{i}_{t\to t+1} \tag{1.2}$$

where $f^{i}_{t\to t+1}$ is the age-specific fertility rate for women aged i at time point t. According to the recent Demographic Health Survey (NIPS, 2012), the TFR in Pakistan and its regions showed a decline from 5.4 in 1990-91 to 3.8 in 2012-13. The maximum TFR observed was in Balochistan region 4.0 while the lowest was in the capital 3.0.

1.2 Fertility Transition

There exist many theories on fertility transition. The basic idea as explained by Mason (1997) is the long term decline in the number of children from four or more per woman to two or fewer. According to Mason (1997), fertility transition theories can be used on three scales, for example, on a millennial time scale, in which the focus of fertility decline is on two hundred years, the second approach is based on a centennial scale, which is based on the question of why fertility declines, and finally on a decadal scale, which is based on the idea of ten years fertility decline. The final concept of decadal scale was introduced in Princeton's European Fertility Project (Coale and Watkins, 1992). Mason (1997) suggest that there are various reasons for fertility transition that includes, institutional, cultural, and environmental conditions that motivate child prevention. Also mortality decline is necessary but not sufficient condition for fertility decline. Casterline (2001) described four key determinants of the pace of fertility decline that are dominated in literature. They are a) the pace of social and economic change, b) the pace of change in economic aspirations and expectations c) the pace of improvement in the provision of birth control services and d) the pace of reduction in the moral and social costs of birth control. He further argued that the pace of fertility decline will have significant role in determining the population size and its regional distribution.

However, whatever the reasons may be, fertility decline can be viewed in three phases generally (Alkema *et al.*, 2011). Phase 1, which is defined as the pre- transition or high fertility phase. Phase 2, the Fertility transition phase and Phase 3, post transition or low fertility phase.

The scenario of Pakistan has an interesting story. According to Sathar (2001) fertility decline in the country has shown significant resistance to change over the past decades. In 1960, the population of Pakistan was growing at the rate of 2.7 percent per year and to overcome this high rate of growth, family planning program was introduced then, but did not show significant impact, despite the fact the growth rate crossed 3 percent per year. However, it has shown a decline and resulted in decline of fertility rate from over 6 to below 4 in the recent studies. Sathar (2001) further pointed out; although most of South Asian countries revealed high fertility and mortality rates, but only some predicted demographic diversity and achieved. India and Srilanka pioneered the fertility transition while Pakistan was the largest block to initiate fertility transition in early 90's (Sathar, 2001).

1.2.1 Phases of Fertility Transition

The start and end of fertility transition has been debated for a long time. In the past, in many of the previous studies it was believed that fertility decline starts when there is a decline of 10 percent below the maximum observed TFR which was criticized by Casterline (2001) and thus Bongaarts (2002) suggested a bench mark for fertility transition to be 5 percent instead of 10 percent. However, Alkema *et al.* (2011) argued it to be stringent, and gave an example of Mozambique by considering the data of UN world population prospects and argued that on the basis of approach mentioned by Bongaarts (2002) the fertility transition has started in 1995-2000, but in reality it has occurred much earlier than this. He further defined the three phases of fertility transition as under:

- **Phase 1** is defined as the stable pre-transition high-fertility phase; the fertility transition has not started yet, and fertility fluctuates around high TFR levels (e.g., around a TFR of about 6 or 7 children).

- **Phase 2** is the period where fertility transition has started and is defined as the most recent period with a local maximum within 0.5 child of the global maximum. If this local maximum is more than 5.5, the corresponding period is defined as the start period of the fertility transition, denoted by τ_c. On the other hand if it is below 5.5, then higher TFR levels have most likely been observed before the beginning of the observation period, in that case $\tau_c <$ 1950–1955.

The reason to mention 1950-55 is because the data considered for current study starts from this period.

- **Phase 3** The start of phase 3 is defined as the period where a two consecutive increments below a TFR of 2 children have been witnessed. The start of phase 3 denoted by λ_c for a country or region c, is defined to be in the middle of the (earliest) two subsequent increments below 2. More specifically, with $f_{c,t}$ the TFR in country c in period t, the start of Phase 3 is the earliest period t for which the following conditions are met. (1) the TFR increased from period $t-1$ to period t, so that $f_{c,t} > f_{c,t-1}$; (2) the TFR increased again from period $ttot+1$, so that $f_{c,t+1} > f_{c,t}$; and (3) the TFR was below 2 in all three periods, so that $f_{c,p} < 2$ for periods $p = t-1$, t, $t+1$ (Alkema *et al.*, 2011).

1.2.2 UN Methodology on TFR Projection

The Population Division of the Department of Economic and Social Affairs of the UN provides accessible population data for all countries and areas of the world. The department also produces projections on TFR that are updated and published in the world population prospects(WPP) (United Nations, Department of Economic and Social Affairs, Population Division 2009). According to the deterministic methodology of UN (United Nations and Social Affairs, 2006) the projection model used is as under

$$f_{c,t+1} = f_{c,t} - d(\theta, f_{c,t}), \tag{1.3}$$

where, $d(\theta, f_{c,t})$ is a parametric decrement function (Alkema *et al.*, 2009, September), whose value is based on the sum of two logistic functions (United Nations and Social Affairs, 2009) which is given in equation 1.6

In the past projections published were deterministic, and were decomposed into projected age specific fertility rates using fertility schedules. But the importance to take into account the uncertainty factor can not be ignored for long, as a result in the world population prospects (WPP2015), the projections have been done using probabilistic approach. To begin with the main idea, it is essential to know about fertility transition models, which is the basis of projection.

1.3 Fertility Transition models

In phase 2 fertility is declining while in phase 3 it is below replacement level and is expected/assumed to reach it in the long run. Since each phase behaves differently, it is therefore essential to use different models for both phases.

1.3.1 Fertility Transition Model for Phase 2

The model used to project fertility rate is a random walk model with drift as described by Alkema *et al.* (2011). It depends on current fertility level at time point t plus a double logistic function and a random error term. The double logistic function is used mostly in non linear regression and in binary response modeling (Lipovetsky, 2010). A regular logistic model is usually defined by four parameters, i.e. location, a parameter of the rate of change, and the lower and upper asymptotic levels, while double logistic or sigmoid model can be constructed as the sum of two different logits, with two parameters of centers, two parameters of the rates, and three parameters of the lower, middle,and upper levels (Lipovetsky, 2010). However, it can also be constructed with five or six parameters as done by Cairns *et al.* (2008). It can also be constructed as product of two logit functions (Roper, 2000). A double logistic function is superior to simple logistic function, specifically for the study of phenomena that experience two phases of logistic growth pulses, either overlapping or sequentially (Meyer, 1994). However, the model to be used for the current study as suggested by Alkema *et al.* (2011) is as under, which is based on the existing UN methodology with the only difference of adding a random error term to it, and is given as:

$$f_{c,t+1} = f_{c,t} - d_{c,t} + \varepsilon_{c,t} \tag{1.4}$$

where, $f_{c,t}$ is the fertility rate for Pakistan and/or its regions at time point t, $d_{c,t}$ is the decrement function or decline function which is used to model decline in fertility transition and about the expected decrease in TFR (Alkema *et al.*, 2015). The subscripts c defines the country (Pakistan or its regions) while t defines the time point at which the decline is reflected. The decrement function is defined as:

$$d_{c,t} = \begin{cases} g(\theta_c, f_{c,t}), & for \quad f_{c,t} > 1, \\ 0 & otherwise. \end{cases} \tag{1.5}$$

where $g(., f_{c,t})$ is a double logistic function, and is given as:

$$g(\theta_c, f_{c,t}) = \frac{d_c}{1 + exp\left(-\frac{2ln(9)}{\Delta_{c3}}(f_{c,t} - (\Delta_{c4} + 0.5\Delta_{c3}))\right)} + \frac{-d_c}{1 + exp\left(-\frac{2ln(9)}{\Delta_{c1}}(f_{c,t} - \sum_i \Delta_{ci} + 0.5\Delta_{c1})\right)} \tag{1.6}$$

The distribution of random error term is as under:

$$\varepsilon_{c,t} \sim \begin{cases} N(m_t, s_t^2), & for \quad t = \tau_c, \\ N(0, \sigma(f_{c,t})^2), & otherwise. \end{cases} \tag{1.7}$$

Here, s_t determines variation in the error term at the beginning of fertility transition, while $\sigma(f_{c,t})$ shows variability at later stages, while

$$\sigma(f_{c,t}) = c_{1975}(t)(\sigma_0 + (f_{c,t} - S)(-aI_{[S,\infty)}(f_{c,t} + bI_{[0,S)}(f_{c,t}))) \tag{1.8}$$

where σ_0 is the maximum standard deviation of the distortions, attained at TFR level S, and a and b are multipliers of the standard deviation, to model the linear decrease for larger and smaller outcomes of the TFR. As proposed by Alkema *et al.* (2011) the constant $c_{1975(t)}$ is added to model the higher error variance of the distortions before 1975, and is given by:

$$c_{1975}(t) = \begin{cases} c_{1975}(t), & t \in [1950 - 55, 1970 - 1975], \\ 1, & t \in [1975 - 1980, \infty]. \end{cases} \tag{1.9}$$

The constant $c_{1975(t)}$ is added because the data of WPP2015 (United Nations and Social Affairs, 2015) for TFR reveals higher variability for most of the countries before 1975.

Parameters $\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$ is a set of decline parameters for Pakistan and its regions. Here the parameter d_c determines maximum possible decrement for the region under study , while Δ_{ci}, where $i = 1, 2, 3, 4$, tells about the ranges of TFR within which fertility decline fluctuates.

Since Pakistan is in the phase of fertility transition, therefore, the change in TFR is modeled as a function of its current level, similar to UN methodology (United Nations and Social Affairs, 2006). The model is formulated as the sum of two logistic functions describing the five year expected decrement in TFR during the fertility transition period. According to Meyer (1994) a three parameter logistic function can be defined as

$$N(t) = \frac{k}{1 + e^{-\alpha t - \beta}} \tag{1.10}$$

where $N(t)$ represents growth rate of a process, α is the rate parameter, β is the location parameter and k is the saturation level or asymptotic value that bounds the function , as $N(t)$ approaches k the growth process approaches to zero resembling the S shaped curve. For the sake of convenience let us define t_m as the mid point of the above growth process, replacing and defining β by t_m or equivalently defining it as β = -$t_m \alpha$. For the sake of simplicity and convenience define a parameter Δt as the length of the time interval required for the $N(t)$ to grow from 10 to 90 percent of the saturation level k. Also from simple algebra $\Delta t = \ln(81)/\alpha$ (Meyer, 1994). Thus equation 1.10 can be simplified as under

$$N(t) = \frac{k}{1 + exp\{\frac{-\ln(81)}{\Delta t}(t - t_m)} \tag{1.11}$$

The same idea can be taken as done by Meyer (1994) in the above case, so let us define logistic function $L(.)$ evaluated at TFR level f and can be defined as

$$L(f) = \frac{d_c}{1 + exp(-\frac{2ln(p)}{\Delta}(f - f_{50\%}))} \tag{1.12}$$

where, $f_{50\%}$ is the midpoint of decrement and is equal to $0.5d_c$. In the given function in equation 1.12 as the process approaches to d_c , the function $L(.)$ decreases from upper bound d_c to lower bound 0 with the decrease of TFR, also Δ represents the length of decrease from 10 to 90 percent of TFR when allowing the decrease from $\dfrac{p}{p+1}d_c$ to $\dfrac{1}{p}d_c$, when p =9, the range of logistic function would be 80%.

In order to model the decrement function, another logistic function similar to above is added to equation 1.12 whose job is to counter balance the opposite force as done by Meyer (1994); Alkema *et al.* (2011). The mid point of second function is greater than the previous one such as balancing the effect in totality, i.e., making the sum of TFR zero. The reason is that the first function decreases towards zero while the second one increases towards zero. The combination of two logistic functions is reflected in equation 1.6. From the discussion above it follows that the in equation 1.6

$\Delta_{c4} + 0.5\Delta_{c3}$ is the mid point of first function , and its 80% range given by Δ_{c3} while

$\sum_{i=2}^{4} \Delta_{ci} + 0.5\Delta_{c1}$ being the midpoint of second function with its 80% range given by Δ_{c1}.

It is important to note that the sum of all Δ_{ci} i.e., $\sum_{i=1}^{4} \Delta_{ci} = U_c$ which determines the starting point of TFR for Pakistan and its regions. At this level, the outcome of second logistic function is -0.9d_c, and has begun movements towards zero, while in the case of first function its value is still above 0.9d_c (Alkema *et al.*, 2011).

1.4 Estimation of Phase II parameters

The format of regional and national level data is prepared keeping in mind the UN available data sets in WPP2015 (United Nations and Social Affairs, 2015). Based on this, we have only few observed points in phase II which causes a challenge for the estimation of parameters $\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$. As we have less than 10 observed declining TFR observations at national and regional level. To overcome the problem of estimation the rational to use Bayesian heirarchical model by Gelman *et al.* (2014) is considered.

1.4.1 Estimation

The estimation of parameters in phase 2 is done in the Bayesian framework. A Markov chain Monte Carlo (MCMC) algorithm is used to derive samples of the posterior distributions of the parameters of the fertility transition model (Alkema *et al.*, 2011). To begin with simulation approach, let us have a brief idea about it as below.

1.4.1.1 Rejecting Sampling approach

A number of approaches are available to generate simulated samples from posterior distribution, however among them the most simple and general one is rejecting sampling approach (Gelman *et al.*, 2014). The idea is to draw random sample from desired posterior distribution or often referred as target distribution, say $P(\theta \mid y)$, we need a positive function say $g(\theta)$, for which the target density is positive and has the following properties:

a. It is possible to generate samples from the density function proportional to $g(\theta)$ which is finite integral not necessarily integrate to 1.

b. The ratio between target density and positive function $g(\theta)$ is bounded by a constant, say M.

Then we can proceed to perform rejecting sampling as under:

1. Draw a candidate sample of θ from positive function $g(\theta)$.

2. Accept θ, if $\frac{P(\theta|y)}{Mg(\theta)} < 1$. Otherwise reject, and repeat step 1 (Gelman *et al.*, 2014).

After being able to generate samples from target distribution $P(\theta/y)$, it becomes easy to draw samples from the predictive distribution of unobserved samples $\tilde{y}$,i.e., $P(\tilde{y} \mid \theta)$ (Gelman *et al.*, 2014).

1.4.1.2 Markov Chain Monte Carlo (MCMC) Approach

The idea of MCMC is gaining much more attention in the class of statisticians in order to simulate complex samples (Chib and Greenberg, 1995)

MCMC is a general approach for drawing values of θ based on approximate distributions, which are corrected in the later setting to better approximate to the target density function, $P(\theta \mid y)$. Th concept of **Markove Chain** is such that for a set of values drawn from a random sample, the distribution of the current observed will depend only on the the most previous observation, i.e., if $\theta^1, \theta^2, \theta^3, ..., \theta^t$, the distribution of θ^t will depend only on θ^{t-1}. Thus it could be stated that the observations drawn in MCMC algorithm are slightly dependedent. The core idea is that after each observed value, the distribution is updated to better approximate to the target distribution (Gelman *et al.*, 2014).

The procedure is stopped once convergence is ensured. For the purpose of convergence the idea of Raftery and Lewis (1992, 1996) is taken into consideration for current study. In order to generate samples in Phase 2, the MCMC algorithm is a combination of Gibbs sampler, Metropolis hosting and slice sampling approach (Alkema *et al.*, 2011)

1.4.1.3 The Gibbs sampler

This approach is adopted for multidimensional problems. Suppose we have a joint distribution of θ as $P(\theta_i), \quad where \quad i = 1, 2, ..., k.$ which is the desired posterior distribution we wish to sample from, provided we know the full conditional distribution of each θ_i conditional on the known information and all other parameters,Liseo (2016), i.e.,

$$P(\theta_i \mid \theta_{\sim[i]}, y)$$

Steps for performing Gibbs sampling:

Let us define a posterior distribution as $P(\theta \mid y)$ and let start to construct the full conditional distribution for each parameter θ as under:

1 Describe the full posterior distribution by ignoring the constant of proportionality.

2 Select a block of parameters , say for instance θ_1 ignoring all that does not depend on θ_1.

3 Use knowledge about distributions to figure out what the normalizing constant is (and thus what is the form of full conditional distribution $P(\theta_1 \mid \theta_{\sim[1]}, y)$.

4 Repeat step 2 and 3 for the rest of parameters. (Gelman *et al.*, 2014; Liseo, 2016).

1.4.1.4 Metropolis-Hostings algorithm

Gelman (1993) noted that Gibbs sampling is a special case of Metropolis-Hastings algorithm (M-H). M-H has been used extensively due to the significant work of Müller (1991); Tierney (1994). As pointed out by Chib and Greenberg (1995) and is a powerful simulation technique which takes the idea of accepting or rejecting the candidates based on the famous acceptance rejection algorithm (also known as rejection sampling

(Gelman *et al.*, 2014)). The concept is useful when posterior distribution from which we wish to sample does not look like a proper distribution, consists of more than two parameters and some (or all) full conditionals don't look like any distribution, then we can use M-H algorithm (Liseo, 2016).

Steps for performing Metropolis-Hostings sampling:

1. Select a starting value say, θ^0 from a target distribution

2. Draw a proposed candidate θ^* from a target density function, say $q_t(\theta^* \mid \theta^{t-1})$.

3. Find out acceptance ratio as under:

$$\alpha = min\left(1; \frac{P(\theta^* \mid y)q_t(\theta^{t-1} \mid \theta^*)}{P(\theta^{t-1} \mid y)q_t(\theta^* \mid \theta^{t-1})}\right)$$

4. Accept θ^* as θ^t with probability α. If not accepted, then consider $\theta^t = \theta^{t-1}$.

5. Keep repeating step 2 to 4 K times, to get K candidates from $P(\theta \mid y)$.

1.4.1.5 Slice Sampling

One problem with usual Gibbs sampling approach is that one needs to develop strategy to sample from nonstandard univariate distribution, and one problem with Metropolis sampling is to find appropriate proposal distribution for efficient sampling, which in result restricts the use of these approaches to sophisticated softwares, in response to which Neal (2003) proposed a method to overcome this problem.

As suggested by Neal (2003) the core idea is to sample a variable of interest say x of dimension R^n with density function $f(x)$ by sampling uniformly from (n + 1) - dimensional region which lies under the plot of $f(x)$. Neal (2003) proposed to introduce an auxiliary variable, say y with a joint distribution which is uniform over the region $U = (x, y) : 0 < y < f(x)$, such as the joint distribution is given by:

(1)

$$P(x,y) = \begin{cases} \mathbf{1/Z}, & if \quad 0 < y < f(x), \\ \mathbf{0}, & otherwise, \end{cases} \tag{1.13}$$

where $Z = \int f(x)dx$ .The marginal density of x becomes

(2)

$$P(x) = \int_0^{f(x)} 1/Zdy = f(x)/Z$$

The procedure is to sample for both (x,y) and then ignore y.

1.4.2 Bayesian Hierarchical Model for Phase II

For the sake of meaningful interpretation based on observed data, the five year decrement has been restricted between 0.25 and 2.5 using logit transformation as under (Alkema *et al.*, 2009, September)

$$d_c^* = log(\frac{dc - 0.25}{2.5 - d_c}), \tag{1.14}$$

The distribution of transformed decrement function is

$$d_c^* \sim N(\chi, \psi^2),$$

where, χ is the mean and ψ^2 is the variance at global level. Similar to this the global distribution of rest of parameters is also done. However, when considering the case of Pakistan and its regions only, the distribution of national level parameters behaves like global distribution. Furthermore, Δ_{c4} has been transformed in such a way to restrict its values between 1 and 2.5 children (Alkema *et al.*, 2011) and is give by

$$\Delta_{c4}^* = log(\frac{\Delta_{c4} - 1}{2.5 - \Delta_{c4}}) \tag{1.15}$$

The distribution of Δ_{c4}^* is given as

$$\Delta_{c4}^* \sim N(\Delta_{c4}, \delta_{c4}^2),$$

As noted by Alkema *et al.* (2011), the actual pace observed is smaller than the decline parameter d_c, which depends on the 4 parameters $\Delta_{ci}, \quad i = 1, 2, 3, 4$, further the maximum value of TFR is observed at U_c which is "6.5" for Pakistan. Alkema *et al.* (2009, September) pointed out that the decline starts at level U_c and at this point its maximum pace is 10 % or $0.1d_C$, while at Δ_{c1} it reaches from value $0.1d_C$ to $0.8d_C$, at range Δ_{c2} it is maximum than the entire ranges and reaches from $0.8d_C$ to d_C, after this it starts to decline at range Δ_{c3} below $0.1d_C$, while parameter Δ_{c4} defines the asymptotic level of TFR at which point the drift term becomes zero (Alkema *et al.*, 2009, September).

For the sake of computational ease, the parameters, i.e., $\Delta_{ci}, \quad i = 1, 2, 3$ on which the decline depends, can be expressed as proportions of the difference between U_c and Δ_{c4} as under

$$p_{ci} = \frac{\Delta_{ci}}{U_c - \Delta_{ci}}, \qquad for \quad i = 1, 2, 3, \tag{1.16}$$

with the condition that $\Sigma_{i=1}^{3} p_{ci} = 1$. Theoretically speaking, we are working with more parameters than the number of observations per paramter, therefore, to cope with this issue the idea of Gelman *et al.* (1996) is taken into consideration by Alkema *et al.* (2011) and thus the equation 1.16 is transformed into a new set of parameters γ_{ci} which gives p_{ci} the following form

$$p_{ci} = \frac{exp(\gamma_{ci})}{\Sigma_{j=1}^{3} exp(\gamma_{cj})}, \tag{1.17}$$

where the distribution of new parameter is Normally distributed with hierarchical mean and variance as

$$\gamma_{ci} \sim N(\alpha_{ci}, \delta_{ci}^2),$$

From the above equations 1.14 to 1.17 , the group of hierarchical parameters may be given by $\gamma_{ci}, U_c, d_c, \Delta_{c4}$, $i = 1, 2, 3$. while from the same set of equations, the hyper-parameters could be defined as $\chi, \psi^2, \Delta_4, \delta_4, \alpha, \delta$ and $a, b, S, \sigma_0, c_{1975}, m_\tau, s_\tau$. The prior distributions on the hyperparameters are given by:

$$
\begin{aligned}
\chi &\sim N(-1.8, 0.6^2)\\
1/\psi^2 &\sim Gamma(1, 0.6^2),\\
\alpha_1 &\sim N(1, -1)\\
\alpha_2 &\sim N(0.5, 1),\\
1/\delta_i^2 &\sim Gamma(1, 1), \quad for \quad i = 1, ..., 3\\
1/\delta_4^2 &\sim Gamma(1.3, 0.8^2),\\
\Delta_4 &\sim N(0.3, 0.8^2),\\
a &\sim U[0, 0.15],\\
b &\sim U[0, 0.15],\\
\sigma_0 &\sim U[0.01, 0.6],\\
c_{1975} &\sim U[0.8, 2],\\
S &\sim U[3.5, 6.5],\\
m_\tau &\sim N(-0.25, 0.4^2),\\
1/s_\tau^2 &\sim Gamma(1, 0.4^2).
\end{aligned}
$$

These prior distributions on the hyperparameters have been selected on the basis of two criteria of which the first one is the least-squares fits to fertility declines and the second one is the guesses of rational outcomes (Alkema *et al.*, 2011). Furthermore, the convergence of all model parameters were evaluated by means of the run length diagnostic of Raftery and Lewis (1992, 1996). The length of the MCMC chain exceeded the required sample size for estimating the 2.5% and 97.5% percentiles of the posterior distributions of all model parameters to within +/-0.0125 accuracy with probability 0.95. Convergence of α_i, $i = 1, 2, 3$, was measured on the transformed scale, i.e. $\alpha_i/\Sigma_{i=1}^3(\alpha_i)$, as these parameters are only weakly identified on their original scale (the likelihood of the data conditional on these parameters, and thus the projections, are not affected due to the addition of a constant to all three α_i. In the same way for the case of γ_{ci}, i = 1, 2, 3, c = 1, . . . , C, convergence was evaluated for $\gamma_{ci}/\Sigma_{j=1}^{3}(\gamma_{cj})$, $i = 1, 2, 3, c = 1, ..., C.$)

.

1.4.3 Post Transition Model for Phase 3

Post-transition phase is the phase in which TFR is at the lowest level, and start to bounce back towards the replacement level. The replacement level for the entire globe is believed to be fixed at 2.1 children (Alkema *et al.*, 2011), however, it depends on the mortality rate and for countries with high mortality rate it is usually considered more than 2.1 (Craig, 1993). For our analysis we assumed varying values of replacement level to see how fast the trajectories would approach to the replacement level. The results are shown in chapter 3.

Only few countries have managed to reach a TFR below 2.1, mostly consists of European union (Alkema *et al.*, 2011). Pakistan is currently in fertility transition phase, but it is assumed that in general, in the long run the future trajectories would eventually fall below replacement level and then bounce back to replacement level (Lee and Tuljapurkar, 1994). For this purpose the model used to project TFR in this phase is a first order autoregressive model, because only few observations are available to model them, as a result higher order of autoregressive model are not considered for the current situation. The model is given as under

$$f_{c,t+1} = (1 - \rho)(\mu - f_{c,t}) + e_{c,t}, \tag{1.18}$$

where, error term is normally distributed with mean zero and variance s^2. The mean in this model is the replacement level which is assumed to be fixed, while $f_{c,t}$ is the TFR for region c at current time point.

1.4.3.1 Estimation of Post transition model

The values of ρ and s^2 are estimated by maximum likelihood approach based on the data at hand. Also the same has been done using MCMC algorithm as done in phase 2 to get an idea of the estimates. To obtain the results using available data the help of *bayesTFR* package has been taken , which is available online in R CRAN (Sevcikova *et al.*, 2017; Ševčíková *et al.*, 2011). It is important to note that the convergence to replacement level depends mostly on the value of ρ, larger the value the more time it will take for the future trajectories to reach to replacement level and vice verse.

1.4.3.2 Estimation of Post transition model in the Bayesian Hierarchical set up

The assumption regarding generalization of fixed replacement value of 2.1 for entire globe is not appropriate for some of the Asian countries (Basten *et al.*, 2012). To overcome this, Bayesian hierarchical modeling approach has been adopted to estimate the values of parameter instead of MLE approach (Raftery *et al.*, 2014a). This allowed to relax the assumption and allowed to consider the possibility of having flexible range for the ultimate level of TFR that is allowed to be slightly higher or lower than the replacement level. The new model is now free to choose values for this based on every country/region data pattern. The new model has the form

$$f_{c,t+1} = (1-\rho)(\mu_c - f_{c,t}) + e_{c,t}, \tag{1.19}$$

where,

$$\begin{aligned} e_{c,t} &\overset{iid}{\sim} Normal(0, \sigma_e^2), \\ \mu_c &\overset{iid}{\sim} Normal_{[0,\infty)}(\bar{\mu}, \sigma_\mu^2), \\ \rho_c &\overset{iid}{\sim} Normal_{[0,1)}(\bar{\rho}, \sigma_\rho^2), \end{aligned}$$

The distributions for μ_c and ρ_c are left truncated, so as not to allow negative values. The prior distributions for the hyperparameters involved in the above equations are given by

$\bar{\mu} \sim U[0, 2.1]$, $\sigma_\mu \sim U[0, 0.318]$, $\bar{\rho} \sim U[0, 1]$, $\sigma_e \sim [0, 0.5]$.

The priors are diffuse except for $\bar{\mu}$ which is restricted to the maximum value of 2.1

(Raftery *et al.*, 2014a).

1.4.4 Projection

The projection of TFR for any region depends on the trajectories of phase 2. These trajectories are considered as samples of predictive distribution obtained using the model of phase 2, which depends on the decrement function, current level TFR (at time point "t") and random error term. Furthermore, the outcomes of the model trajectories depend on the set of values estimated from the set of parameters $\theta_{c,t}$. To get clear idea of how it works, let us assume we wish to estimate the TFR trajectory for Pakistan for the year 2018, as the country of interest is already in phase 2, so the predictive distribution is shown by a sample $\left\{f^{(i)}_{PK,2018} : i = 1, 2, 3, ..., I\right\}$, and the final outcome of the trajectory depends on ith value of the sample, thus the future trajectory for Pakistan for 2018 would be given as under:

$$f_{PK,2018} = f^{(i)}_{PK,2017} - d^{(i)}_{PK,2017} + \varepsilon^{(i)}_{PK,2017},$$

where, $f^{(i)}_{PK,2017}$ is the ith member of sample of TFR outcomes for Pakistan for the year 2017, $d^{(i)}_{PK,2017}$ is the ith member of expected decrement function evaluated at the $f^{(i)}_{PK,2017}$ and $\theta^{(i)}_{PK}$ (ith sample of the parameter vector θ_{PK}), while $\varepsilon^{(i)}_{PK,2017}$ is the random error term drawn from normal distribution, with mean zero and variance $(\sigma(f_{PK,2017})^2)^{(i)}$. Since Pakistan is currently in phase 2, but as per assumption it will eventually reach to phase 3 once for any trajectory the TFR has decreased to replacement level TFR (Lee and Tuljapurkar, 1994; Alkema *et al.*, 2011). As noted by Alkema *et al.* (2011) the earliest possible starting period "t" of phase 3 is supposed to occur when i) $min_t(f_{c,t})^{(i)} \leqslant \Delta^{(i)}_{c4}$, and ii) $f^{(i)}_{c,t} > f^{(i)}_{c,t-1}$, The TFR level $\Delta^{(i)}_{c4}$ is restricted to be between 1 and 2.5 (as noted in equation 1.15) so as to ensure that fertility transition ends around replacement level. So the beginning of phase 3 depends on the trajectories of phase 2 and eventually on the pace of decrement function and random error term. As noted earlier, the beginning of phase 2 for Pakistan is given by U_c, so an additional prior distribution is made for the projected TFR as such to bound its limits between realistic values and is given by $f_{c,t+1} \sim U[0, U_c]$ (Alkema *et al.*, 2011). Based on the algorithms of *bayesTFR* package (Ševčíková *et al.*, 2011) many trajectories are generated and the best one is given by median trajectories of TFR for each period.

1.5 Mortality

Another important demographic component which is most important is mortality. It is the natural state of being mortal or dead and is calculated as a measure of number of deaths over a period of time (Yusuf *et al.*, 2014). Mortality is measured usually as the number of deaths to that of average population in a location of interest. It has significant role in shaping the structure of a society. The other interesting fact everybody wishes to know is how long they could expect to live? which is determined by means of **Life Expectancy**. According to the publication of social indicators of Pakistan 2016, the life expectancy at birth has improved significantly from 61.2 years in 1990 to 66.6 years in 2013, whereas for females it improved from 61.9 years to 67.5 as compare to 60.5 years to 65.7 years for males for the same period (Government of Pakistan, 2016). The concept of **Life Expectancy** is very influential and is defined in detail in the next section.

1.6 Life Expectation

One important demographic concept which tells the world about the good quality of health system in a country is usually done with the help of life expectancy or life expectation. Its use is not restricted to population studies and health studies only, but actuarial scientist also uses the concept of life tables and life expectancy for estimation of insurance statements. It is defined as the average number of years a synthetic cohort is expected to live if exposed to period mortality throughout their life (Preston *et al.*, 2000). Usually for men and women it is calculated separately and is estimated through life tables at birth or other age points (Yusuf *et al.*, 2014). However, in current study the life expectancy at birth for both genders is considered which makes the basis for the projection of population size at later stages.

1.6.1 UN Methodology on Life Expectancy

The UN projects life expectancy at birth deterministically (Raftery *et al.*, 2013), where the model used depends on current level of life expectancy for a specific gender plus the gain in life expectancy. The gain in life expectancy is given by a double logistic function. The model used by UN is as given below and is developed in the same way as done in the case of TFR:

$$l_{c,t+1} = l_{c,t} + g(l_{c,t}|\theta^{(c)}) \tag{1.20}$$

where, $l_{c,t}$ determines the life expectancy for a specific gender(male or female) at time point t for a specific country/region c, while $l_{c,t+1}$ determines the projected life expectancy for the same gender considered in the model at time point $t+1$ and for the same country/region c, where as $g(l_{c,t}|\theta^{(c)})$ determines the gain in life expectancy for a specific region c at a specific time point t and is given as under

$$g(l_{c,t}|\theta^{(c)}) = \frac{k^c}{1 + exp\left(-\frac{A_1}{\Delta_2^C}(l_{c,t} - \Delta_1^c - A_2\Delta_2^c)\right)} + \frac{z^c - k^c}{1 + exp\left(-\frac{A_1}{\Delta_4^c}(l_{c,t} - \sum_{i=1}^{3} \Delta_i^c - A_2\Delta_4^c)\right)} \tag{1.21}$$

with country/region specific parameters , $\theta^{(c)} = (\Delta_1^c, \Delta_2^c, \Delta_3^c, \Delta_4^c, k^c, z^c)$.

The first four parameters, $\Delta_1^c, \Delta_2^c, \Delta_3^c, \Delta_4^c$ tells about the range within which gain in life expectancy is changing, parameter k^c describes the approximate maximum gain in life expectancy, These are six parameters of double logistic function whose values are chosen by UN experts from the five possibilities, that is if the country for which they are projecting shows i) very slow gain in life expectancy, or ii) slow gain, or iii) medium gain, or iv) fast gain, or v) very fast gain in life expectancy (Raftery *et al.*, 2013). Furthermore, A_1 and A_2 are two constants whose values are chosen in such a way that these parameters $\{\Delta_i^c$, where $i=$ 1,2,3,4$\}$ are easily interpretable.

1.6.2 Probabilistic model

Since the model provided by UN does not take into account the uncertanity factor, so Raftery *et al.* (2013) suggested an addition to the model in terms of adding the random perturbation term, which makes the model looks like a random walk model with drift term, where the drift term is the gain in life expectancy. The second thing done by Raftery *et al.* (2013) is to allow the parameter of the model to vary based on pooled information obtained from each country and its neighbouring countries. This allows to use the Bayesian hierarchical model to estimate the parameters of the model based on a range of prior distributions for these parameters. The conditional distributions of the parameters of logistic function are truncated normal with different truncated ranges and are given below:

$$\begin{aligned}
\Delta_1^c/\sigma_{\Delta_1} &\overset{iid}{\sim} Normal_{[0,100]}(\Delta_1, \sigma^2_{\Delta_1}) \\
\Delta_2^c/\sigma_{\Delta_2} &\overset{iid}{\sim} Normal_{[0,100]}(\Delta_2, \sigma^2_{\Delta_2}) \\
\Delta_3^c/\sigma_{\Delta_3} &\overset{iid}{\sim} Normal_{[-20,50]}(\Delta_3, \sigma^2_{\Delta_3}) \\
\Delta_4^c/\sigma_{\Delta_4} &\overset{iid}{\sim} Normal_{[0,100]}(\Delta_4, \sigma^2_{\Delta_4}) \\
k^c/\sigma_k &\overset{iid}{\sim} Normal_{[0,10]}(k, \sigma^2_k) \\
z^c/\sigma_z &\overset{iid}{\sim} Normal_{[0,1.15]}(z, \sigma^2_z) \\
\varepsilon_{c,t} &\overset{iid}{\sim} Normal(0, (\omega \times f(l_{c,t-1}))^2)
\end{aligned}$$

The distributions for the six parameters of double logistic function are truncated so as to get positive values for all the parameters (Chunn *et al.*, 2010) with an exception for Δ_3 (Ševčíková and Raftery, 2011). These ranges have been set based on entire globe data pattern. The first five parameters, i.e., $(\Delta_1^c, \Delta_2^c, \Delta_3^c, \Delta_4^c, k^c)$ correspond to the values of life expectancy and maximum gain, while the sixth parameter z^c is the parameter showing asymtotic gain in life expectancy for a specific country/region. The findings of Oeppen and Vaupel (2002) suggested an estimated value of 1.11 per year gain for best practicing countries, but Raftery *et al.* (2013) suggested that for any given country the asymptotic rate of gain will not exceed the value of 1.15.

Furthermore, the distribution of error term $\varepsilon_{c,t}$ is also normally distributed with a mean of zero and standard deviation proportional to the regression spline fitted to the absolute residuals (Raftery *et al.*, 2013).

1.6.3 Estimation of Parameters using Bayesian Hierarchical model

The estimation of parameters is done under Bayesian Hierarchical modeling approach. For this purpose the prior distribution of Pakistan parameters is defined which is proper but diffused than the posterior distributions (Raftery *et al.*, 2013). The Bayesian framework is set as below:

$$\begin{aligned}
\Delta_i &\sim Normal_{[0,100]}(a_i, \sigma_i^2) \qquad i = 1, 2, 4, \\
\Delta_3 &\sim Normal_{[-20,50]}(a_3, \sigma_3^2) \\
k &\sim Normal_{[0,10]}(a_5, \sigma_5^2) \\
z &\sim Normal_{[0,1.15]}(a_6, \sigma_6^2
\end{aligned}$$

For females the values of prior parameters a_i, where $i = 1, 2, .., 6$ are based on the UN medium-pace and are given by (13.22, 41.07, 9.24, 17.60, 2.84, 0.38), where as the values of σ_i, where $i = 1, 2, .., 6$ are given by the UN variant for females, and are given by (14.78, 16.28, 133.13, 31.80, 0.81, 0.16).

For the national level variance parameters, $\sigma^2_{\Delta_i}$, where $i = 1, 2, 3, 4$, σ_k^2 and σ_z^2, the prior distribution used is inverse-gamma with four degrees of freedom, while for ω a diffuse Uniform[0,10] is used (Raftery *et al.*, 2013).

The posterior distributions of national level and regional level parameters was approximated by MCMC algorithm using either Gibbs sampling (Gelfand *et al.*, 1990) , Metropolis-Hastings (Hastings, 1970; Chib and Greenberg, 1995) , or Slice sampling (Neal, 2003) was implemented in **R**. A freely available **R** software package called *bayeslife* (Sevcikova *et al.*, 2017) was implemented. We changed the ranges of prior distributions based on our regional data pattern and allowed the inclusion of regional level hierarchy in the model so as to cope with the current situation and requirements. We used three chanins of length 70000 with a burn-in of 10000 scans. Based on convergence test due to Raftery and Lewis (1992, 1996) and Gelman *et al.* (2014) revealed that the chains have converged.

1.6.4 Joint Projection of male and female life expectancy based on modeling the Gap in life expectancy

An interesting relationship between male and female life expectancy exist, which in most countries reveal higher female life expectancy than male life expectancy at birth. This has also been observed for the case of Pakistan and its regions. But Trovato and Odynak (2011) argued that the gap between female and male life expectancy is narrowing for most of the develpoed coutries which could be generalised for entire world, besides Bongaarts (2009) pointed out that this trend can be generalised/hypothesised for developing countries too.

To get the joint projections for female and male life expectancy the approach is to build a relationship between them by modeling the difference/gap between the two using linear regression approach by taking into account the Bayesian hierarchical model female life expectancy projection as a covariate (Raftery *et al.*, 2014b). The male life expectancy is obtained by taking the difference between simulated value of gap projection and simulated value of female life expectancy projection. The female life expectancy projection values are already obtained at the first step before starting the procedure for gap modeling.

The gap model which is basically the difference between female and male life expectancy for a specific region c at time point t is given by $G_{c,t}$ is dependent on the four terms. The first term is the gap at previous lag time point, that is $G_{c,t-1}$, second term is the female life expectancy at birth at the first quinquennium, i.e, 1950-55, and is denoted by $e^{f}_{0,c,1953}$, the third term is the female life expectancy at birth in the same quinquennium and is given by, $e^{f}_{0,c,t}$, and the fourth term is the number of years from where the female

life expectancy at birth in the same time point would exceed a specific threshold denoted by τ, which defines the point at which the difference/gap between life expectancy of male and female stops expending and starts narrowing, namely $(e^f_{0,c,t} - \tau)_+$.
To sum up, the model looks like,

$$G_{c,t} = \quad min\{max\{G^*_{c,t}, L\}, U\}, \tag{1.22}$$

where,

$$G^*_{c,t} = \begin{cases} \beta_0 + \beta_1 e^f_{0,c,1953} + \beta_2 G_{c,t-1} + \beta_3 e^f_{0,c,t} + \beta_4 (e^f_{0,c,t} - \tau)_+ \varepsilon^{(1)}_{c,t}, & if \quad e^f_{0,c,t} \leq A, \\ G_{c,t-1} + \varepsilon^{(2)}_{c,t} & if \quad e^f_{0,c,t} > A, \end{cases}$$

where,

$$\varepsilon^{(1)}_{c,t} \quad \overset{iid}{\sim} \quad t(\mu = 0, \sigma^{(1)2}, \nu = 2),$$

$$\varepsilon^{(2)}_{c,t} \quad \overset{iid}{\sim} \quad N(\mu = 0, \sigma^{(2)2}),$$

In equation 1.22, t distribution is used so as to allow the possibility of generating outliers as being observed in the data (Raftery *et al.*, 2013). The model reveals that, gap in the current time point $G_{c,t}$ depends on the gap in the preceding point $G_{c,t-1}$ together with female life expectancy at birth $e^f_{0,c,t}$ and the difference between female life expectancy at birth and the possible threshold $(e^f_{0,c,t} - \tau)$. Estimates of the parameters involved in above equations are estimated through maximum likelihood estimation procedure based on data of world population prospects (WPP2015 and WPP2017) (United Nations and Social Affairs, 2015, 2017) together with the regional data of Pakistan obtained from Pakistan Bureau of Statistics and the demographic health survey program. In the equation 1.22, L and U defines the upper and lower possible values of observed gap estimated based on the data at hand, while A represents the highest observed levels of female life expectancy and beyond which is also estimated by MLE approach. All estimates for the above equations are obtained using R package hett (Taylor, 2009).

1.7 Reconstruction of Population based on fertility, mortality and migration

The basic approach adopted for projection of population is based on demographic balancing equation given by

End Population = Starting Population ±Natural increase ± Net Migration, where natural increase = births − deaths, and net migration = immigrants − emigrants

or

$$P_{c,t} = P_{c,t-1} + B_{c,t} - D_{c,t} + M_{c,t}, \tag{1.23}$$

where, B denotes the number of births, D the number of deaths and M net migration (Preston *et al.*, 2000) (Ševcıková *et al.*, 2014). The equation is mostly solved deterministically using cohort component method by decomposing fertility, mortality and migration rates into age sex specific components (Whelpton, 1928, 1936). UN is also adopting the same procedure by converting these components into age sex specific components (Ševcıková *et al.*, 2014). The UN approach before adopting probabilistic approach was to convert the projected TFR into age specific fertility rates using fertility distribution and schedules, while the life expectancy rates were converted using the variant of Lee-Carter method (Lee and Carter, 1992), and after which the cohort component model is applied (Ševcıková *et al.*, 2014).

1.7.1 Population Projection using Bayesian Hierarchical model

To account for the uncertainty factor instead of deterministic projected TFR and life expectancy for male and female, the probabilistic versions of TFR and life expectancy at birth for male and female are incorporated in the model (Raftery *et al.*, 2012). The conversion of TFR and Life expectancy into age specific pattern is done similar to what has been done by UN (Raftery *et al.*, 2012). The conversion of TFR is done using models of fertility schedules (Coale and Trussell, 1974; GENERAL, n.d), while the conversion of mortality rates into age specific mortality rates is done using variant of Lee-Carter method (Lee and Carter, 1992). The procedure starts by simulating TFR trajectories probabilistically for entire globe and specifically for Pakistan and its regions under the guidelines of Alkema *et al.* (2011), and then generating equal number of male and female life expectancy trajectories (Raftery *et al.*, 2014b). The next step is to use fertility schedules and variant of Lee-Carter method similar to UN to get age sex specific rates (Raftery *et al.*, 2012). Finally these trajectories are converted into future age specific population numbers with the help of cohort component model (Raftery *et al.*, 2012). The details of cohort component method can be viewed from Preston *et al.* (2000). The method takes into account the projected TFR and Life expectancy at birth that are obtained in the first step probabilistically and combines them with the net migration rates that are non-stochastic as available in WPP2012 (United Nations and Social Affairs, 2014).

Chapter 2

Data

2.1 Demographics of Pakistan

Pakistan is the fifth most populous country in the world at South Asia sharing borders with India to the east, Afghanistan to the west, Iran to the south west, and China to the north east. The country consists of four provinces namely Punjab, Sindh, KhyberPakhtoonkhwa (KPK), and Baluchistan, one federal capital territory, a group of federally administered tribal area (FATA), and two autonomous and disputed territories. The most populous province based on the 1998 census (Government of Pakistan, 2013) is Punjab with a population share of 55%, Sindh 23%, KPK 13% and Balochistan 5%. Furthermore, the fertility rate in the country showed a declining pattern based on DHS surveys where at the national level the TFR was observed as in 1990-91 which dropped to 3.8. A similar pattern was observed at regional levels as well with Balochistan region having the highest TFR (NIPS, 2012). The life expectancy at birth calculated based on demographic surveys also show improvement in the survival rate (Government of Pakistan, 2003, 2006, 2007). The country hosts one of the largest refugee population in the world (UNHCR, 2012; Rafi, 2015). Map by UNHCR showing the overview of internally disputed population and refugees can be seen from the link `http://unhcrpk.org/wp-content/uploads/2012/04/Pakistan_Refugees_Overview_August16_Ver1.pdf`.

Furthermore, according to the labor force survey of Pakistan (2013-14), the highest number of internal migration by most of the regions is to the province of Punjab followed by Sindh and KPK. The map of Pakistan and its regions [1] together with bordering [2] countries is shown below:

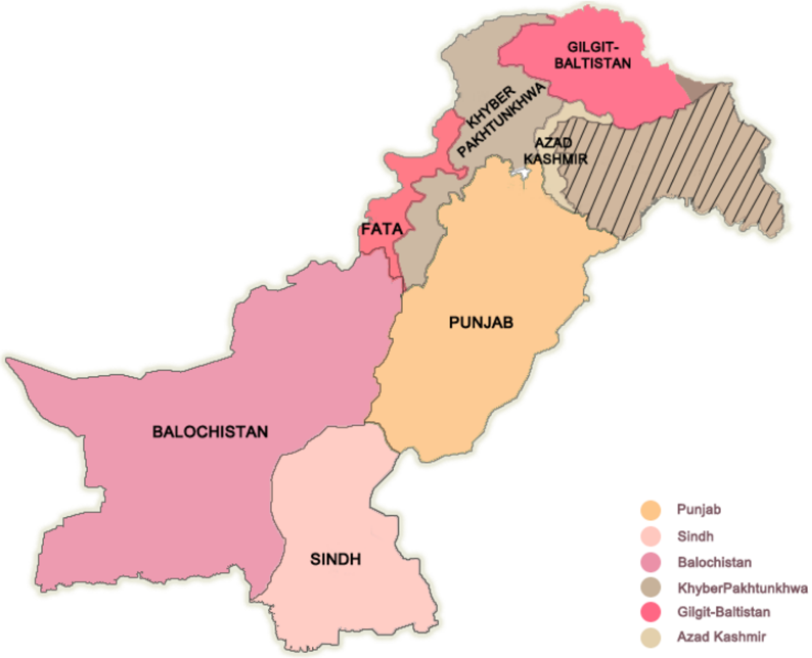


FIGURE 2.1: Pakistan map at regional level

FIGURE 2.2: Pakistan map with bordering countries

---

[1] source of fig 2.1: `http://wildlife.pk/index.php?cmd=introduction&action=pakistan`

[2] source of fig 2.2: `https://lizardpoint.com/geography/pakistan-quiz.php`

2.2 Fertility Data

Computation of fertility rates depend on the population of mothers belonging to the reproductive life span that is, from 15-49 years and the total number of births of children in this age interval. It is administratively and financially a hectic procedure, therefore needs to be done by state responsible agencies together with international organizations responsible for the collection and dissemination of needful statistical data.

Thus for the said purpose the required data has been obtained from Pakistan demographic surveys, Demographic and Health Surveys and World Population Prospects (United Nations and Social Affairs, 2015). Pakistan Demographic surveys are carried out by Pakistan Bureau of Statistics (PBS), state responsible federal agency of Pakistan, Demographic and health survey (DHS) are carried out by USAID through The DHS Program and World Population Prospects, the 2015 revision is a database updated every two year by United Nations department of economic and social affairs Population division, covering the important demographic data on a number of essential demographic indicators for entire world. The UN WPP2015 (United Nations and Social Affairs, 2015) database covers estimated data at national level for entire world in 5 years format starting from 1950-55 to 2010-15. The limitation of WPP2015 is that it does not give detail of estimates at sub national level. To overcome this, national and specifically regional level data was collected from different surveys whose detail is such that; for the periods 1984 to 1998, and 2000, the data from the publications of Pakistan demographic surveys was collected from the library of Pakistan Bureau of Statistics, Islamabad headquarter, while for the periods 2001 (Government of Pakistan, 2001), 2003 (Government of Pakistan, 2003), 2005 (Government of Pakistan, 2005), 2006 (Government of Pakistan, 2006) and 2007 (Government of Pakistan, 2007) were retrieved from the website of PBS. On the other hand till now only three demographic and health surveys for Pakistan have been done through the DHS program for the periods 1990-91, 2006-07 and 2012-13 and are accessible from the online database of the DHS program `http://dhsprogram.com/data/available-datasets.cfm`.

The missing data for the periods 1999, 2002, 2004 and before 1984 and after 2007 to

2011 and 2013 to 2015 have been interpolated and extrapolated using linear regression approach so as to come up like UN WPP2015 database format. For the case of missing observations, the UN adopted experts based openion by reviewing each observation analytically and by automated statistical approaches like local regression or cubic splines etc.(United Nations and Social Affairs, 2015). We kept the same idea and viewed every observation analytically and statically by keeping a bench mark of UN database pattern.

2.3 Mortality Data

Mortality rates are computed as the ratio of deaths to that of average mid year population and expressed per 1000. For this purpose, the needful data has been collected from Pakistan Bureau of Statistics and from WPP2015 online data base `https://esa.un.org/unpd/wpp/Download/Standard/Mortality/`. The rates are computed and later used for the construction of life tables [3], that are used for the projection of life expectancy at birth separately for both sexes at sub-national level. For some of the periods like 2003, 2006 and 2007 abridged life tables at national level were used from Pakistan Demographic survey reports available online on PBS website `www.pbs.gov.pk/demographic-and-research-publications` together with life expectancy data sets from WPP2015 data base. However, it was a hectic job to retrieve and reproduce the same data for both sexes, and for each province as available in the library of PBS, since only five years data is available online. Care has been taken into account to ensure the accurate reproduction of data, thus the reproduced data has been checked on every step to avoid any kind of error. An example of data used can be seen from the links `http://www.pbs.gov.pk/sites/default/files/population_satistics/publications/pds2006/tables/t01.pdf` and `http://www.pbs.gov.pk/sites/default/files/population_satistics/publications/pds2006/tables/t14.pdf`. Mortality rates for both sexes were used not only for projection of life expectancy but also for the projection of Population along with other data sets.

---

[3] Formulas used for the computation of these rates were obtained from the books of Preston *et al.* (2000); Yusuf *et al.* (2014)

2.4 Migration Data

Migration data is key determinant for the population count. It is the population of those who leave and enter the country. The population of those leaving the country is known as out migrants or emigrants while those entering the country are in migrants. Due to the complex nature of in and out migrants, it is a challenge to count exactly the correct number of in and out migrants. As a result, alternate approach is adopted to count the number of migrants.

2.4.1 Net Migration Rate

Since it is hard to find exactly the population of immigrants and emigrants, because it depends on a number of sources which are very difficult to cover up as a whole, as a result indirect approach is adopted by United Nations experts and also by (Azose and Raftery, 2015) to overcome this difficulty. It is computed based on indirect estimation of migration through residual methods. The method cannot provide information about in and out migrants but is a good estimate for the movements of all migrants (Siegel and Hamilton, 1952). The method is such that the difference between in and out migration is equal to net migration rate which is obtained by taking the difference of populations at two different census along with the difference of births and deaths.

$$(I - O) = (P_t - P_0) - (B - D)$$

Where, $(I - O)$ is net migration,

I = In migrants

O = Out migrants

P_t = Population at time “0”

$(P_t - P_0)$ is the difference between two populations at tow different time points.

P_0 = Population at time “t”

$(B - D)$ is an estimate of natural increase during the period from the net change

B = Number of births

D= Number of deaths

Hence, net migration is estimated as the difference in population change and natural increase over an intercensal period of time (Winkler *et al.*, 2013).

At national level the data is available in the WPP2015 (United Nations and Social Affairs, 2015), but at subnational level it is not available. To overcome this, we made assumptions on redistributing the migration rates in to subnational levels based on the labor force survey data. The distribution is based on the proportionate distribution of migrant population by place of existing and previous residence as reported in various labor force surveys, for instance, table 12 of 2013-14 labor force survey `http://www.pbs.gov.pk/sites/default/files//Labour%20Force/publications/lfs2013-14/t12-pak-fin.pdf`.

The data quality is validated at various stages by the UN experts and are cross validated for the purpose of better analysis (United Nations and Social Affairs, 2015), while the reliability of estimates for Pakistan and its regions is reported in the methodology section of labour force survey (Government of Pakistan, 2015).

2.5 Population Data

Population data consist of total head counts at each age group for both sexes for entire country. It is one of the biggest data collection activity carried out every ten years by most of countries in the world. However, for the case of Pakistan, this practice is not considered into account due to a number of factors, like political instability, separation of east Pakistan, law and order situation (TARIQ, 2016). The series of census conducted in Pakistan after her independence in 1947 is such that; 1951, 1961, 1972, 1988, and 1998 (Government of Pakistan, nd). The sixth population and housing census just carried out from march till may, 2017, and currently is in the phase of compilation. The counting of population in Pakistan is usually done based on both de facto ad de jure approach (Government of Pakistan, 2017). De facto is the approach in which population count is done where ever they are found without considering if they reside or do not reside at

that place at the time of census. On the other hand, de jure approach is just opposite of de facto, in its population of individuals is enumerated based on their place of residence irrespective of the place where they are found at the time of census.
As it is obvious that there exist missing data for few years, which is estimated so as to reflect the similarity pattern similar to that of UN.The approach adopted for this purpose is same as described in section 2.2. Furthermore, the data on growth rate can also be viewed at the link `https://esa.un.org/unpd/wpp/Download/Standard/Population/`.

2.6 Dependency Ratio Data

It is defined as the ratio of population not working at all to that group of population which is working. In technical terms it is defined as the ratio of population not in the labor force to that which is in labor force. Usually the population not working consists of age groups 0 - 14 and 65+ years. On the other hand, the working population belongs to the age group 15-64 years. That is:

$$Dependancy \quad Ratio = \frac{No. \quad of \quad people \quad aged \quad 0-14 \quad and \quad 65 \quad above}{No. \quad of \quad people \quad aged \quad 15-64} \quad X \quad 100$$

or

$$\text{Dependancy} \quad \text{Ratio} = \frac{P_{0-14} + P_{65+}}{P_{15-64}} \quad X \quad 100$$

The estimates are calculated from the population totals estimates, which are already obtained in the prior steps.

2.7 Percent Age Specific Fertility Rate Data

One more essential data set used for computation of population totals is percent age specific fertility rate, which is obtained as the percentage of age specific fertility rate with respect to total fertility rate. The rates are computed based on the estimates of demographic surveys. These rates play key role in determining age specific population totals. The data for the said purpose is collected from different demographic surveys whose detail is mentioned in the section 2.2. A table[4] showing percent age specific fertility rate is also shown as under:

	Year						
Age	**1980-1985**	**1985-1990**	**1990-1995**	**1995-2000**	**2000-2005**	**2005-2010**	**2010-2015**
15-19	8.04	7.56	6.71	5.98	5.43	5.23	5.49
20-24	20.84	21.26	21.81	22.08	22.33	23.15	24.48
25-29	24	24.5	25.42	26.53	27.96	29.31	30.08
30-34	20.82	21.02	21.38	21.83	22.23	22.7	23.37
35-39	15.41	15.39	15.06	14.53	13.72	12.64	11.39
40-44	8.5	8.19	7.61	7.08	6.35	5.23	3.93
45-49	2.39	2.09	2.01	1.97	1.98	1.75	1.27

Table 2.1: Percent Age Specific Fertility Rate for Pakistan for selected years

2.8 Sex Ratio at Birth Data

It is defined as the ratio of male population to female population at the time of birth. Mathematically

$$\text{Sex Ratio} = \frac{Number \quad of \quad males}{Number \quad of \quad females} \quad X \quad 100$$

or

$$\text{Sex Ratio} = \frac{P^M}{P^F} \quad X \quad 100$$

The rate plays key role in determining the number of males to female in the projection of population. Both census [5] and survey data are used for the computation of this rate.

A table [6] showing the Sex Ratio at birth at national level is shown sa well.

---

[4] Source of data: `https://esa.un.org/unpd/wpp/Download/Standard/Fertility/`

[5] `http://www.pbscensus.gov.pk/sites/default/files/DISTRICT_WISE_CENSUS_RESULTS_CENSUS_2017.pdf`

[6] source of table: `https://esa.un.org/unpd/wpp/Download/Standard/Population/`

Year	**Sex Ratio**	**Year**	**Sex Ratio**
1980	109.7	2010	105.8
1985	108.5	2015	105.6
1990	107.6	2020	105.6
1995	106.9	2025	105.5
1970	112.2	2030	105.3
1975	111.0	2035	105.1
2000	106.6	2040	104.9
2005	106.4	2045	104.7

Table 2.2: Sex Ratio at birth for Pakistan starting from 1980 to 2015.

Chapter 3

Results and Discussion

This chapter describes the output of fertility rates, life expectancy at birth for male and female and population counts projection for Pakistan and its regions in the form of tables and plots. The chapter also describes the trace plots of the parameters and parameter density plots.

3.1 Fertility Projection

The projection of total fertility rate for Pakistan and its regions was done based on the methodology described in chapter 1. The first step was to estimate the regional level data for Pakistan so as to come as close as possible to what has been reported by UN in the publication of WPP2015 and WPP2017 (United Nations and Social Affairs, 2015, 2017). For the case of fertility the total fertility rates were estimated from Pakistan demographic surveys together with the estimates compiled by UN and National Institue of Population Studies Pakistan in demographic health survey reports (NIPS, 2012). As described in section 1.2 there are three phases of fertility transition. The first phase which is the pre-transition phase is not modeled due to high or stable fertility rate. The second stage is transition phase, where fertility rate has started to decline.

In the study under consideration, the choice for the start period of phase II is defined to be a period where fertility rate is more than 5.5 (Alkema *et al.*, 2011), however, for the case of Pakistan and its regions the starting point of phase II is above the value of

6.6. This phase is modeled using the approach described in section 1.3.1. After knowing about the point from where phase II starts, the posterior samples for the parameters of phase II model are obtained using MCMC algorithm. Several checks were carried out for deciding the minimum number of iterations needed to have best estimates of parameters, which concluded with the final result of 70,000 iterations of three chains. The trace plot and diagnostics test suggested conversion of parameters. Once the TFR starts to decline, there reaches a time where two consecutive increases above two children are observed which is the start of phase III. Pakistan is currently in phase II, but as assumed there would eventually happen a time when future trajectories would enter this phase (Alkema *et al.*, 2011). After obtaining the posterior samples in phase II, the model in phase III is estimated as described in section 1.4.3.1. The final step leads to generation of future trajectories of TFR. We generated 1000 trajectories of which the best one was picked up by median trajectory together with 75, 80 and 95 percent confidence intervals.

One of the parameter in phase III is μ which describes the replacement level. The value for this parameter is 2.1. However, we tried to test different values for μ to check what would happen if the replacement level was different than what has been hypothesized.

3.1.1 Tables and Graphs of TFR Trajectories

The following table describes the trajectories of TFR for Pakistan and its regions starting from the year 2013 to 2098. The projections are presented at five years gap due to the data format of UN. The UN has collected data for entire globe starting from 1950 till recent years, and are presented in the format for the year t to $t+5$ from 1st July to 30th June, with centering at $t+3$ of 1st January (Alkema *et al.*, 2011). The data is updated every two years and is published in world population prospects, while the most recent available data is of WPP2017 (United Nations and Social Affairs, 2017).

Detailed tables and graphs showing TFR trajectories with 75, 80 and 95 percent confidence intervals and graphs are shown in Appendix.

Year	**Pakistan**	**Punjab**	**Sindh**	**KPK**	**Balochistan**
2013	3.578	3.468	3.900	3.836	3.586
2018	3.257	3.133	3.574	3.470	3.199
2023	2.990	2.883	3.300	3.193	2.892
2028	2.790	2.673	3.079	2.950	2.652
2033	2.607	2.498	2.896	2.755	2.452
2038	2.470	2.359	2.737	2.584	2.267
2043	2.347	2.233	2.600	2.436	2.127
2048	2.249	2.150	2.483	2.312	2.009
2053	2.155	2.048	2.357	2.193	1.923
2058	2.066	1.978	2.256	2.106	1.860
2063	1.994	1.935	2.172	2.040	1.810
2068	1.948	1.905	2.107	1.982	1.784
2073	1.910	1.873	2.046	1.939	1.791
2078	1.881	1.860	2.008	1.911	1.789
2083	1.859	1.851	1.963	1.890	1.795
2088	1.853	1.851	1.932	1.880	1.808
2093	1.840	1.853	1.925	1.872	1.814
2098	1.836	1.852	1.915	1.866	1.827

TABLE 3.1: TFR Trajectories for Pakistan and its regions starting from year 2013 to 2098 based on BHM approach.

The table above describes the TFR trajectories for Pakistan and its regions. The best trajectories out of 1000 trajectories are represented by median value. These trajectories are generated based on Bayesian hierarchical modeling approach (BHM) for phase III. For the sake of comparison, we also generated trajectories based on classical AR(1) approach, where the parameters were estimated using MLE approach. We found significant difference in the trajectories generated based on classical approach and based on Bayesian hierarchical approach, which is contradictory to what has been reported by Raftery *et al.* (2014a), who suggested there was no such difference even if fix value of replacement level were considered for eastern Asian countries as a response to the concerns raised by Basten *et al.* (2012).

In the classical approach we fixed the replacement level to 2.1, while in the case of BHM the data suggested possible value for replacement level. There is an abrupt drop in TFR value in the next five years from now, if BHM approach is used, while the scenario is different based on classical approach. It was also observed that at national level the time period taken by trajectories to go below a TFR of 2.1 was observed after 2053 based on BHM approach while based on classical approach it reaches there after 2063. The table generated based on classical AR(1) approach, is shown in Appendix.

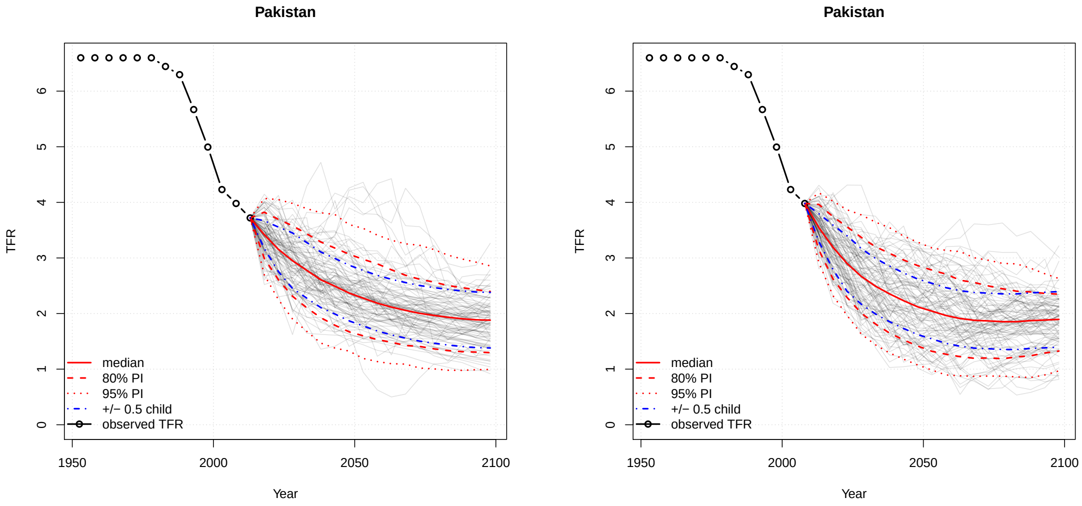
The table 3.1 also describes TFR trajectories for the four provinces namely, Punjab, Sindh, KPK and Balochistan. Interesting thing was to see different behaviour of trajectories based on classical AR(1) approach in comparison to BHM approach. As in the case of classical approach in phase III, we were free to choose replacement level values based on our own judgment, and different combinations of variable number of countries. It was shown when entire world level data was used, the results of Punjab, Sindh and KPK shown a bit similarity to national level, while Balochistan shown different result. However, based on BHM approach, there was no such similar pattern as observed in the case of classical approach.

The table 3.2 gives comparison of TFR when variable number of countries combinations were used in the model. The most important feature of Bayesian Hierarchical model is that, it is used when we don't have sufficient information to support the estimation process. In this case it borrows information from the neighbouring sources, making it possible to have good estimates. It can be seen from the table below, that for insufficient information in the case of case i) and ii), the trajectories are showing different pattern however for case iii) and iv) there exist a bit similarity in the long run. It could be said that for Asian countries and entire world data, the behaviour of trajectories pattern is very close in the long run. This suggests, instead of using entire globe data, Asian countries data is sufficient to use, which in terms saves time and increases the speed of computation as less time is taken for simulation study when less number of countries are used.

Year	**PAK ONLY**	**S.ASIA**	**ASIA**	**WORLD**
2018	3.433	3.394	3.378	3.381
2023	3.168	3.134	3.092	3.105
2028	2.964	2.916	2.866	2.885
2033	2.785	2.728	2.696	2.705
2038	2.631	2.577	2.531	2.556
2043	2.495	2.450	2.396	2.422
2048	2.376	2.334	2.281	2.309
2053	2.263	2.243	2.174	2.222
2058	2.170	2.152	2.119	2.149
2063	2.091	2.078	2.066	2.076
2068	2.026	2.027	2.013	2.030
2073	1.978	1.982	1.980	1.994
2078	1.927	1.942	1.962	1.961
2083	1.885	1.909	1.944	1.940
2088	1.851	1.883	1.932	1.935
2093	1.832	1.864	1.938	1.929
2098	1.799	1.861	1.935	1.915

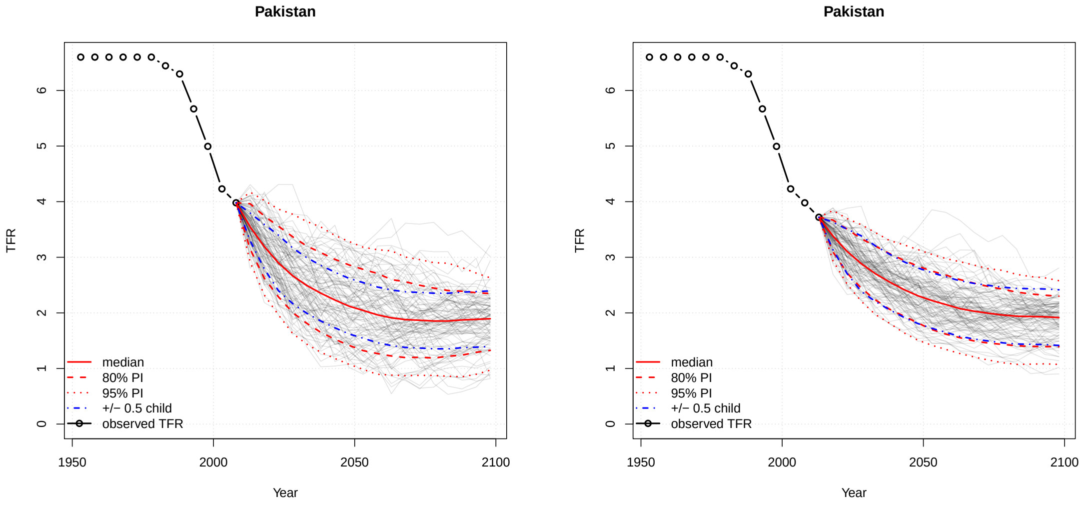
TABLE 3.2: Median Trajectories based on four combinations of data sets for, i) Pakistan and its regions only, ii) S.Asian countries, iii) Asian countries, and iv) complete world data.

Figures 3.1(a) to 3.1(d) shows the trajectories for four combinations as described in table 3.2. Although it is not so obvious to see the difference from the figures due to very fractional differences but they are obvious in table 3.2. In the figures TFR trajectories at national level are plotted against years, where the best trajectories are presented by Median trajectory, together with 80% and 90% projection intervals.



(a) TFR Trajectories based on Pakistan data only

(b) TFR Trajectories based S.Asian countries data



(c) TFR Trajectories based on Asia data

(d) TFR Trajectories based on world data

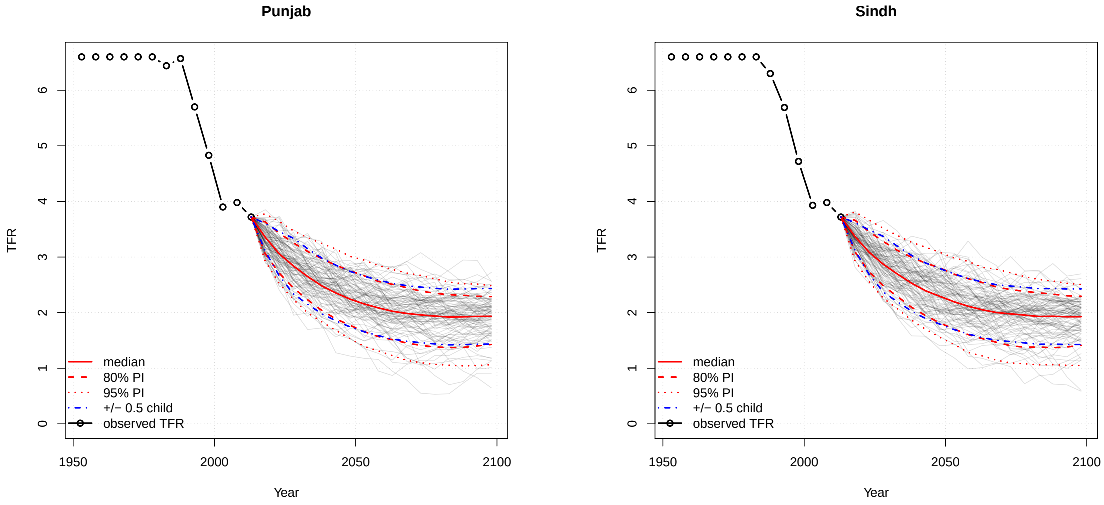
FIGURE 3.1: graphs showing TFR trajectories at national(Pakistan) level based on data from i) Pakistan and its regions, ii) Southern Asian countries, iii) Asian countries, and iv) entire world data.

The table 3.3 describes the behaviour of different values of replacement level on the trajectories of Pakistan. The four possible combinations considered are i) a value of $\mu = 1.85$, ii) $\mu = 2.0$, iii) $\mu = 2.10$, and iv) $\mu = 2.20$. For larger values, i.e., more than 2.0, the conversion of trajectories to the value of μ takes long time while for smaller value, the conversion is comparatively fast.

Year	2.2	2.1	2.0	1.85
2018	3.3818	3.3806	3.3737	3.3821
2023	3.1089	3.1049	3.1092	3.1064
2028	2.8864	2.8854	2.8862	2.8833
2033	2.7093	2.7048	2.7053	2.6992
2038	2.5622	2.5560	2.5480	2.5542
2043	2.4199	2.4223	2.4275	2.4271
2048	2.3026	2.3086	2.3203	2.3190
2053	2.2048	2.2222	2.2213	2.2104
2058	2.1351	2.1488	2.1299	2.1114
2063	2.0689	2.0758	2.0678	2.0411
2068	2.0194	2.0295	2.0126	1.9808
2073	1.9788	1.9940	1.9607	1.9352
2078	1.9557	1.9614	1.9372	1.8969
2083	1.9488	1.9403	1.9076	1.8680
2088	1.9393	1.9350	1.8979	1.8555
2093	1.9298	1.9291	1.8917	1.8374
2098	1.9348	1.9147	1.8847	1.8271

Table 3.3: TFR Trajectories based on different values of replacement level for the case of Pakistan starting from year 2018 to 2098

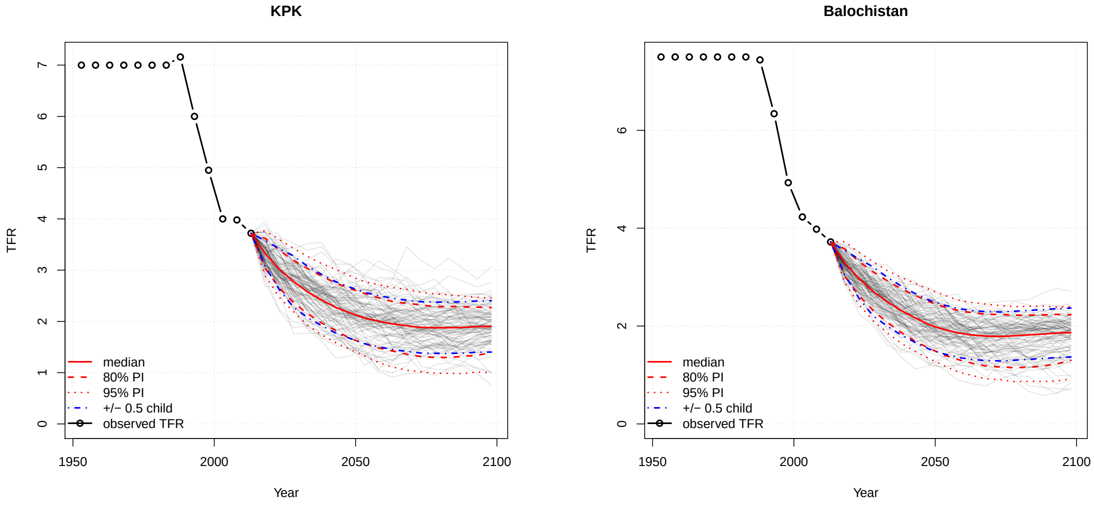
Figure 3.2(a) to 3.2(d) describes the trajectories for four provinces. The Double logistic curves are also shown from figure 3.3(a) to 3.3(d). The double logistic curves are plotted against the decreasing TFR, where the pre transition phase and transition phase are also pointed out. The decline function is shown on y-axis, while TFR rates are reversed so as to show declining pattern on x-axis. The trajectories for decline function are drawn with 80% projection interval.


Punjab

Sindh

(a) TFR Trajectories for Punjab province

(b) TFR Trajectories for Sindh province

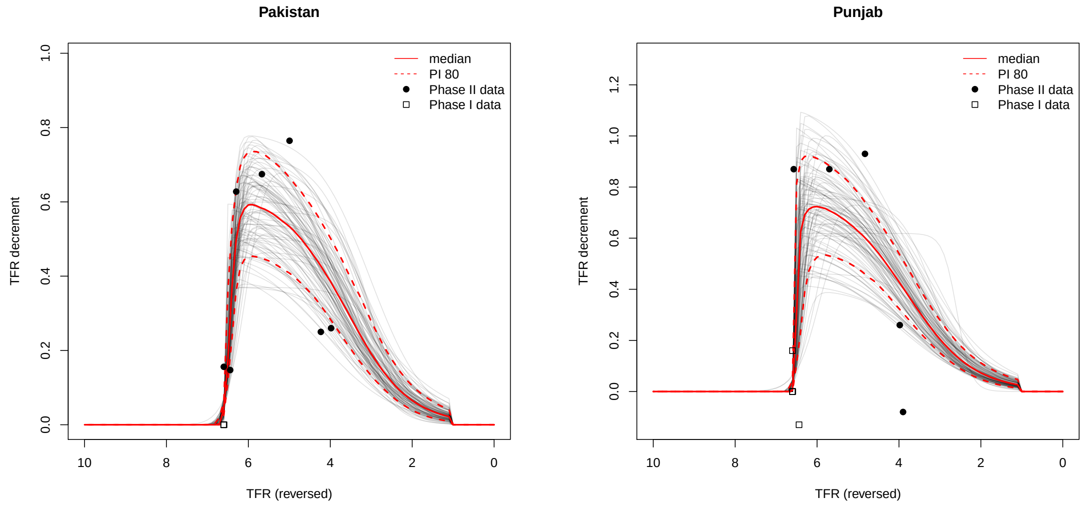

KPK

Balochistan

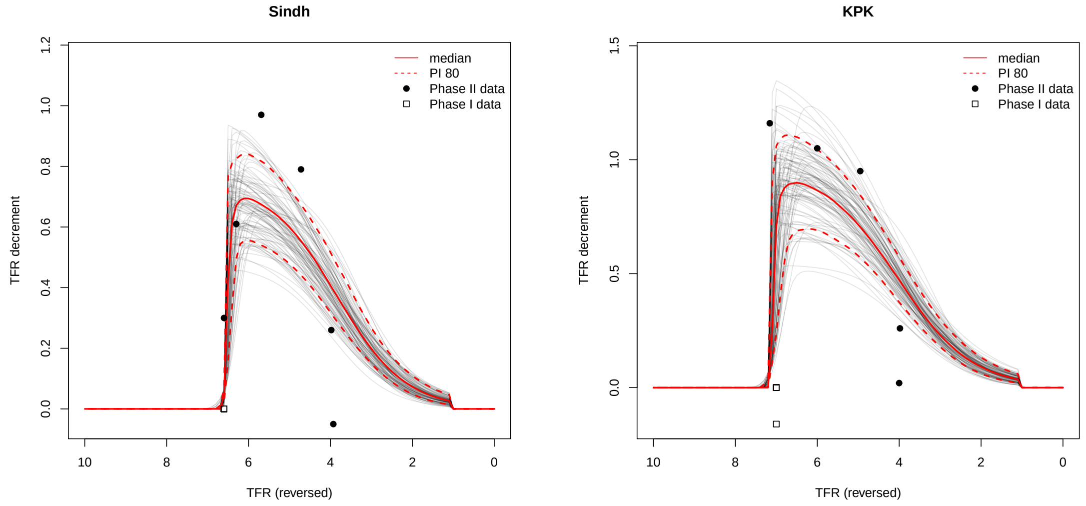
(c) TFR Trajectories for KPK province

(d) TFR Trajectories for Balochistan province

Figure 3.2: Graphs showing TFR trajectories at regional level for i)Punjab, ii) Sindh, iii) K.P.K, and iv) Balochistan



(a) DL Curve for Pakistan (b) DL Curve for Punjab



(c) DL Curve for Sindh (d) DL Curve for KPK

FIGURE 3.3: Graphs showing Double Logistic Curve at national as well as regional level.

3.1.2 TFR Parameters estimates

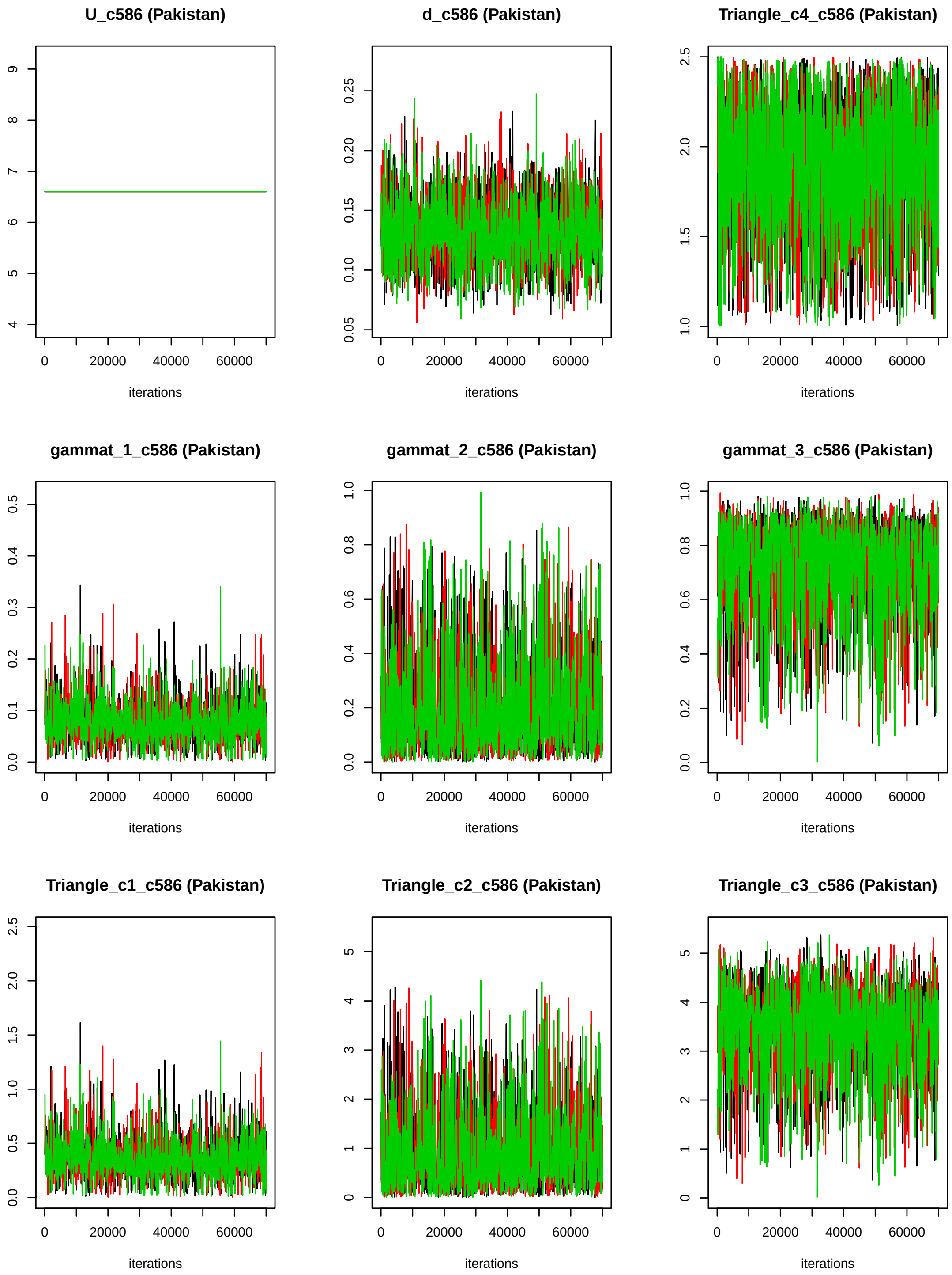
The table below gives summary statistics for each variable involved in the simulation process. The table provides empirical sample mean and sample standard deviation for each variable together with standard error of mean. In the columns below SD estimates the square root of the variance of posterior samples. The standard error gives a measure of the precision of the sample mean as a point estimate for the posterior mean, which depends on the number of iterates and the autocorrelation within samples. The table provides three different estimates of standard error. The first one is empirical SD, the second one, Naive S.E which is obtained by dividing the square root of number of iterates over the empirical S.D, while time series S.E gives the asymptotic standard error obtained as a result when the square root of the spectral density estimate is divided by the sample size.

Trace plots of the parameters revealed conversion which were also justified from diagnostic tests of Raftery and Lewis (1992, 1996). The trace plots are shown for each region separately in appendix. The first trace plot shows the starting value of TFR for each region separately, and is represented by $U_{c,region-number}$. This is the general format which is based on international coding system (Armingeon *et al.*, 2013). The second trace plot is given by $d_{c,region-number}$, which defines the maximum possible pace of decrement rate. The third parameter is presented by $\Delta_{c4,region-number}$ which defines the TFR range, in which the pace of the fertility decline changes (Alkema *et al.*, 2009, September). The fourth, fifth and sixth parameters are named as $p_{ci} = \frac{exp(\gamma_{ci})}{\Sigma_{j=1}^{3} exp(\gamma_{cj})}, \quad for \quad i = 1, 2, 3$ and are introduced for computational purpose only, while parameters seven to nine are given by $\Delta_{ci,reion-number}$, where $i = 1, 2, 3$, also define the TFR ranges in which the pace of the fertility decline changes (Alkema *et al.*, 2011). The mean and standard deviation of hyperparameters involved in the model are given in the table 3.4. Details about prior distribution of these parameters is given in subsection 1.4.2.

Table and plot at national level are shown here, while regional level results are given in appendix.

Variable	**Mean**	**SD**	**Naive SE**	**Time-series SE**
δ_1	1.0695	0.2375	0.0017	0.0109
δ_2	1.1902	0.3392	0.0024	0.0237
δ_3	0.8932	0.2101	0.0015	0.0107
Δ_4	0.3812	0.3025	0.0021	0.0101
δ_4	1.3844	0.5001	0.0035	0.0296
ψ	1.1191	0.0892	0.0006	0.0017
χ	-1.9231	0.1219	0.0008	0.0046
a	0.0365	0.0258	0.0002	0.0036
b	0.0155	0.0090	0.0001	0.0011
c_{1975}	1.6327	0.0584	0.0004	0.0022
S	4.8701	1.1653	0.0081	0.1859
σ_0	0.2100	0.0093	0.0001	0.0003
m_τ	-0.2200	0.0204	0.0001	0.0002
s_τ	0.3052	0.0146	0.0001	0.0001
U_{Pak}	6.6000	0.0000	0.0000	0.0000
d_{Pak}	0.1319	0.0286	0.0002	0.0004
Δ_{Pak4}	1.8866	0.3776	0.0026	0.0046
p_{pak1}	0.0783	0.0402	0.0003	0.0005
p_{pak2}	0.1948	0.1759	0.0012	0.0028
p_{pak3}	0.7269	0.1753	0.0012	0.0026
Δ_{Pak1}	0.3665	0.1854	0.0013	0.0021
Δ_{Pak2}	0.9171	0.8350	0.0058	0.0133
Δ_{Pak3}	3.4298	0.8895	0.0062	0.0130

TABLE 3.4: Empirical mean and standard deviation for each variable at national level (Pakistan), plus standard error of the mean


U_c586 (Pakistan)

d_c586 (Pakistan)

Triangle_c4_c586 (Pakistan)

gammat_1_c586 (Pakistan)

gammat_2_c586 (Pakistan)

gammat_3_c586 (Pakistan)

Triangle_c1_c586 (Pakistan)

Triangle_c2_c586 (Pakistan)

Triangle_c3_c586 (Pakistan)

FIGURE 3.4: Trace Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$

(gammat) at national level.

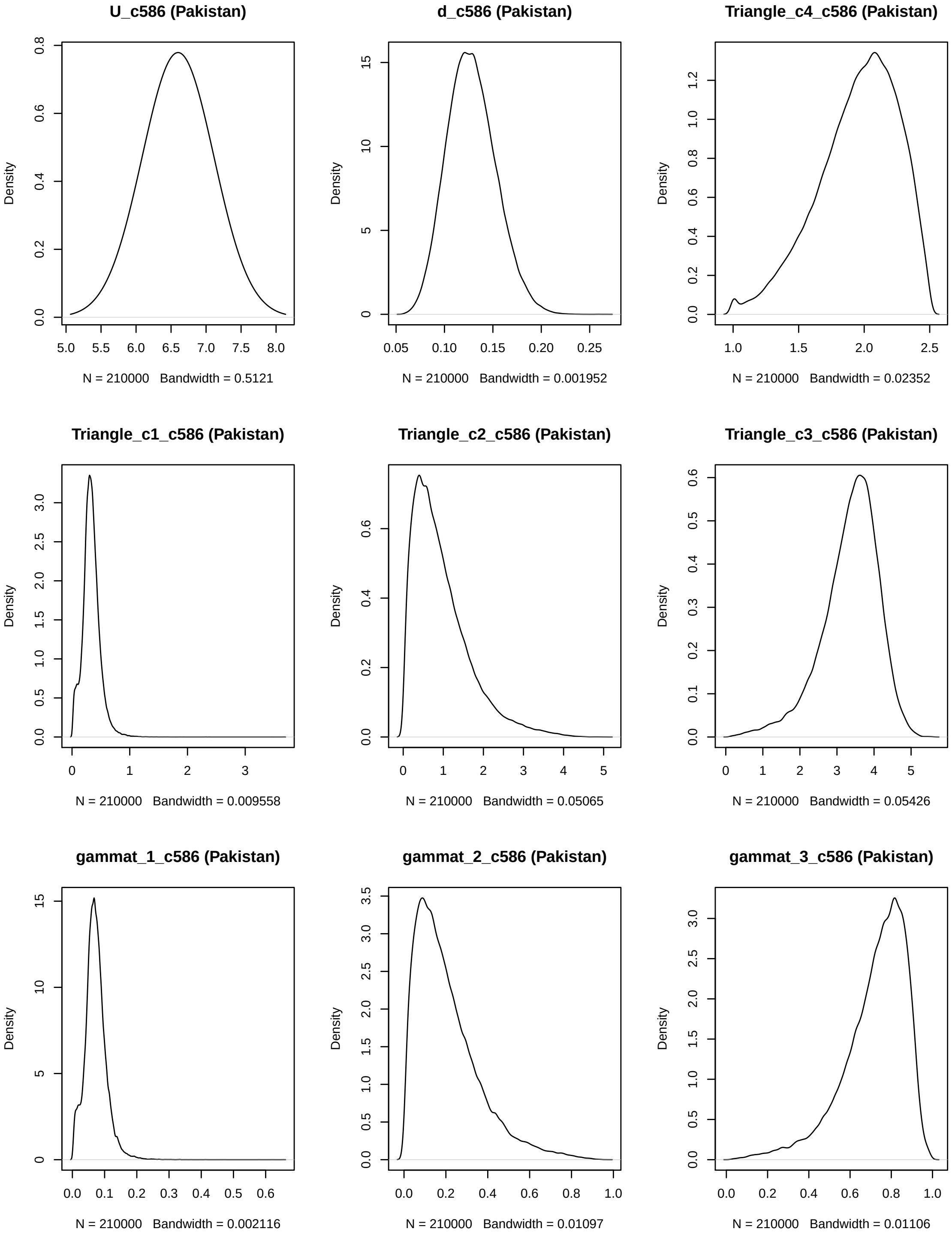


FIGURE 3.5: Density Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$

(gammat) for Pakistan.

3.1.3 Out of Sample Validation for TFR

An out of sample validation was performed to check the validity of the model fitting as done by Alkema *et al.* (2011). We used Bayesian Projection model to construct projections for the period 1990-2015 based on the data of UN and Pakistan Bureau of Statistics (PBS) estimates. We compared the results obtained with that of the estimates of UN and PBS. The results are shown in the table below.

Data untill 1990	**MSE**	**Above Median**	**Proportion of Obs. Above 95% PI**	**Proportion of Obs. Below 95% PI**
1990 – 1995	0.13	0.51	0.05	0.01
1995 – 2000	0.21	0.45	0.04	0.05
2000 – 2005	0.30	0.43	0.03	0.07
1995 – 2000	0.42	0.46	0.02	0.10
2000 – 2005	0.47	0.42	0.03	0.08
2005 – 2010	0.61	0.37	0.04	0.10
2010 – 2015	0.77	0.39	0.03	0.08

Table 3.5: Out of Sample validation for TFR: Mean Square error and proportion of left-out UN estimates that fall above the median projected TFR, and above or below their 95% projection interval in future periods.

As a rule of thumb, if the projection model is valid, we expect on the average about 2.5% observations to fall below the 95% PI and 2.5% to fall above the 95% PI (Alkema *et al.*, 2011). Looking from the results, the projection intervals appeared to be well calibrated in the out of sample projections.

3.2 Life Expectancy Projections

The projections of life expectancy for males and females separately and combined were generated based on the methodology defined in section 3.7. Similar to the procedure of TFR, we first obtained estimates of life expectancy for both genders at regional level based on the data of demographic surveys. The first step was to create life tables by using mortality estimates, published in demographic surveys, which were later on incorporated in the WPP2015 and WPP2017 (United Nations and Social Affairs, 2015, 2017), because the UN data base does not provide sub-national level estimates. The next step was to run the simulation study where 70,000 iterations of three chains were enough to provide good estimates for the posterior predictive distribution.

3.2.1 Tables and Graphs of Life Expectancy Trajectories

Table 3.6 gives life expectancy at birth for females at national and regional level, while table 3.7 gives life expectancy at birth for male at national and regional level. The projections reveal nearly similar pattern for three provinces, i.e. Punjab, Sindh and KPK, but different for Balochistan. We also tried to investigate the effect of variable number of countries on the trajectories with the same combination of countries as used in the case of TFR. The result had same effect as reported in the case of TFR. Furthermore, we plotted the combined graph of males and females based on the methodology described in subsection 1.6.4, and observed that the gap between males and females in the region of Balochistan is comparatively more than other regions while for the region of Punjab the gap is comparatively small than the rest of the regions. The mean and standard deviation of the hyperparameters is shown in appendix.

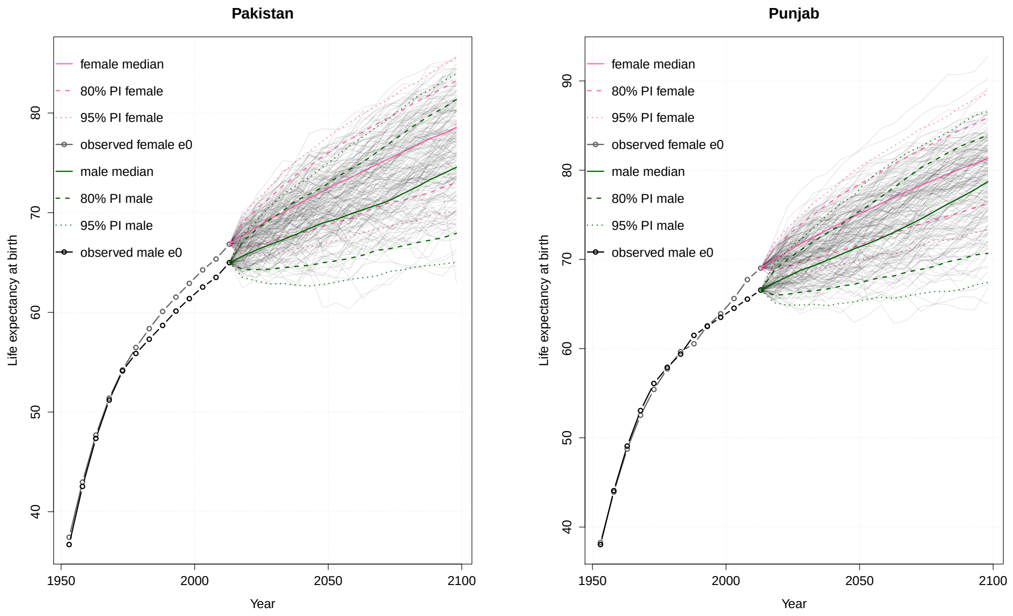
Year	**Pakistan**	**Punjab**	**Sindh**	**KPK**	**Balochistan**
2018	67.70	70.00	70.98	70.19	66.40
2023	68.55	70.91	71.82	70.98	67.07
2028	69.34	71.78	72.64	71.73	67.75
2033	70.00	72.51	73.41	72.42	68.42
2038	70.75	73.32	74.21	73.10	68.96
2043	71.47	74.09	74.87	73.72	69.48
2048	72.14	74.85	75.63	74.42	70.07
2053	72.87	75.53	76.24	75.02	70.66
2058	73.46	76.26	76.94	75.63	71.17
2063	74.15	76.97	77.63	76.25	71.73
2068	74.81	77.65	78.14	76.85	72.20
2073	75.45	78.29	78.78	77.43	72.78
2078	76.10	78.93	79.43	78.06	73.41
2083	76.77	79.57	80.06	78.53	73.91
2088	77.42	80.19	80.65	79.15	74.50
2093	77.94	80.71	81.22	79.77	75.03
2098	78.53	81.32	81.84	80.29	75.60

TABLE 3.6: Median trajectories of Female Life Expectancy at birth for Pakistan and its regions.

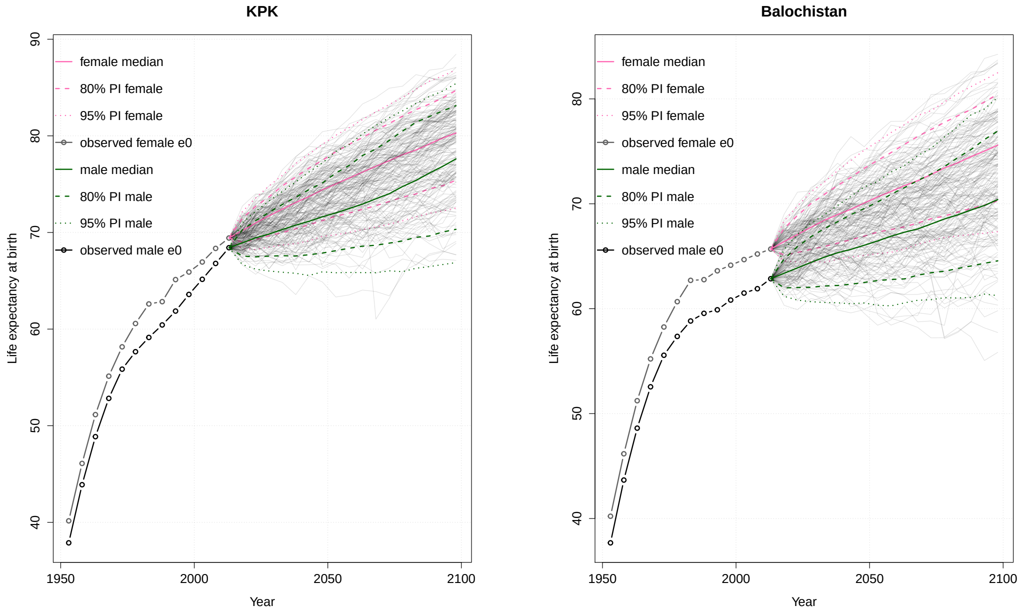
Year	**Pakistan**	**Punjab**	**Sindh**	**KPK**	**Balochistan**
2018	65.65	67.35	67.83	68.93	63.38
2023	66.25	68.07	68.43	69.45	63.89
2028	66.78	68.79	69.09	69.83	64.41
2033	67.29	69.31	69.63	70.31	64.90
2038	67.83	69.93	70.28	70.76	65.33
2043	68.43	70.55	70.74	71.16	65.65
2048	68.98	71.19	71.33	71.63	66.05
2053	69.33	71.74	72.00	72.02	66.46
2058	69.83	72.38	72.66	72.47	66.91
2063	70.29	73.05	73.34	72.94	67.29
2068	70.75	73.78	74.02	73.49	67.60
2073	71.23	74.54	74.94	74.01	68.07
2078	71.91	75.36	75.75	74.75	68.55
2083	72.57	76.21	76.62	75.39	68.95
2088	73.30	76.97	77.51	76.14	69.40
2093	73.91	77.80	78.30	76.84	69.84
2098	74.56	78.71	79.12	77.62	70.40

TABLE 3.7: Median trajectories of Male Life Expectancy at birth for Pakistan and its regions.

Figure 3.6(a) to 3.6(d) shows life expectancy for both genders together on same plot. These were generated based on joint probabilistic projection approach (Raftery *et al.*, 2014b). The results show that the gap between male and female life expectancy at birth for Balochistan province is greater as compare to rest of the country. Trace plots for Pakistan are shown in figure 3.7 and density plot of parameters are shown in figure 3.8. Details of male and female life expectancy tables and graphs separately for each region are shown in appendix.

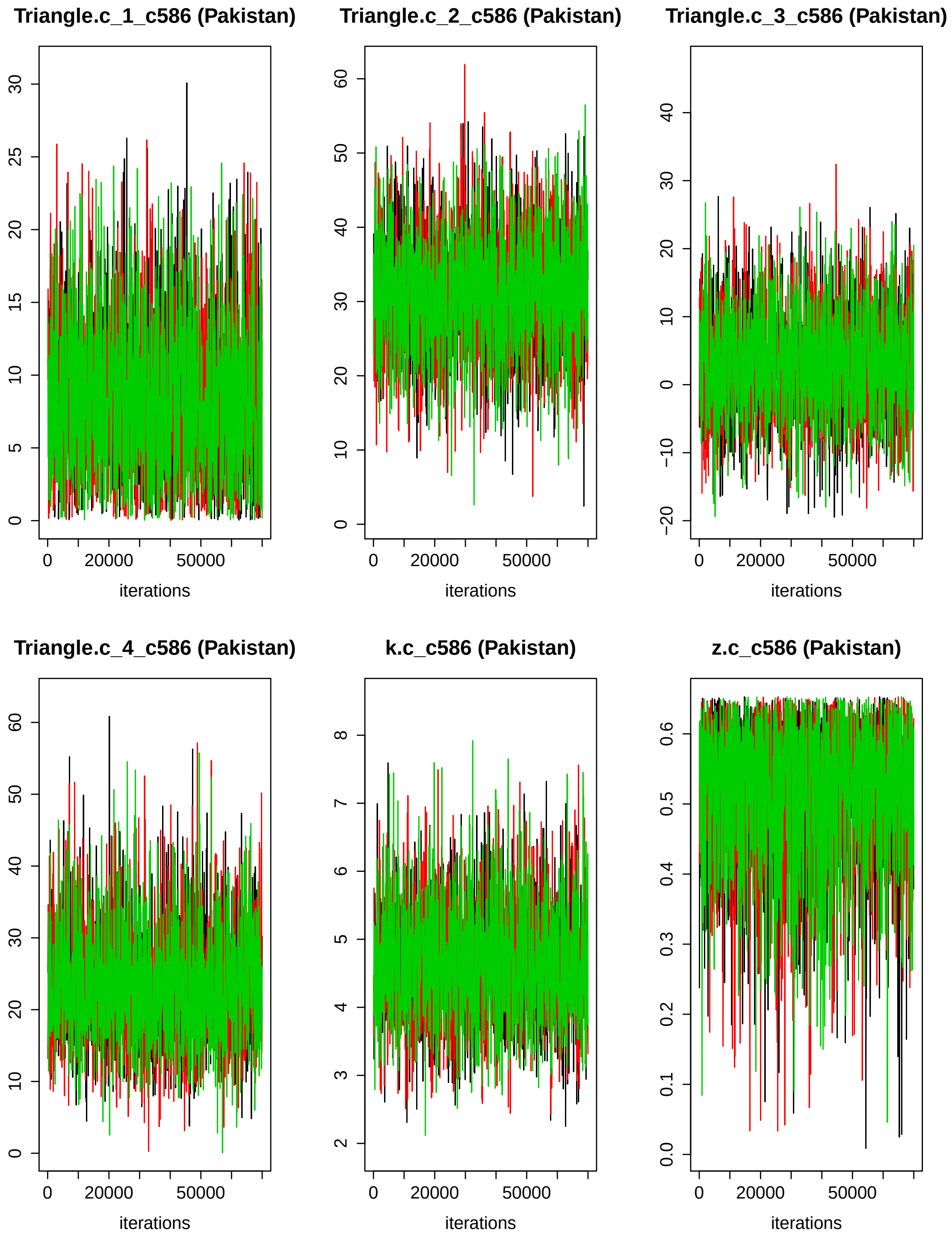


(a) Life Expectancy plot for Pakistan (b) Life Expectancy plot for Punjab



(c) Life Expectancy plot for K.P.K (d) Life Expectancy plot for Balochistan

FIGURE 3.6: Graphs showing Median trajectories of Life Expectancy at birth for both genders along with 80% and 95% projection intervals at national as well as selected regional levels i)Punjab, ii) K.P.K, and iii) Balochistan.



Triangle.c_1_c586 (Pakistan)

Triangle.c_2_c586 (Pakistan)

Triangle.c_3_c586 (Pakistan)

Triangle.c_4_c586 (Pakistan)

k.c_c586 (Pakistan)

z.c_c586 (Pakistan)

FIGURE 3.7: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k^c, iii)z^c at national level.

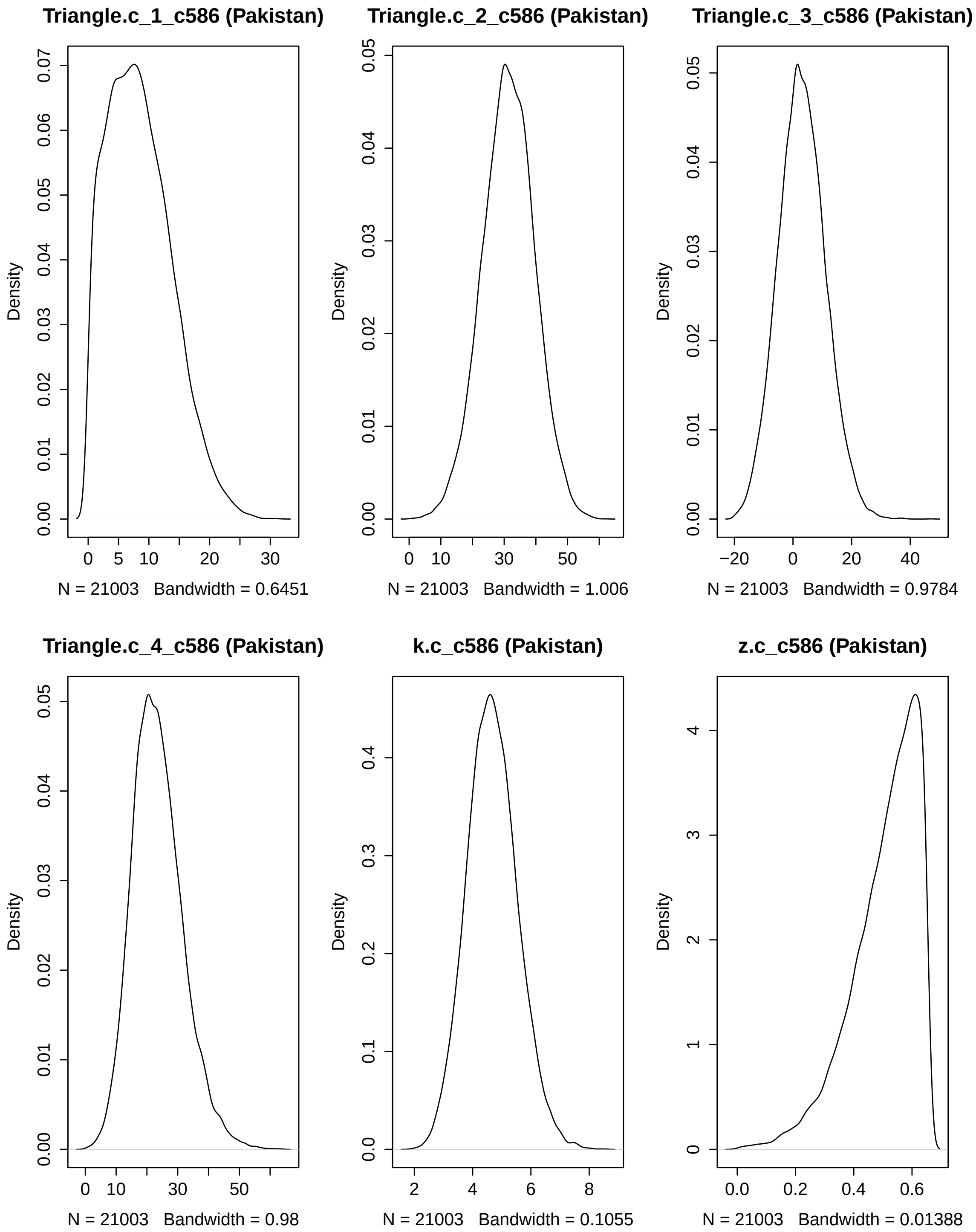


FIGURE 3.8: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k_c, iii)z_c at national level.

3.2.2 Out of Sample Validation for Life Expectancy

An out of sample validation was performed to check the validity of the model fitting as done in the case of TFR by Alkema *et al.* (2011); Raftery *et al.* (2013). The procedure adopted is the same as done previously in the case of Total Fertility Rates.

Data untill 1990	**MSE**	**Above Median**	**Proportion of Obs. Above 95% PI**	**Proportion of Obs. Below 95% PI**
1990 – 1995	0.19	0.61	0.04	0.02
1995 – 2000	0.31	0.55	0.05	0.09
2000 – 2005	0.35	0.43	0.03	0.07
1995 – 2000	0.44	0.45	0.03	0.09
2000 – 2005	0.57	0.52	0.02	0.08
2005 – 2010	0.62	0.47	0.04	0.10
2010 – 2015	0.77	0.39	0.03	0.07

Table 3.8: Out of Sample validation for Life Expectancy: Mean Square error and proportion of left-out UN estimates that fall above the median projected Life Expectancy, and above or below their 95% projection interval in future periods.

Similar to the results of TFR, the resultS of the projection intervals appeared to be well calibrated in the out of sample projections.

3.3 Sensitivity Analysis

Different assumptions are structural part of UN model and we tried to relax some of them and tested how sensitive the model is to those changes. We did two changes, i) by testing different combinations of number of countries and ii) by considering different values of parameters in the autoregressive model. We tested four different combinations of number of countries and tried to examin their impact on the behaviour of model as provided in the table 3.2. Furthermore, we also examined the impact of different values of μ to see how they may affect the behaviour of trajectories in the long run. We concluded that, for slight changes in the above said conditions, the model is not very sensitive.

3.4 Population Projection

The results are generated by taking into account the probabilistic projections of TFR, probabilistic projections of life expectancy for men and women separately and deterministic migration rates[1]. It should be noted that the subnational level population totals when aggregated would not sum up to what has been reported at national level because of missing data for other parts of the country. Due to this reason the projections were carried out independently for each region.

Year	**Pakistan**	**Punjab**	**Sindh**	**KPK**	**Balochistan**
2018	207806	113907	46658	26531	10138.6
2023	226270	123792	51096	28955	11021.5
2028	243861	133051	55212	31318	11858.5
2033	260997	142033	59351	33720	12728.7
2038	277867	151065	63735	36168	13619.0
2043	293888	159807	68089	38457	14424.2
2048	308358	167704	71994	40452	15067.4
2053	320966	174452	75310	42142	15551.9

TABLE 3.9: Population projection at national as well as regional level in thousands.

The results reveal a population of 207 million for the country, which is almost coinciding with the recently revealed provisional results of population census by Pakistan Bureau of Statistics. The major difference in results for the province of KPK is observed which is basically due to the inconsistent data as we did not have complete access to all data sets, another appealing reason might be the influx of Afghan refugees who were counted in the current census, and in the final revision there might be some variation in

---

[1] Note: We are also working to produce probabilistic projections for migration as done by Azose and Raftery (2015), which when done would reveal results similar to what have been produced by Azose *et al.* (2016).

the census figures (Mubarak, 2017). The reason for the Afghan refugees in a big number is due to the ongoing war situation in Afghanistan since 9/11, as a result millions of Afghans moved to Pakistan. However, this number has dropped with the passage of time, but still a significant number of Afghans still exist in Pakistan. Most of them are found in Pushtun dominated areas of Pakistan, specifically in the KPK province, and in the city of Quetta in the northern Balochistan, because these regions provide easy access to the country due to common border between the two countries.

This difference could be minimized in future based on availability of sufficient and reliable data sets. It may be noticed that at national level the results are exactly coinciding with what has been reported in the recent provisional figures because at national level the data sets are of reliable nature in contrast to the data at provincial level. Furthermore, a comparison with deterministic projections of world bank and UN non-Stochastic model is also shown in the table below.

TABLE 3.10: comparison of Population totals (in thousands) with different sources

Sources	Thesis	UN	WB	PBS
Population totals	207806	197016	196744	207774

Here Thesis represents the result obtained based on approach adopted in this thesis, UN [1] represents the Non-Stochastic population totals, WB [2] represents the world bank, while PBS[3] represents the provisonal figures revealed by Pakistan Bureau of Statistics.

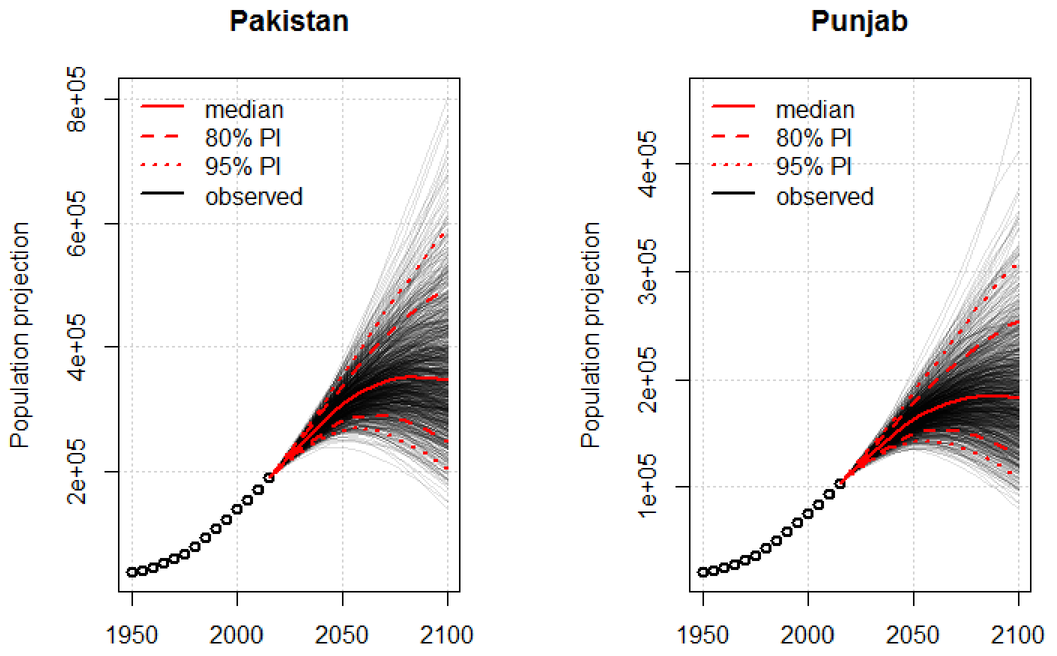
Besides, the graphs of population total trajectories are shown in figure 3.9(a) to figure 3.9(d), while the figure 3.10 shows the pyramid for the projected year 2030. The pyramid also shows the population in the year 2010, along with 80% and 90% credibility intervals. The population of males and females with different credibility intervals is shown in appendix.

---

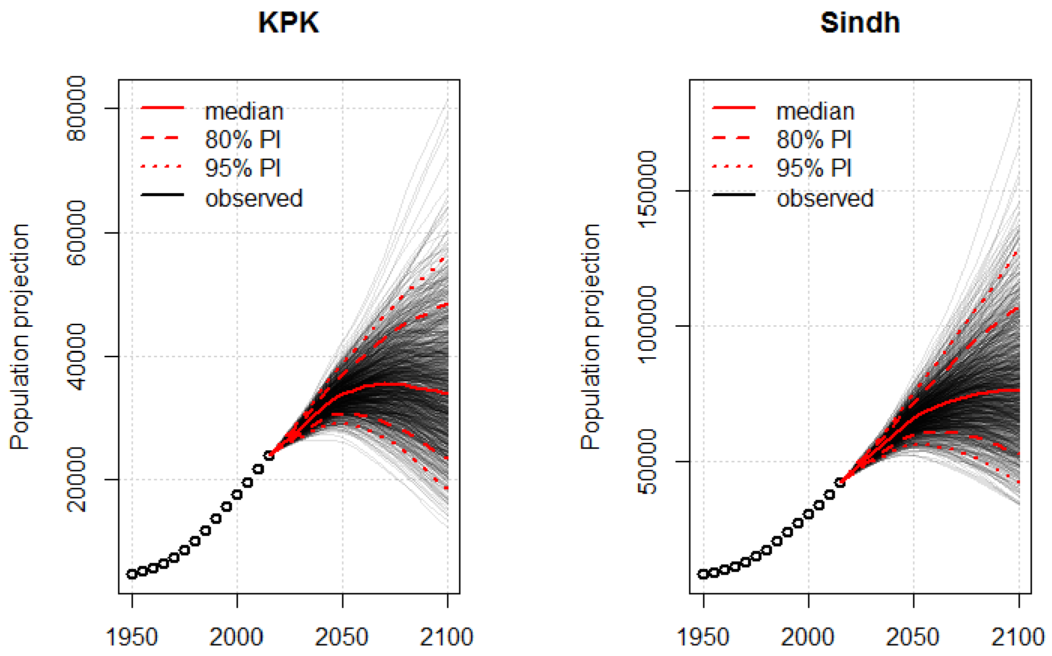
[1] `https://esa.un.org/unpd/wpp/Download/Standard/Population/`

[2] `http://datatopics.worldbank.org/health/population`

[3] `http://www.pbscensus.gov.pk/content/provisional-summary-results-6th-population-and-housing-c`



(a) Population of Pakistan (b) Population of Punjab



(c) Population of KPK (d) Population of Sindh

FIGURE 3.9: Graphs showing Median trajectories of Population totals along with 80% and 95% projection intervals at national as well as selected regional levels i)Punjab, ii) K.P.K, and iii) Sindh.

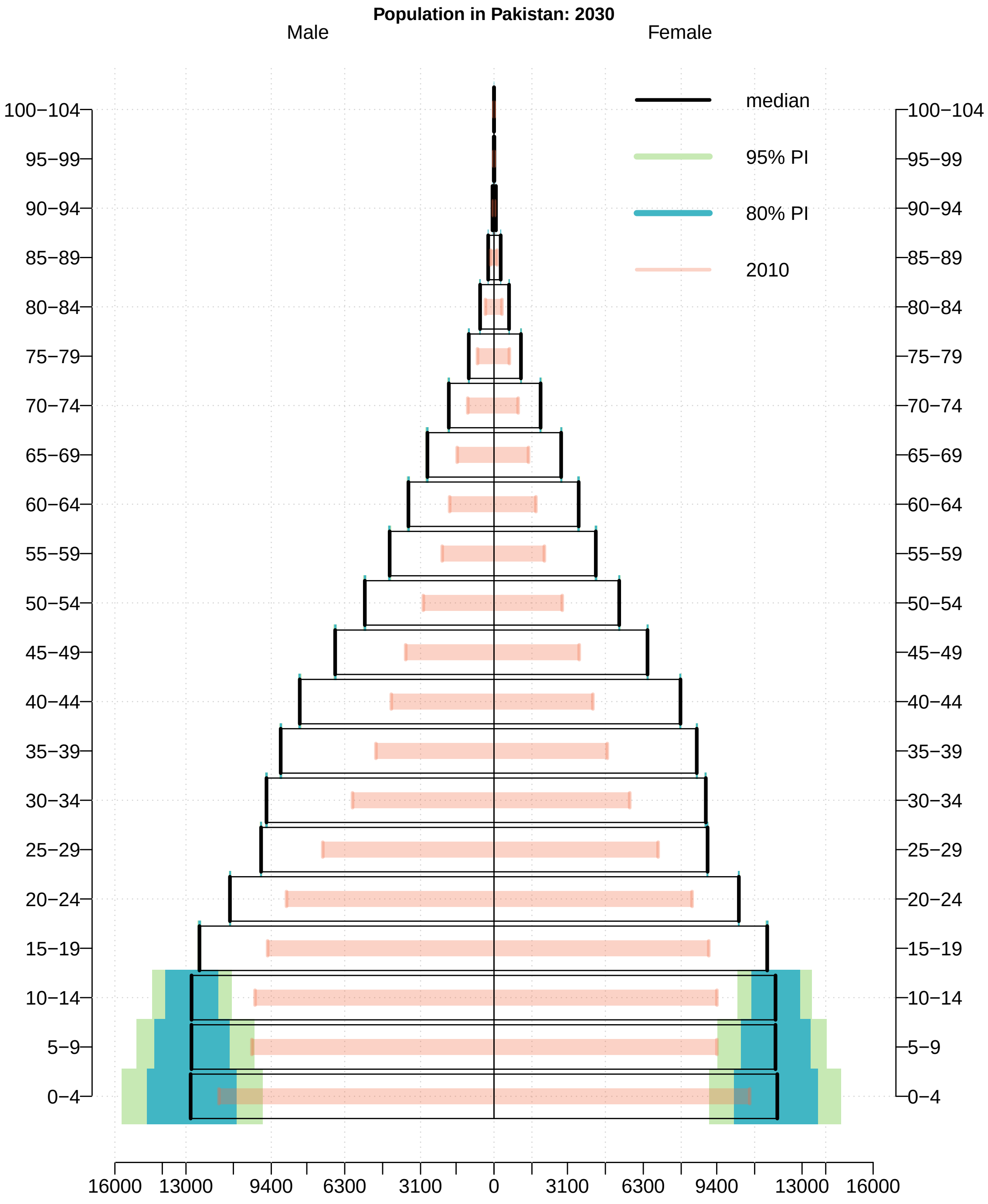


FIGURE 3.10: Graph showing male female projected population pyramid at national level. The pink lines show the population in the year 2010, while blue and green lines show 80% and 90% confidence bounds.

3.4.1 Out of Sample Validation for Population totals

Out of sample validation for Population totals was performed to check the validity of the model fitting as done in the cases of TFR and Life expectancy. The procedure adopted is the same as done previously. The detailed information is given in the table. Similar to the cases of TFR and Life Expectancy, the results for Population toals also suggested a well calibrated projection intervals.

Data untill 1990	**MSE**	**Above Median**	**Proportion of Obs. Above 95% PI**	**Proportion of Obs. Below 95% PI**
1990 – 1995	0.24	0.51	0.04	0.07
1995 – 2000	0.31	0.55	0.05	0.09
2000 – 2005	0.45	0.49	0.03	0.06
1995 – 2000	0.57	0.67	0.03	0.09
2000 – 2005	0.62	0.52	0.02	0.07
2005 – 2010	0.66	0.47	0.04	0.06
2010 – 2015	0.77	0.49	0.03	0.10

Table 3.11: Out of Sample validation for Population toals : Mean Square error (MSE) and and proportion of left-out UN estimates that fall above the median projected Population total, and above or below their 95% projection interval in future periods.

Chapter 4

Conclusions

We have obtained probabilistic projections for fertility, life expectancy at birth for males and females and population totals. UN has produced in the past trajectories based on deterministic approach, and recently due to the Alkema *et al.* (2011); Raftery *et al.* (2012, 2013, 2014b), revised its methodology and started producing probabilistic projections. We tried to extend the same concept to regional levels and produced the probabilistic projections at sub-national/regional levels in Pakistan. The importance of probabilistic projections can not be ignored, because not only they give us the best possible trajectories by taking into account the uncertainty factor, but also give much reliable credibility intervals.

Regional level forecasts play vital role in deciding the policy measures. The estimates of TFR are based on demographic surveys, however, the frequency of these survey is not consistent and time bound, as a result policy makers mostly rely on forecasts. But deterministic forecasts do not ascertain the uncertainty factor that may alter the results. We tried to overcome this part and produced these forecasts by taking the same concept of Alkema *et al.* (2011) with some modifications based on available data.

We produced sub-national level future forecasts for life expectancy by extending the concept of Raftery *et al.* (2013, 2014b) to sub-national level. The sub-national level projections for life expectancy for Pakistan are not available which could take into account the uncertainty factor as well. This would be a helping result for the authorities who are responsible for making future plans and strategies in the field of health and other related areas.

The sixth population census in the country just finished whose provisional results are released and are nearly similar to what we produced based on probabilistic approach. According to the web content of National Institute of Population Studies (NIPS)[1], the results are far different than what have been observed in the recent provisional census, while our projected results are quite near to the provisionally released figures.[2]

4.1 Discussion

The TFR projections for the regions of Pakistan are generated by taking into account the classical AR(1) approach as well as Bayesian Hierarchical modeling (BHM) approach, adopted in phase III of TFR model prediction. We tried to check the performance of model based on different values of μ (replacement level of fertility) for which the most generalized value for entire globe is assumed to be 2.1 children and concluded that for values larger than 2.0 the conversion takes long time than the smaller values. We also observed in our analysis that the results for BHM which has the quality to suggest ultimate level of replacement based on the pattern of data for specific region, were quite different from the classical approach.

As there were not sufficient decrements observed for any particular region, so BHM approach borrows information from the entire globe to support the analysis. We tried to investigate if we could restrict this influence by minimizing the number of countries in the model, and observed that for a specific region, the best information comes from the neighbouring source, thus it was better to restrict the number of borrowers to that continent only where the country exist, which in result saves time of computation as well as reduces the influence of other countries.

---

[1] NIPS is one of the main organization in Pakistan for producing population projections.

[2] It should be noticed that final results are under preparation and might be slightly different than the provisional results.

We made adjustment in the prior distribution ranges based on our regions information, because the ranges provided for the model were generalized for entire globe, however, region to region the scenario might be different. This in result provided much concise results than the most generalized one. We also observed that for the case of life expectancy the gap in general is not narrowing and is predicted to increase in the long run. At the current time point the gap between male and female life expectancy at national as well as regional level with the exception of Balochistan region is between 2 to 3 years, where as for the Balochistan region it is 4 years, which in future is expected to reach to 5 years.

One big challenge we faced was pertaining to the availability of data at regional level specifically for migration. To overcome this we made an assumption and redistributed the net migration rates at national level reported in the world population prospect in to regions based on the proportionate distribution reported in the labour force survey. These results were combined with the probabilistic trajectories of TFR and Life Expectancy to generate future population projections.

4.2 Future directions of research

We are currently working to extend this model to fully probabilistic one by projecting the migration rates in the similar way and then to incorporate them in the model as done by Azose *et al.* (2016). The migration rates used in the thesis are deterministic, and do neglect a substantial rate of uncertainty. This issue would be resolved by extending and modifying the approach of Azose and Raftery (2015) to sub-national level, so as to overcome the missing element of uncertainty in migration estimates.

One more step which we want to do is to estimate the parameters of joint probabilistic model also by Bayesian Hierarchical modeling approach, because these parameters are estimated based on maximum likelihood approach for the sake of ease. This raises the question that, the estimates obtained at each step do not take into account the uncertainty part. We would like to implement BHM approach at this step also to avoid the uncertainty part.

We would extend this work by incorporating more sophisticated approaches like Alkema *et al.* (2008) to deal with data problems and to get more precise results than what we have. The question of data quality are of greater value and we would try to answer this part by implementing the approach adopted by Alkema *et al.* (2008) to compare if there is any discrepancy in the results. BHM approach is gaining much more importance these days and can work efficiently for unreliable or missing data sets (Bijak and Bryant, 2016). We would take into account these approaches to resolve data relavent issue.

Social and economic variable like education, inflation, unemployment etc., are also very important in determining the demographic characteristics who's effect if studied would give a further insight into the problem of high fertility and mortality rates. We would add the hierarchies of these characteristics in the model and investigate their impact on the outcomes of demographic components. The study of inflation would be interesting because, for the capital Islamabd, the most expensive city for living has high consumer price index (CPI) and least TFR, while Balochistan has high TFR and low CPI in Quetta city, which is the capital of Balochistan province (NIPS, 2012; Government of Pakistan, August, 2017). An insight in to this factor could be helpful for policy makers.

Appendix

The following table describes the trajectories of TFR for Pakistan and its regions separately together with 75, 80 and 95 percent confidence intervals.

TFR Trajectories Tables

Year	**Median**	**0.025**	**0.10**	**0.125**	**0.875**	**0.90**	**0.975**
2018	3.257	2.629	2.859	2.912	3.607	3.642	3.840
2023	2.990	2.249	2.534	2.576	3.404	3.444	3.667
2028	2.790	1.973	2.272	2.334	3.234	3.279	3.528
2033	2.607	1.790	2.091	2.140	3.077	3.141	3.420
2038	2.470	1.612	1.911	1.974	2.970	3.016	3.307
2043	2.347	1.465	1.787	1.850	2.855	2.912	3.225
2048	2.249	1.330	1.676	1.739	2.750	2.818	3.126
2053	2.155	1.236	1.566	1.623	2.662	2.716	3.080
2058	2.066	1.140	1.479	1.534	2.584	2.657	2.999
2063	1.994	1.044	1.408	1.459	2.522	2.586	2.945
2068	1.948	1.013	1.349	1.412	2.451	2.515	2.865
2073	1.910	0.971	1.303	1.361	2.388	2.451	2.802
2078	1.881	0.955	1.262	1.341	2.321	2.402	2.712
2083	1.859	0.954	1.265	1.342	2.284	2.337	2.710
2088	1.853	0.948	1.264	1.329	2.242	2.298	2.658
2093	1.840	0.957	1.274	1.332	2.222	2.271	2.593
2098	1.836	0.960	1.284	1.345	2.191	2.243	2.564

TABLE .1: TFR Trajectories for Pakistan starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval based on Bayesian hierarchical approach.

Year	**Median**	**0.025**	**0.10**	**0.125**	**0.875**	**0.90**	**0.975**
2018	3.377	2.911	3.080	3.112	3.640	3.674	3.846
2023	3.100	2.507	2.716	2.756	3.455	3.496	3.706
2028	2.872	2.219	2.433	2.475	3.280	3.330	3.597
2033	2.702	1.976	2.223	2.265	3.137	3.198	3.481
2038	2.538	1.799	2.041	2.093	3.006	3.071	3.362
2043	2.401	1.630	1.894	1.955	2.882	2.947	3.286
2048	2.300	1.503	1.785	1.837	2.782	2.847	3.211
2053	2.205	1.399	1.685	1.732	2.706	2.762	3.121
2058	2.132	1.326	1.604	1.649	2.627	2.696	3.035
2063	2.070	1.254	1.535	1.577	2.570	2.631	2.974
2068	2.014	1.182	1.469	1.529	2.507	2.560	2.922
2073	1.963	1.114	1.441	1.495	2.438	2.491	2.843
2078	1.944	1.094	1.397	1.459	2.377	2.432	2.803
2083	1.925	1.061	1.381	1.445	2.345	2.396	2.711
2088	1.922	1.043	1.378	1.441	2.321	2.371	2.672
2093	1.919	1.002	1.400	1.455	2.297	2.339	2.661
2098	1.921	0.995	1.405	1.476	2.289	2.336	2.602

Table .2: TFR Trajectories for Pakistan starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval based on classical AR(1) approach, by considering replacement level of 2.1.

The following tables describes TFR trajectories for the four provinces namely, Punjab, Sindh, KPK and Balochistan. The trajectories were generated based on Bayesian Hierarchical (BHM)approach. Interesting thing was to see different behaviour of trajectories based on classical AR(1) approach. As in the case of classical approach in phase III, we were free to choose replacement level values based on our own judgment, and different combinations of variable number of countries. It was shown when entire world level data was used, then the results of Punjab, Sindh and KPK shown a bit similarity to national level, while Balochistan shown different result. However, based on BHM approach, there was no such similar pattern as observed with classical approach.

Year	Median	0.025	0.10	0.125	0.875	0.90	0.975
2018	3.133	2.547	2.760	2.795	3.452	3.493	3.682
2023	2.883	2.240	2.447	2.494	3.254	3.295	3.537
2028	2.673	1.945	2.219	2.264	3.086	3.144	3.390
2033	2.498	1.773	2.017	2.067	2.956	3.004	3.261
2038	2.359	1.598	1.848	1.897	2.823	2.870	3.159
2043	2.233	1.420	1.703	1.770	2.717	2.765	3.048
2048	2.150	1.278	1.597	1.662	2.611	2.678	2.972
2053	2.048	1.186	1.501	1.567	2.533	2.590	2.897
2058	1.978	1.096	1.423	1.482	2.445	2.502	2.833
2063	1.935	1.014	1.349	1.412	2.388	2.441	2.745
2068	1.905	0.972	1.305	1.358	2.322	2.375	2.659
2073	1.873	0.935	1.263	1.338	2.266	2.314	2.605
2078	1.860	0.927	1.248	1.307	2.238	2.283	2.562
2083	1.851	0.902	1.261	1.330	2.211	2.260	2.480
2088	1.851	0.912	1.276	1.348	2.182	2.224	2.471
2093	1.853	0.922	1.295	1.372	2.170	2.209	2.445
2098	1.852	0.910	1.321	1.393	2.150	2.190	2.425

TABLE .3: TFR Trajectories for Punjab province starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.125**	**0.875**	**0.90**	**0.975**
2018	3.574	3.004	3.186	3.225	3.925	3.959	4.153
2023	3.300	2.615	2.859	2.909	3.714	3.761	3.988
2028	3.079	2.320	2.589	2.642	3.523	3.575	3.858
2033	2.896	2.099	2.379	2.439	3.371	3.415	3.690
2038	2.737	1.915	2.199	2.252	3.238	3.299	3.590
2043	2.600	1.761	2.026	2.083	3.121	3.184	3.513
2048	2.483	1.613	1.893	1.954	3.014	3.070	3.398
2053	2.357	1.458	1.774	1.833	2.905	2.965	3.284
2058	2.256	1.374	1.687	1.744	2.815	2.880	3.213
2063	2.172	1.293	1.571	1.635	2.733	2.797	3.135
2068	2.107	1.194	1.512	1.582	2.655	2.726	3.069
2073	2.046	1.147	1.463	1.536	2.582	2.659	2.998
2078	2.008	1.096	1.441	1.505	2.513	2.590	2.937
2083	1.963	1.091	1.410	1.472	2.449	2.535	2.887
2088	1.932	1.045	1.385	1.452	2.405	2.459	2.813
2093	1.925	1.014	1.367	1.441	2.363	2.421	2.743
2098	1.915	0.997	1.365	1.440	2.320	2.377	2.736

TABLE .4: TFR Trajectories for Sindh province starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.125**	**0.875**	**0.90**	**0.975**
2018	3.470	2.871	3.074	3.115	3.815	3.860	4.063
2023	3.193	2.494	2.732	2.777	3.593	3.636	3.855
2028	2.950	2.213	2.460	2.510	3.389	3.441	3.702
2033	2.755	1.988	2.258	2.304	3.220	3.272	3.555
2038	2.584	1.777	2.056	2.111	3.070	3.129	3.446
2043	2.436	1.598	1.873	1.934	2.935	3.000	3.311
2048	2.312	1.442	1.746	1.809	2.829	2.881	3.201
2053	2.193	1.307	1.637	1.695	2.718	2.788	3.099
2058	2.106	1.185	1.538	1.599	2.630	2.680	3.039
2063	2.040	1.104	1.457	1.530	2.541	2.590	2.956
2068	1.982	1.065	1.399	1.457	2.470	2.531	2.840
2073	1.939	1.014	1.353	1.423	2.397	2.441	2.814
2078	1.911	0.975	1.304	1.391	2.331	2.385	2.750
2083	1.890	0.925	1.283	1.363	2.295	2.350	2.678
2088	1.880	0.921	1.284	1.353	2.261	2.313	2.666
2093	1.872	0.938	1.294	1.364	2.237	2.281	2.590
2098	1.866	0.963	1.310	1.386	2.205	2.245	2.530

TABLE .5: TFR Trajectories for K.P.K province starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.125**	**0.875**	**0.90**	**0.975**
2018	3.199	2.646	2.845	2.875	3.504	3.544	3.731
2023	2.892	2.279	2.493	2.535	3.248	3.288	3.494
2028	2.652	2.002	2.195	2.242	3.047	3.081	3.298
2033	2.452	1.726	1.975	2.025	2.865	2.915	3.131
2038	2.267	1.537	1.777	1.833	2.702	2.750	3.007
2043	2.127	1.331	1.613	1.671	2.585	2.632	2.889
2048	2.009	1.175	1.461	1.521	2.456	2.508	2.799
2053	1.923	1.055	1.337	1.391	2.359	2.414	2.698
2058	1.860	0.978	1.252	1.310	2.282	2.325	2.582
2063	1.810	0.880	1.194	1.255	2.239	2.288	2.524
2068	1.784	0.840	1.142	1.218	2.182	2.222	2.473
2073	1.791	0.833	1.140	1.196	2.159	2.197	2.436
2078	1.789	0.821	1.129	1.208	2.138	2.177	2.384
2083	1.795	0.803	1.169	1.225	2.128	2.169	2.377
2088	1.808	0.804	1.182	1.265	2.123	2.169	2.353
2093	1.814	0.824	1.219	1.309	2.122	2.163	2.345
2098	1.827	0.902	1.288	1.365	2.119	2.157	2.328

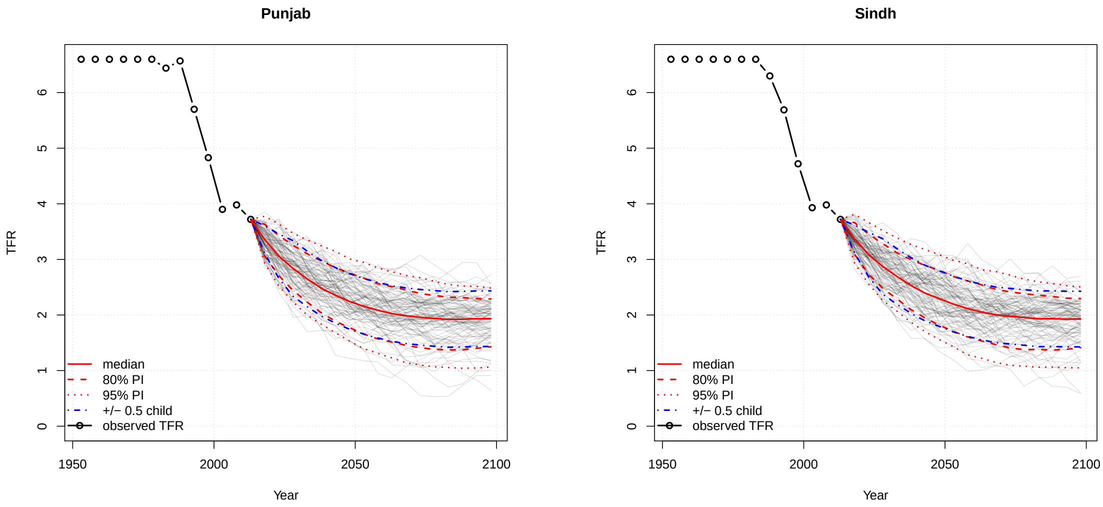
Table .6: TFR Trajectories for Balochistan province starting from 2018 to 2098, with median, 75%, 80%, 95% confidence interval.

Year	mean	SD	RSD	2.50%	5%	10%	25%	50%	75%	90%	95%	97.50%
2018	3.25	0.307	0.094	2.63	2.74	2.86	3.05	3.26	3.47	3.64	3.75	3.84
2023	2.99	0.361	0.121	2.25	2.40	2.53	2.76	2.99	3.24	3.44	3.56	3.67
2028	2.78	0.398	0.143	1.97	2.12	2.27	2.52	2.79	3.05	3.28	3.42	3.53
2033	2.61	0.415	0.159	1.79	1.94	2.09	2.34	2.61	2.90	3.14	3.30	3.42
2038	2.47	0.433	0.175	1.61	1.75	1.91	2.19	2.47	2.75	3.02	3.17	3.31
2043	2.35	0.444	0.189	1.47	1.61	1.79	2.06	2.35	2.64	2.91	3.10	3.22
2048	2.24	0.451	0.201	1.33	1.49	1.68	1.95	2.25	2.54	2.82	2.99	3.13
2053	2.14	0.459	0.214	1.24	1.38	1.57	1.84	2.15	2.45	2.72	2.89	3.08
2058	2.06	0.464	0.225	1.14	1.29	1.48	1.76	2.07	2.36	2.66	2.82	3.00
2063	2.00	0.464	0.232	1.04	1.22	1.41	1.70	1.99	2.29	2.59	2.78	2.94
2068	1.94	0.458	0.236	1.01	1.16	1.35	1.66	1.95	2.23	2.51	2.68	2.87
2073	1.90	0.453	0.238	0.97	1.12	1.30	1.62	1.91	2.18	2.45	2.65	2.80
2078	1.86	0.442	0.238	0.96	1.09	1.26	1.59	1.88	2.13	2.40	2.59	2.71
2083	1.84	0.430	0.234	0.95	1.09	1.27	1.57	1.86	2.09	2.34	2.53	2.71
2088	1.82	0.419	0.230	0.95	1.09	1.26	1.55	1.85	2.07	2.30	2.48	2.66
2093	1.81	0.406	0.224	0.96	1.09	1.27	1.56	1.84	2.05	2.27	2.42	2.59
2098	1.80	0.394	0.219	0.96	1.10	1.28	1.57	1.84	2.05	2.24	2.39	2.56

TABLE .7: Table showing TFR along with S.D, Relative Standar Deviation (also known as CV, a measure of reliablity) and confidence limits for Pakistan

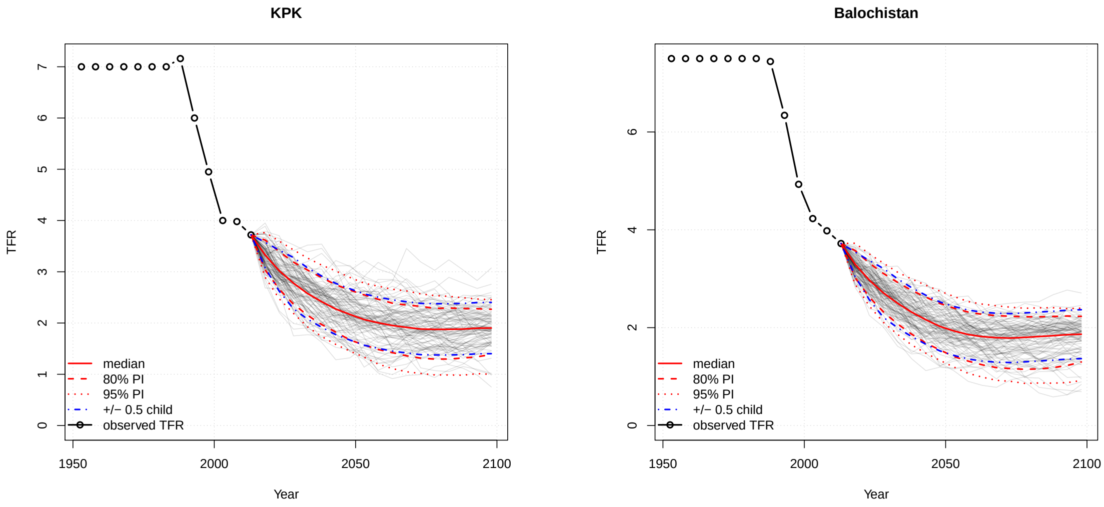
Note: RSD is given by the formula: $\frac{SD}{mean}$, while CV is given by $\frac{SD}{mean}X100$

TFR Graphs



(a) TFR Trajectories for Punjab province

(b) TFR Trajectories for Sindh province

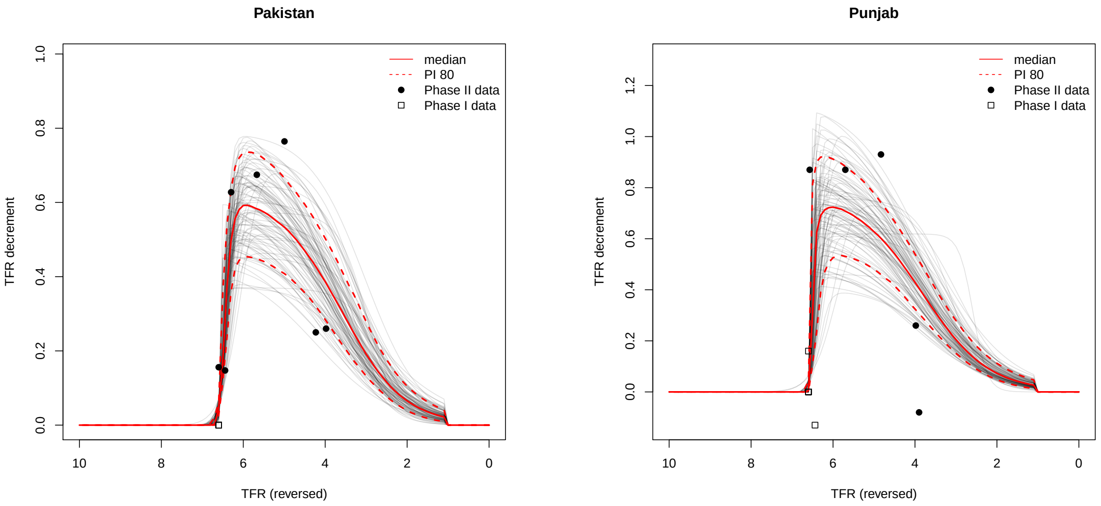


(c) TFR Trajectories for K.P.K province

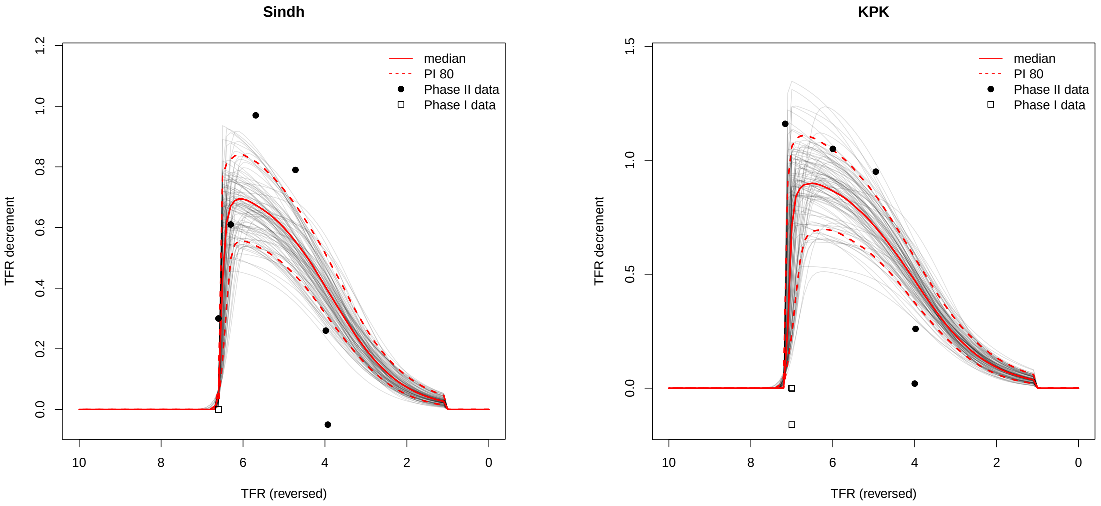
(d) TFR Trajectories for Balochistan

Figure .1: Graphs showing TFR trajectories at regional level for i)Punjab, ii) Sindh, iii) K.P.K, and iv) Balochistan

Logistic Curves



(a) DL Curve for Pakistan (b) DL Curve for Punjab



(c) DL Curve for Sindh (d) DL Curve for KPK

FIGURE .2: Graphs showing Double Logistic Curve at national as well as regional level.

Variable	**Mean**	**SD**	**Naive SE**	**Time-series SE**
δ_1	1.06948	0.2375	0.001651	0.0108994
δ_2	1.19024	0.339245	0.002358	0.0236619
δ_3	0.89321	0.210147	0.001461	0.010695
Δ_4	0.38122	0.302514	0.002103	0.0101045
δ_4	1.38436	0.500135	0.003476	0.0296386
ψ	1.11913	0.089238	0.000620	0.0016839
χ	-1.92307	0.121915	0.000847	0.0046063
a	0.03654	0.02584	0.000180	0.0036116
b	0.01545	0.009015	0.000063	0.0011015
c_{1975}	1.63271	0.058388	0.000406	0.002205
S	4.87005	1.165272	0.008099	0.1859473
σ_0	0.20999	0.00925	0.000064	0.0002655
m_τ	-0.22002	0.020408	0.000142	0.0001529
s_τ	0.30523	0.014573	0.000101	0.0001011
U_{Pun}	6.5700	0.0000	0.0000	0.0000
d_{Pun}	0.17056	0.035931	0.000250	0.0004231
Δ_{Pun4}	1.99742	0.352932	0.002453	0.0039579
p_{pun1}	0.06159	0.054994	0.000382	0.0004965
p_{pun2}	0.14336	0.125155	0.000870	0.0020596
p_{pun3}	0.79505	0.133827	0.000930	0.0017782
Δ_{Pun1}	0.28006	0.248442	0.001727	0.0022156
Δ_{Pun2}	0.65428	0.576048	0.004004	0.0094225
Δ_{Pun3}	3.63824	0.691046	0.004803	0.0089284

Table .8: Empirical mean and standard deviation for each variable at provincial level (Punjab), plus standard error of the mean

Variable	**Mean**	**SD**	**Naive SE**	**Time-series SE**
δ_1	1.06948	0.2375	0.00165	0.0108994
δ_2	1.19024	0.339245	0.00236	0.0236619
δ_3	0.89321	0.210147	0.00146	0.010695
Δ_4	0.38122	0.302514	0.00210	0.0101045
δ_4	1.38436	0.500135	0.00348	0.0296386
ψ	1.11913	0.089238	0.00062	0.0016839
χ	-1.92307	0.121915	0.00085	0.0046063
a	0.03654	0.02584	0.00018	0.0036116
b	0.01545	0.009015	0.00006	0.0011015
c_{1975}	1.63271	0.058388	0.00041	0.002205
S	4.87005	1.165272	0.00810	0.1859473
σ_0	0.20999	0.00925	0.00006	0.0002655
m_τ	-0.22002	0.020408	0.00014	0.0001529
s_τ	0.30523	0.014573	0.00010	0.0001011
U_{Sind}	6.60000	0.00000	0.00000	0.00000
d_{Sind}	0.13228	0.027779	0.00019	0.0004243
Δ_{Sind4}	1.97283	0.361643	0.00251	0.0041991
p_{Sind1}	0.03934	0.031913	0.00022	0.0002665
p_{Sind2}	0.13001	0.134339	0.00093	0.0019766
p_{Sind3}	0.83065	0.138812	0.00096	0.0018906
Δ_{Sind1}	0.18103	0.145745	0.00101	0.0012074
Δ_{Sind2}	0.60236	0.629957	0.00438	0.0092069
Δ_{Sind3}	3.84377	0.715423	0.00497	0.0090043

Table .9: Empirical mean and standard deviation for each variable at provincial level (Sindh), plus standard error of the mean

Variable	**Mean**	**SD**	**Naive SE**	**Time-series SE**
δ_1	1.06948	0.23750	0.00165	0.010899
δ_2	1.19024	0.339245	0.00236	0.023662
δ_3	0.89321	0.210147	0.00146	0.010695
Δ_4	0.38122	0.302514	0.00210	0.010105
δ_4	1.38436	0.500135	0.00348	0.029639
ψ	1.11913	0.089238	0.00062	0.001684
χ	-1.92307	0.121915	0.00085	0.004606
a	0.03654	0.02584	0.00018	0.003612
b	0.01545	0.009015	0.00006	0.001102
c_{1975}	1.63271	0.058388	0.00041	0.002205
S	4.87005	1.165272	0.00810	0.185947
σ_0	0.20999	0.00925	0.00006	0.000266
m_τ	-0.22002	0.020408	0.00014	0.000153
s_τ	0.30523	0.014573	0.00010	0.000101
U_{KPK}	7.16000	0.00000	0.00000	0.000000
d_{KPK}	0.16778	0.037787	0.00026	0.000580
Δ_{KPK4}	2.00279	0.350938	0.00244	0.004110
p_{KPK1}	0.05864	0.052469	0.00036	0.000471
p_{KPK2}	0.13566	0.124189	0.00086	0.001761
p_{KPK3}	0.8057	0.133309	0.00093	0.001542
Δ_{KPK1}	0.30145	0.268843	0.00187	0.002400
Δ_{KPK2}	0.70039	0.649696	0.00452	0.008860
Δ_{KPK3}	4.15537	0.750391	0.00522	0.008192

TABLE .10: Empirical mean and standard deviation for each variable at provincial level (K.P.K), plus standard error of the mean

Variable	**Mean**	**SD**	**Naive SE**	**Time-series SE**
δ_1	1.069481	0.237500	0.001651	0.010900
δ_2	1.190242	0.339245	0.002358	0.023660
δ_3	0.893209	0.210147	0.001461	0.010700
Δ_4	0.381220	0.302514	0.002103	0.010100
δ_4	1.384361	0.500135	0.003476	0.029640
ψ	1.119130	0.089238	0.000620	0.001684
χ	1.923068	0.121915	0.000847	0.004606
a	0.036539	0.025840	0.000180	0.003612
b	0.015451	0.009015	0.000063	0.001102
c_{1975}	1.632710	0.058388	0.000406	0.002205
S	4.870047	1.165272	0.008099	0.185900
σ_0	0.209987	0.009250	0.000064	0.000266
m_τ	0.220017	0.020408	0.000142	0.000153
s_τ	0.305234	0.014573	0.000101	0.000101
U_{Bal}	7.500000	0.000000	0.000000	0.000000
d_{Bal}	0.260823	0.031671	0.000220	0.000484
Δ_{Bal4}	2.079149	0.313952	0.002182	0.004048
p_{Bal1}	0.009082	0.004361	0.000030	0.000053
p_{Bal2}	0.079311	0.074397	0.000517	0.001552
p_{Bal3}	0.911608	0.074447	0.000517	0.001532
Δ_{Bal1}	0.048937	0.023022	0.000160	0.000277
Δ_{Bal2}	0.427787	0.399633	0.002778	0.007983
Δ_{Bal3}	4.944127	0.518161	0.003601	0.009001

TABLE .11: Empirical mean and standard deviation for each variable at provincial level (Balochistan), plus standard error of the mean

Trace Plots for TFR at regional level

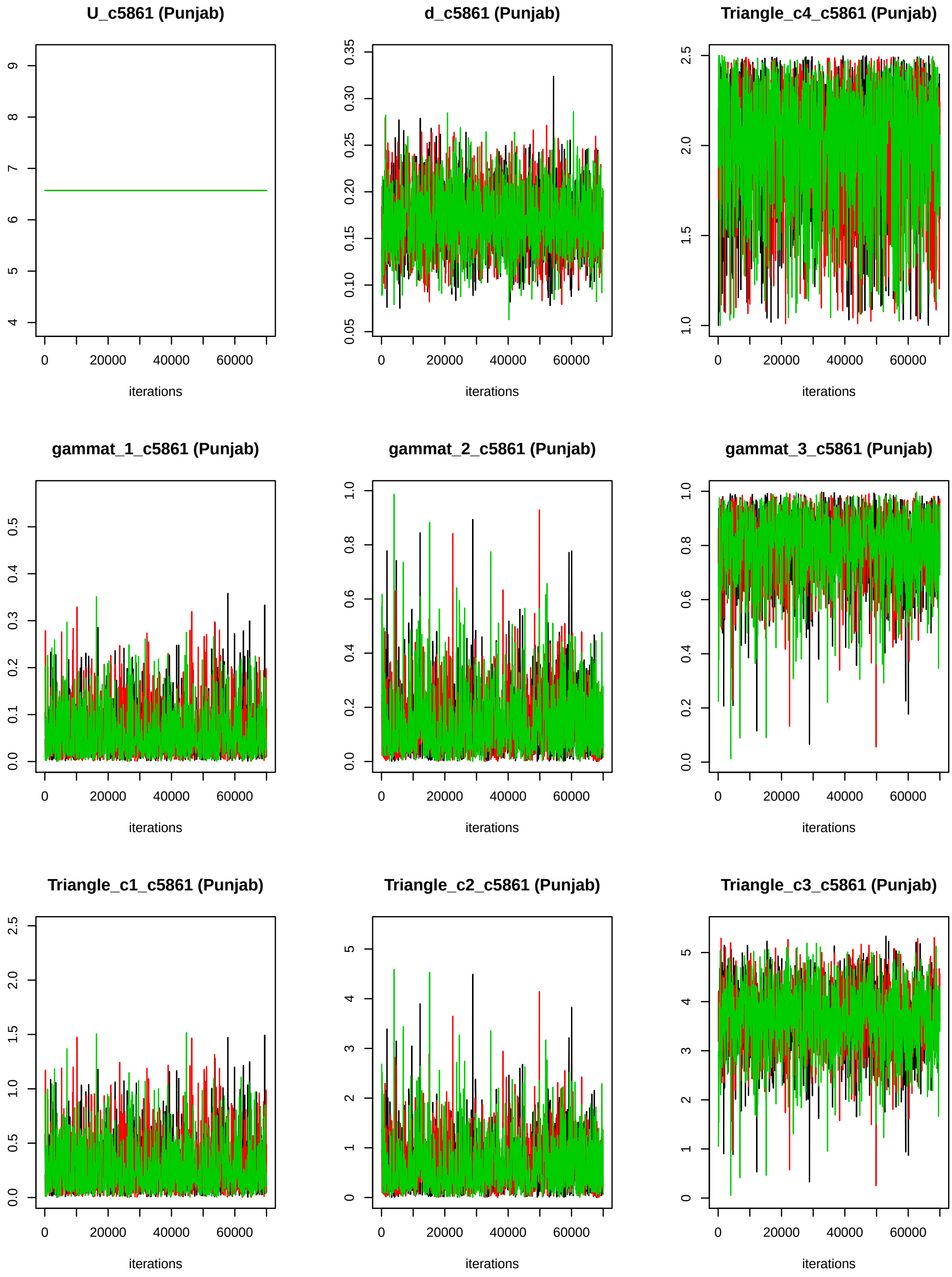


FIGURE .3: Trace Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Punjab.

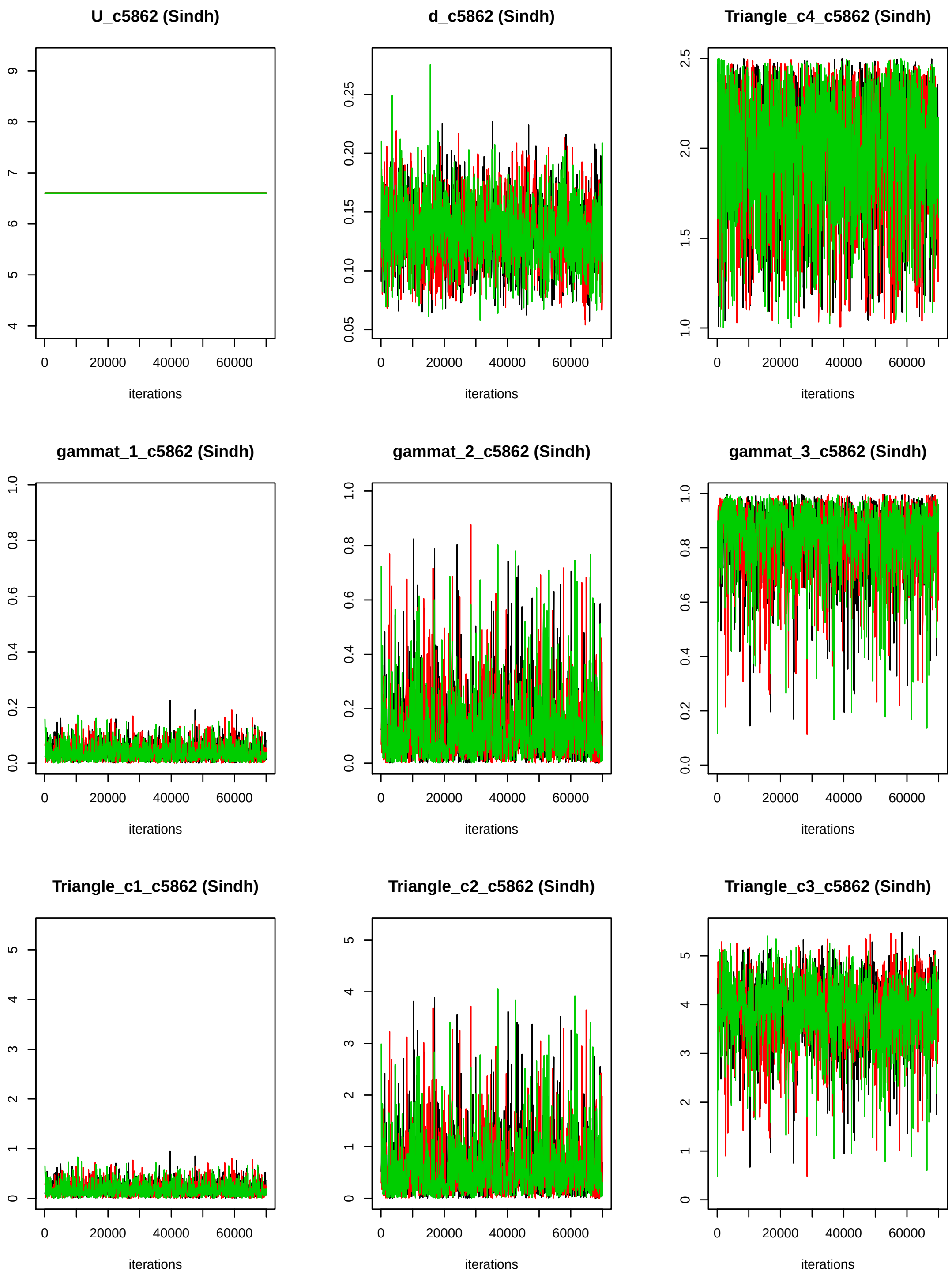


FIGURE .4: Trace Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Sindh.

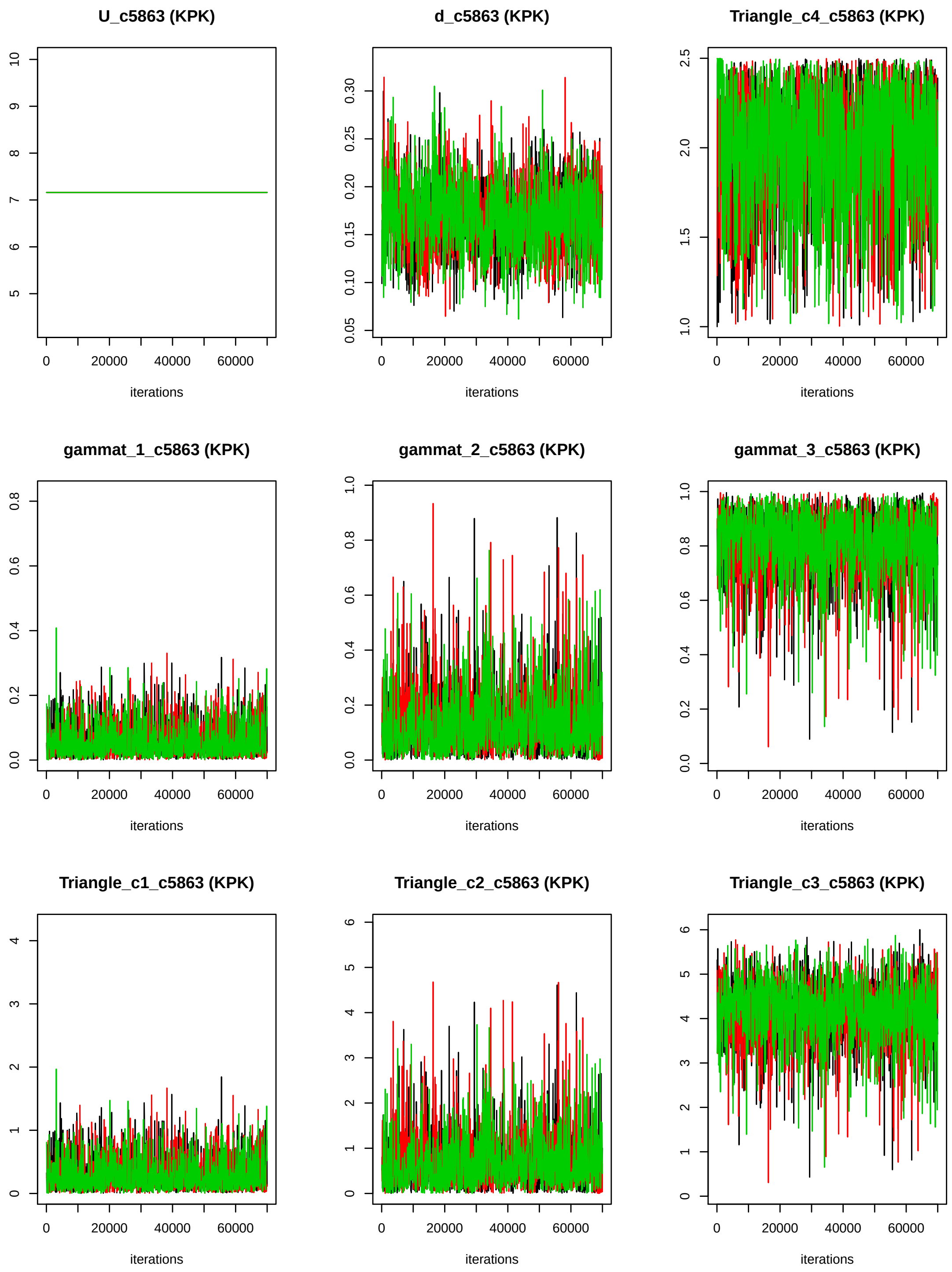


FIGURE .5: Trace Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for KPK.

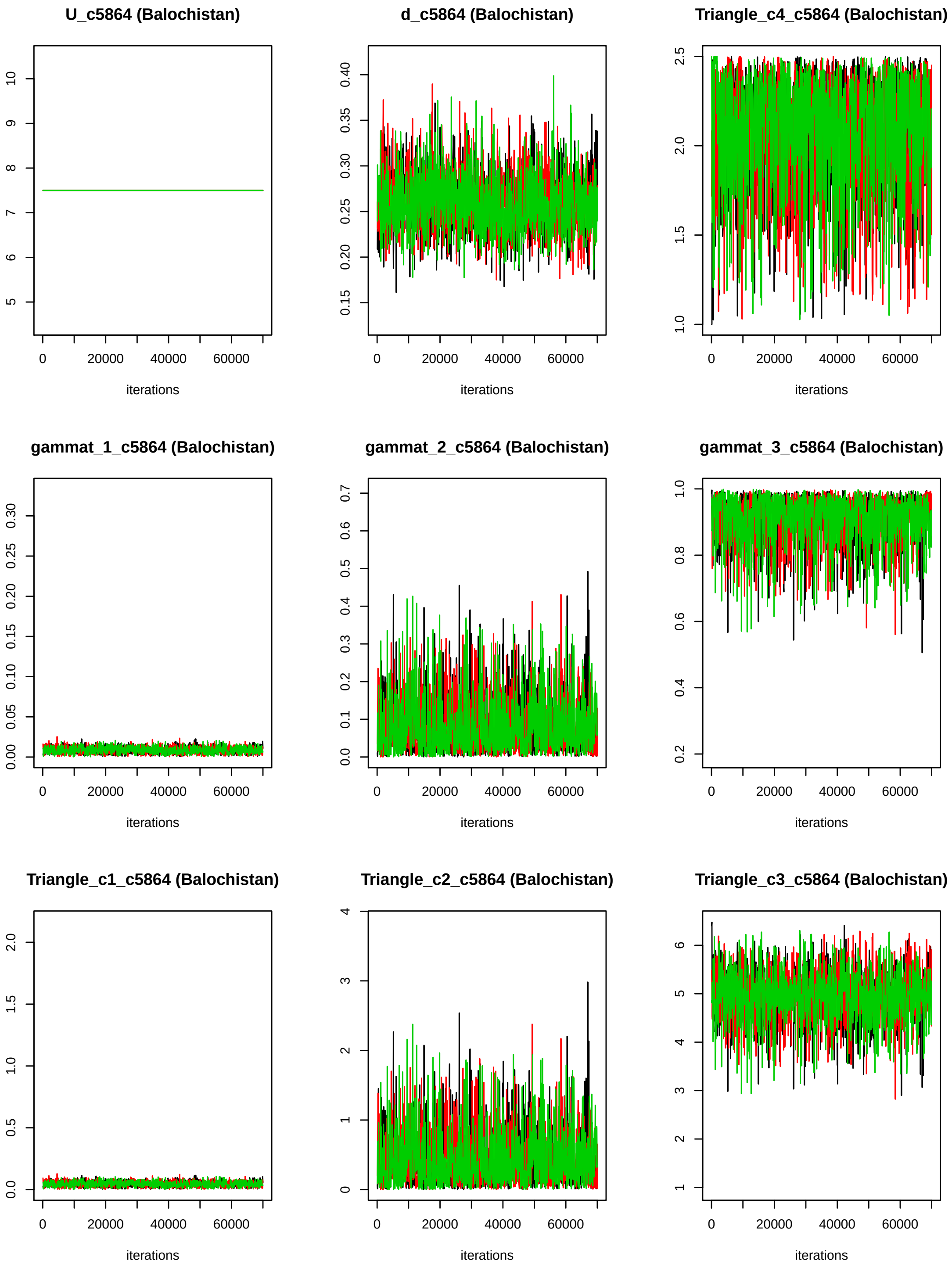


FIGURE .6: Trace Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Balochistan.

Density Plot for TFR at regional level

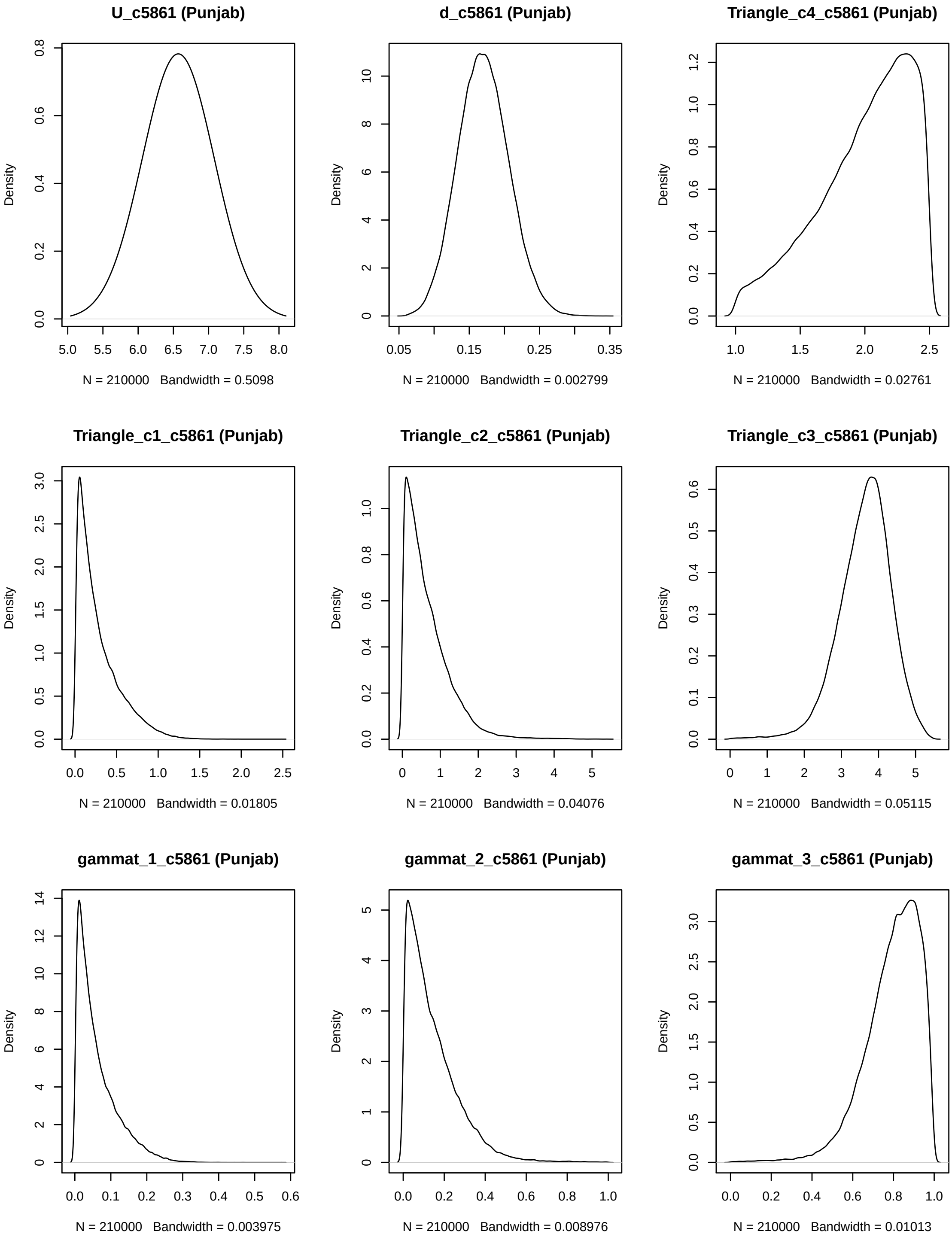


FIGURE .7: Density Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Punjab.

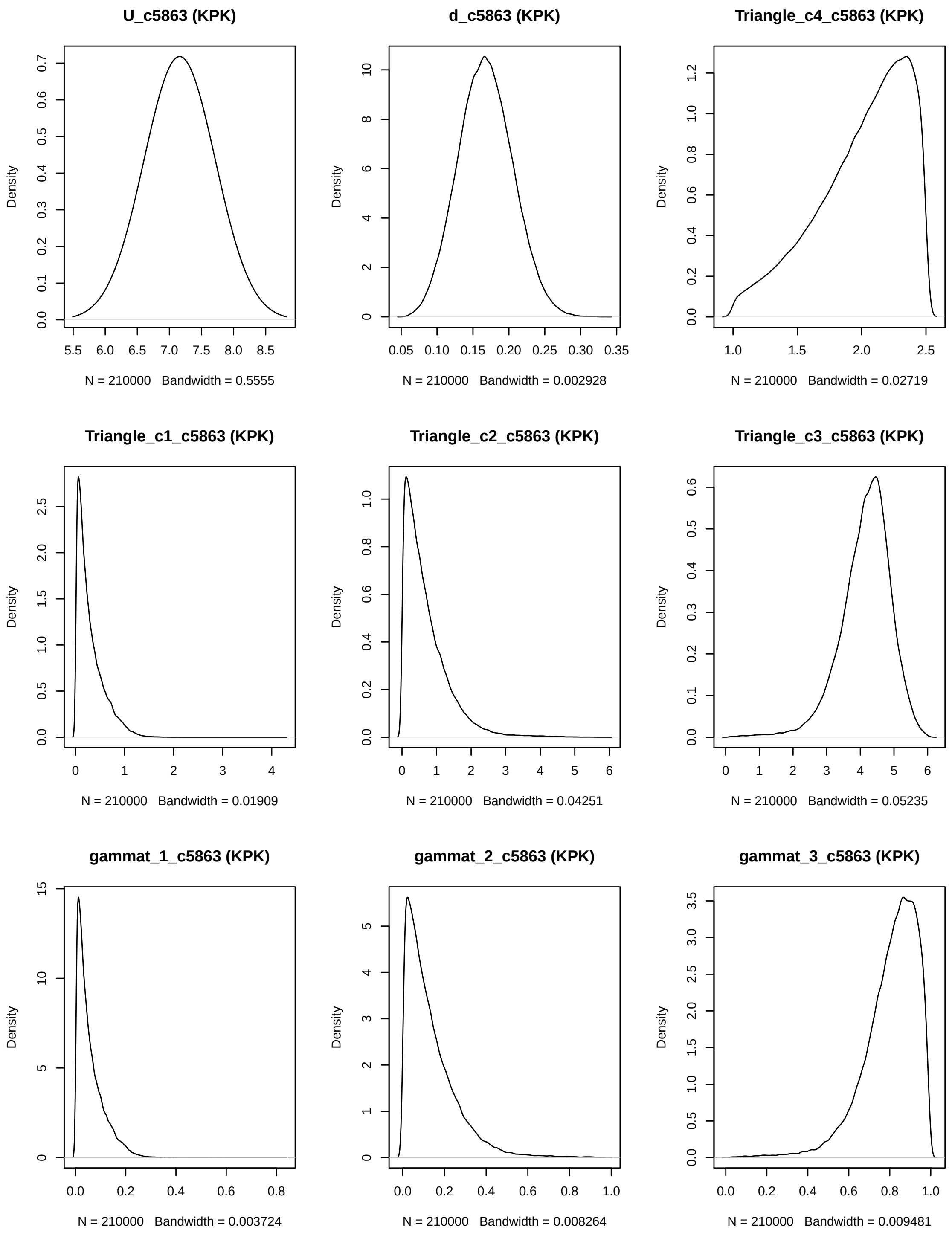


FIGURE .8: Density Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for KPK.

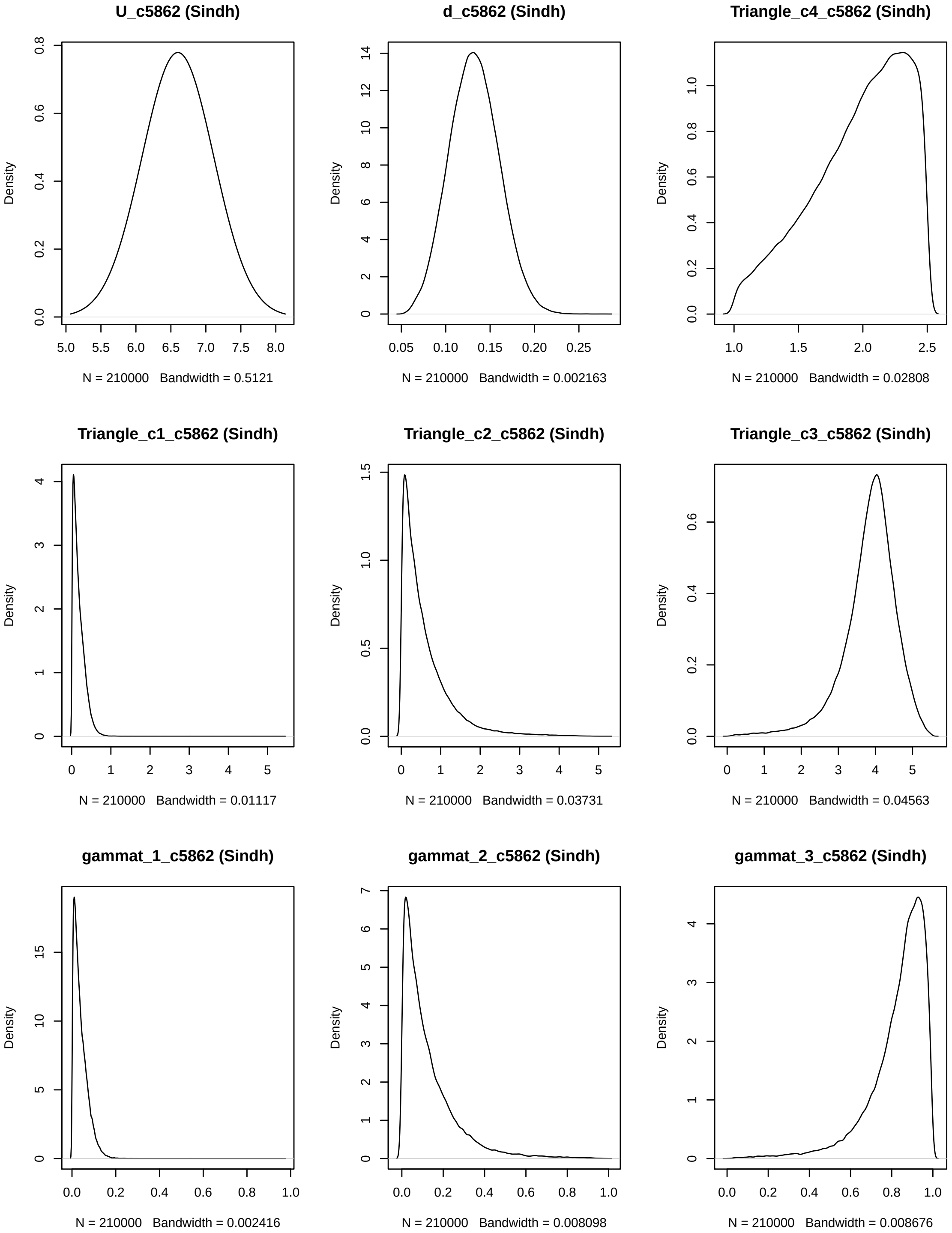


FIGURE .9: Density Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Sindh.

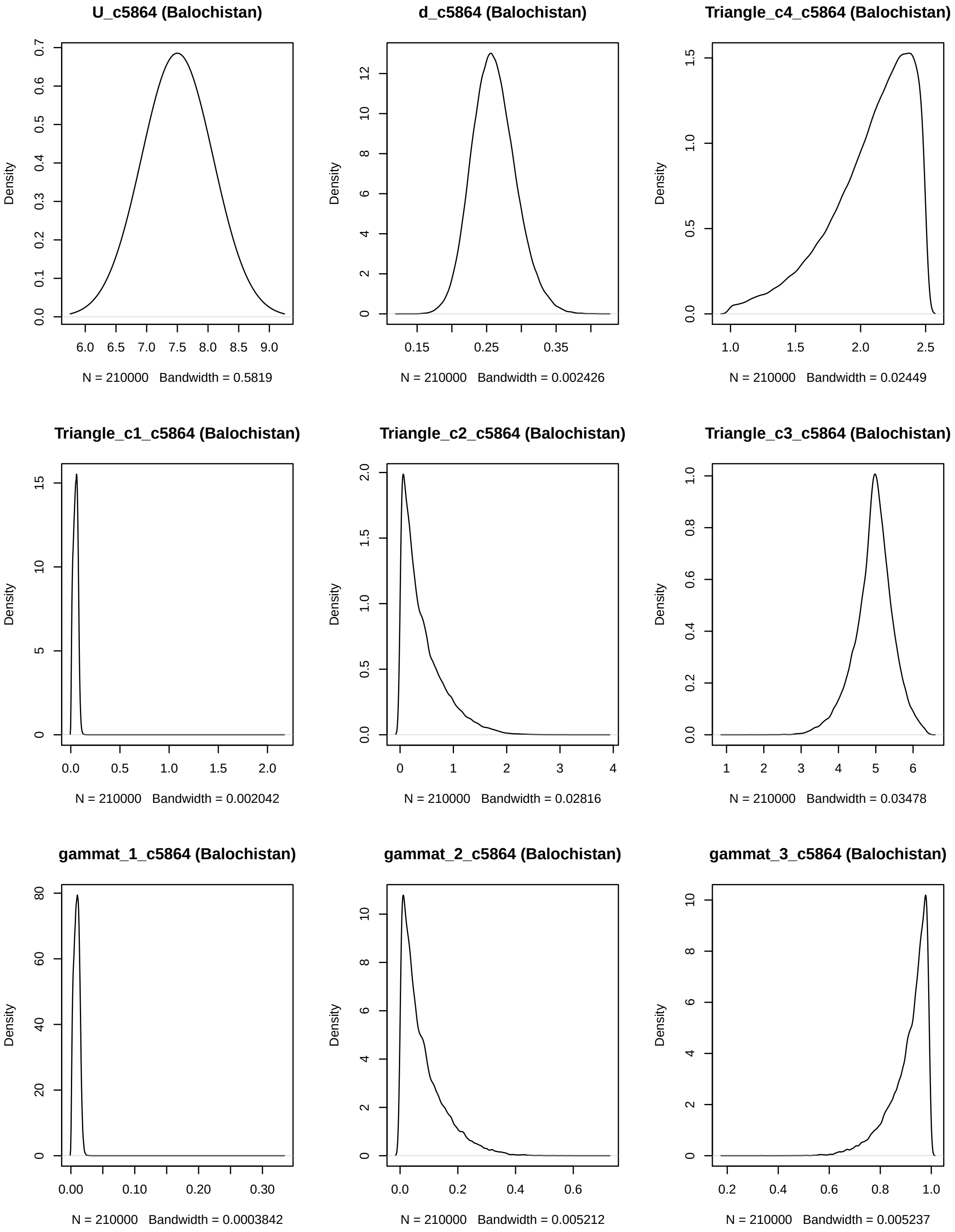


FIGURE .10: Density Plots of parameters i) U_c, ii) d_c, iii)$\Delta_{ci}, i = 1, 2, 3, 4$, iv) $p_{ci}, i = 1, 2, 3$ (gammat) for Balochistan.

Life Expectancy Tables

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	67.70	65.71	66.45	69.02	69.62
2023	68.55	65.72	66.71	70.33	71.25
2028	69.34	65.77	67.08	71.47	72.54
2033	70.00	66.12	67.46	72.68	74.02
2038	70.75	66.37	67.79	73.78	75.18
2043	71.47	66.62	68.19	74.69	76.13
2048	72.14	66.95	68.63	75.62	77.21
2053	72.87	67.24	69.05	76.56	78.19
2058	73.46	67.56	69.39	77.33	79.04
2063	74.15	68.05	69.99	78.11	79.90
2068	74.81	68.31	70.56	78.86	80.69
2073	75.45	68.68	71.03	79.61	81.44
2078	76.10	68.97	71.23	80.33	82.42
2083	76.77	69.22	71.66	81.15	83.05
2088	77.42	69.72	72.01	81.78	83.97
2093	77.94	69.94	72.56	82.48	84.82
2098	78.53	70.11	73.02	83.25	85.51

Table .12: Female Life Expectancy at birth for Pakistan showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	65.65	63.50	64.28	67.08	67.79
2023	66.25	63.16	64.29	68.31	69.18
2028	66.78	62.85	64.33	69.25	70.56
2033	67.29	62.89	64.49	70.18	71.80
2038	67.83	62.75	64.61	71.22	72.74
2043	68.43	62.64	64.72	71.86	73.54
2048	68.98	62.80	64.98	72.58	74.59
2053	69.33	63.02	65.17	73.50	75.61
2058	69.83	63.31	65.42	74.34	76.70
2063	70.29	63.48	65.73	75.08	77.62
2068	70.75	63.64	66.22	76.03	78.59
2073	71.23	64.11	66.38	76.98	79.52
2078	71.91	64.26	66.67	77.94	80.52
2083	72.57	64.57	66.99	78.89	81.60
2088	73.30	64.72	67.24	79.73	82.31
2093	73.91	64.85	67.50	80.49	83.18
2098	74.56	65.02	67.93	81.35	83.95

TABLE .13: Male Life Expectancy at birth for Pakistan showing median trajectories with 80% and 95% confidence interval.

Year	Median	0.025	0.10	0.90	0.975
2018	70.00	68.02	68.78	71.24	71.84
2023	70.91	68.22	69.17	72.65	73.57
2028	71.78	68.42	69.69	73.95	74.95
2033	72.51	68.82	70.10	75.08	76.36
2038	73.32	69.10	70.49	76.18	77.49
2043	74.09	69.41	70.99	77.23	78.72
2048	74.85	69.69	71.40	78.09	79.92
2053	75.53	70.17	71.95	78.93	80.93
2058	76.26	70.62	72.54	79.78	81.82
2063	76.97	70.59	72.94	80.71	82.73
2068	77.65	71.14	73.54	81.48	83.67
2073	78.29	71.61	73.97	82.31	84.67
2078	78.93	72.19	74.44	83.11	85.25
2083	79.57	72.12	74.86	83.80	86.07
2088	80.19	72.52	75.29	84.53	87.12
2093	80.71	72.98	75.76	85.18	87.88
2098	81.32	73.31	76.26	85.88	88.58

TABLE .14: Female Life Expectancy at birth for Punjab province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	67.35	65.15	65.97	68.74	69.39
2023	68.07	64.87	66.09	69.98	70.99
2028	68.79	64.89	66.33	71.06	72.26
2033	69.31	64.96	66.50	72.04	73.58
2038	69.93	64.81	66.68	73.00	74.98
2043	70.55	65.07	67.01	73.88	76.27
2048	71.19	65.23	67.21	75.01	77.21
2053	71.74	65.55	67.62	75.98	78.85
2058	72.38	65.76	67.98	76.99	80.00
2063	73.05	65.63	68.16	78.08	81.03
2068	73.78	65.92	68.61	79.21	81.94
2073	74.54	66.18	68.87	80.28	82.71
2078	75.36	66.41	69.21	81.10	83.44
2083	76.21	66.42	69.73	81.93	84.22
2088	76.97	66.61	70.05	82.71	85.19
2093	77.80	67.10	70.49	83.30	86.15
2098	78.71	67.44	70.69	83.93	86.55

TABLE .15: Male Life Expectancy at birth for Punjab province showing median trajectories with 80% and 95% confidence interval.

Year	Median	0.025	0.10	0.90	0.975
2018	70.98	69.14	69.77	72.32	72.90
2023	71.82	69.19	70.03	73.65	74.69
2028	72.64	69.22	70.39	74.82	76.08
2033	73.41	69.45	70.70	75.89	77.10
2038	74.21	69.80	71.23	76.87	78.28
2043	74.87	70.12	71.78	77.76	79.33
2048	75.63	70.48	72.15	78.69	80.24
2053	76.24	70.67	72.68	79.50	81.26
2058	76.94	71.08	73.19	80.37	82.13
2063	77.63	71.60	73.71	81.16	82.96
2068	78.14	71.71	74.07	81.94	83.96
2073	78.78	72.07	74.66	82.77	84.83
2078	79.43	72.26	75.12	83.46	85.75
2083	80.06	72.75	75.54	84.23	86.39
2088	80.65	73.40	76.01	84.88	87.00
2093	81.22	73.54	76.42	85.64	87.64
2098	81.84	73.79	76.76	86.28	88.51

Table .16: Female Life Expectancy at birth for Sindh province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	67.83	65.64	66.43	69.25	69.99
2023	68.43	65.35	66.47	70.44	71.49
2028	69.09	65.34	66.57	71.47	73.00
2033	69.63	65.17	66.69	72.49	73.87
2038	70.28	65.42	66.92	73.43	75.25
2043	70.74	65.56	67.28	74.36	76.34
2048	71.33	65.41	67.52	75.48	77.45
2053	72.00	65.48	67.86	76.56	78.58
2058	72.66	65.65	68.20	77.59	79.93
2063	73.34	66.02	68.53	78.49	81.01
2068	74.02	66.41	68.81	79.58	81.94
2073	74.94	66.41	69.32	80.58	83.01
2078	75.75	66.47	69.68	81.32	83.60
2083	76.62	66.82	70.04	82.25	84.55
2088	77.51	67.16	70.46	82.85	85.06
2093	78.30	67.53	70.96	83.53	85.77
2098	79.12	67.66	71.53	84.04	86.54

TABLE .17: Male Life Expectancy at birth for Sindh province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	70.19	68.33	68.99	71.46	72.15
2023	70.98	68.24	69.25	72.71	73.57
2028	71.73	68.48	69.67	73.69	74.82
2033	72.42	68.71	69.98	74.83	75.92
2038	73.10	68.85	70.39	75.76	77.13
2043	73.72	69.13	70.81	76.67	78.13
2048	74.42	69.60	71.23	77.47	79.13
2053	75.02	69.76	71.68	78.35	80.06
2058	75.63	70.00	72.01	79.13	80.85
2063	76.25	70.31	72.40	79.82	81.69
2068	76.85	70.56	72.78	80.59	82.49
2073	77.43	70.86	73.20	81.28	83.24
2078	78.06	71.21	73.66	82.04	83.98
2083	78.53	71.63	73.98	82.71	84.81
2088	79.15	71.98	74.52	83.28	85.65
2093	79.77	72.23	74.84	83.92	86.21
2098	80.29	72.53	75.41	84.70	86.84

TABLE .18: Female Life Expectancy at birth for KPK province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	68.93	66.68	67.56	70.27	71.01
2023	69.45	66.20	67.48	71.23	72.19
2028	69.83	66.01	67.55	72.09	73.25
2033	70.31	65.89	67.59	72.90	74.25
2038	70.76	65.78	67.56	73.64	75.24
2043	71.16	65.55	67.70	74.45	76.26
2048	71.63	65.93	67.87	75.13	77.20
2053	72.02	65.83	68.11	76.15	78.39
2058	72.47	65.86	68.37	76.94	79.10
2063	72.94	65.84	68.58	77.96	80.19
2068	73.49	65.84	68.68	78.64	81.07
2073	74.01	66.03	68.90	79.52	81.88
2078	74.75	65.94	69.13	80.50	82.59
2083	75.39	66.30	69.50	81.41	83.35
2088	76.14	66.47	69.63	81.99	84.07
2093	76.84	66.68	70.05	82.58	84.63
2098	77.62	66.88	70.33	83.12	85.42

TABLE .19: Male Life Expectancy at birth for KPK province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	66.40	64.55	65.12	67.65	68.37
2023	67.07	64.45	65.36	68.88	69.78
2028	67.75	64.50	65.66	70.01	71.13
2033	68.42	64.42	65.93	70.89	72.04
2038	68.96	64.71	66.24	71.73	73.16
2043	69.48	64.92	66.58	72.53	74.12
2048	70.07	65.08	66.92	73.38	75.11
2053	70.66	65.25	67.30	74.13	75.92
2058	71.17	65.37	67.55	74.87	76.78
2063	71.73	65.68	67.80	75.65	77.54
2068	72.20	66.14	68.16	76.42	78.42
2073	72.78	66.47	68.57	77.08	78.87
2078	73.41	66.79	68.83	77.76	79.70
2083	73.91	66.93	69.21	78.39	80.49
2088	74.50	67.08	69.56	78.99	81.17
2093	75.03	67.20	69.88	79.70	81.82
2098	75.60	67.34	70.29	80.46	82.49

Table .20: Female Life Expectancy at birth for Balochistan province showing median trajectories with 80% and 95% confidence interval.

Year	**Median**	**0.025**	**0.10**	**0.90**	**0.975**
2018	63.38	61.18	61.99	64.74	65.66
2023	63.89	60.82	62.02	65.92	67.07
2028	64.41	60.67	62.04	66.85	68.13
2033	64.90	60.59	62.17	67.62	68.95
2038	65.33	60.56	62.21	68.28	69.84
2043	65.65	60.47	62.34	68.97	70.71
2048	66.05	60.55	62.50	69.55	71.61
2053	66.46	60.33	62.68	70.17	72.12
2058	66.91	60.27	62.79	70.90	73.02
2063	67.29	60.55	62.84	71.53	73.64
2068	67.60	60.80	63.19	72.21	74.59
2073	68.07	60.85	63.43	72.83	75.29
2078	68.55	61.05	63.57	73.54	76.37
2083	68.95	61.02	63.83	74.40	77.34
2088	69.40	61.07	64.07	75.18	78.49
2093	69.84	61.39	64.33	76.12	79.03
2098	70.40	61.25	64.57	76.94	80.14

TABLE .21: Male Life Expectancy at birth for Balochistan province showing median trajectories with 80% and 95% confidence interval.

Year	mean	SD	RSD	2.50%	5%	10%	25%	50%	75%	90%	95%	97.50%
2018	67.7	1.00	0.01	65.70	66.10	66.50	67.10	67.70	68.40	69.00	69.30	69.60
2023	68.5	1.42	0.02	65.70	66.20	66.70	67.60	68.50	69.40	70.30	70.80	71.30
2028	69.3	1.74	0.03	65.80	66.40	67.10	68.10	69.30	70.40	71.50	72.00	72.50
2033	70.0	2.03	0.03	66.10	66.80	67.50	68.60	70.00	71.40	72.70	73.40	74.00
2038	70.7	2.28	0.03	66.40	67.10	67.80	69.20	70.80	72.20	73.80	74.60	75.20
2043	71.5	2.49	0.03	66.60	67.30	68.20	69.80	71.50	73.20	74.70	75.50	76.10
2048	72.1	2.70	0.04	67.00	67.60	68.60	70.30	72.10	74.00	75.60	76.40	77.20
2053	72.8	2.85	0.04	67.20	68.00	69.10	70.90	72.90	74.80	76.60	77.40	78.20
2058	73.5	2.98	0.04	67.60	68.50	69.40	71.50	73.50	75.50	77.30	78.20	79.00
2063	74.1	3.11	0.04	68.00	69.00	70.00	72.00	74.20	76.20	78.10	79.20	79.90
2068	74.7	3.24	0.04	68.30	69.40	70.60	72.50	74.80	77.00	78.90	80.00	80.70
2073	75.3	3.36	0.04	68.70	69.50	71.00	73.10	75.50	77.70	79.60	80.60	81.40
2078	75.9	3.48	0.05	69.00	70.10	71.20	73.50	76.10	78.40	80.30	81.50	82.40
2083	76.6	3.61	0.05	69.20	70.30	71.70	74.10	76.80	79.00	81.20	82.30	83.00
2088	77.1	3.73	0.05	69.70	70.70	72.00	74.70	77.40	79.70	81.80	83.00	84.00
2093	77.7	3.85	0.05	69.90	71.10	72.60	75.20	77.90	80.40	82.50	83.70	84.80
2098	78.3	3.95	0.05	70.10	71.40	73.00	75.70	78.50	81.00	83.30	84.60	85.50

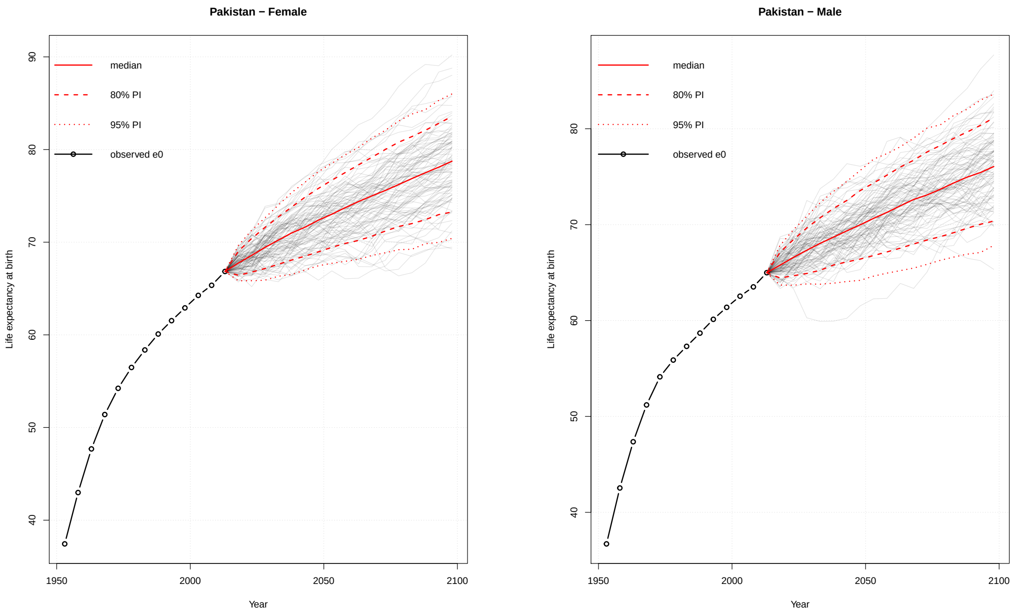
Table .22: Table showing mean life expectancy for females along with S.D, R.S.D and different confidence limits at national level.

Life Expectancy Parameters Estimates

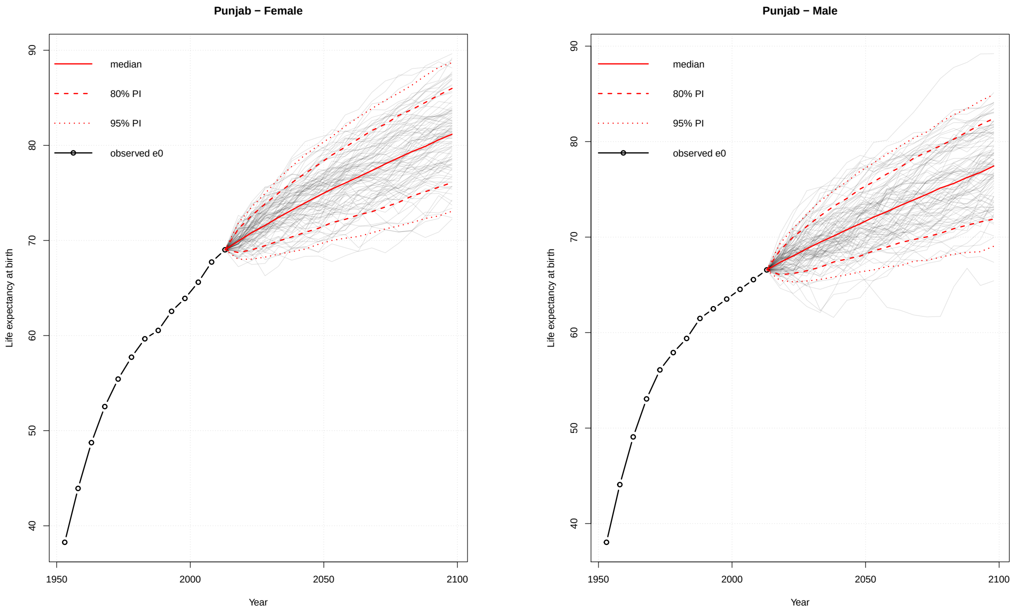
Pakistan	**Mean**	**SD**	**Naive SE**	**Time-series SE**
Δ_1^{PK}	8.47380	5.25040	0.03676	0.03957
Δ_2^{PK}	31.32900	8.18630	0.05732	0.07514
Δ_3^{PK}	3.07750	7.97700	0.05585	0.07328
Δ_4^{PK}	23.51480	8.17330	0.05722	0.05764
k^{PK}	4.67820	0.86430	0.00605	0.00645
z^{PK}	0.50720	0.11290	0.00079	0.00084
Punjab	**Mean**	**SD**	**Naive SE**	**Time-series SE**
Δ_1^{PB}	8.53330	5.26710	0.03688	0.03900
Δ_2^{PB}	31.42430	8.20230	0.05743	0.07381
Δ_3^{PB}	3.49080	8.17490	0.05724	0.07300
Δ_4^{PB}	25.66280	8.36300	0.05855	0.05906
k^{PB}	4.70890	0.87680	0.00614	0.00630
z^{PB}	0.51030	0.11170	0.00078	0.00080
Sindh	**Mean**	**SD**	**Naive SE**	**Time-series SE**
Δ_1^{SD}	8.73580	5.32720	0.03730	0.04238
Δ_2^{SD}	31.83640	8.07580	0.05654	0.07742
Δ_3^{SD}	4.25970	7.85440	0.05499	0.07468
Δ_4^{SD}	24.19200	8.05460	0.05639	0.05829
k^{SD}	4.89250	0.85550	0.00599	0.00624
z^{SD}	0.50960	0.11230	0.00079	0.00081
KPK	**Mean**	**SD**	**Naive SE**	**Time-series SE**
Δ_1^{KP}	9.03700	5.41990	0.03795	0.04716
Δ_2^{KP}	32.41900	8.00790	0.05607	0.08461
Δ_3^{KP}	5.15500	7.67690	0.05375	0.07891
Δ_4^{KP}	20.73600	7.25680	0.05081	0.05323
k^{KP}	5.00900	0.82520	0.00578	0.00613
z^{KP}	0.50500	0.11430	0.00080	0.00082
Balochistan	**Mean**	**SD**	**Naive SE**	**Time-series SE**
Δ_1^{BL}	9.16620	5.43290	0.03804	0.04861
Δ_2^{BL}	33.43580	8.10410	0.05674	0.09934
Δ_3^{BL}	6.06360	7.67750	0.05375	0.09454
Δ_4^{BL}	14.63810	6.15660	0.04311	0.04900
k^{BL}	5.04030	0.81560	0.00571	0.00647
z^{BL}	0.48740	0.12100	0.00085	0.00087

TABLE .23: Empirical mean and standard deviation for each variable, plus standard error of the mean at national and provincial level

Life Expectancy Graphs

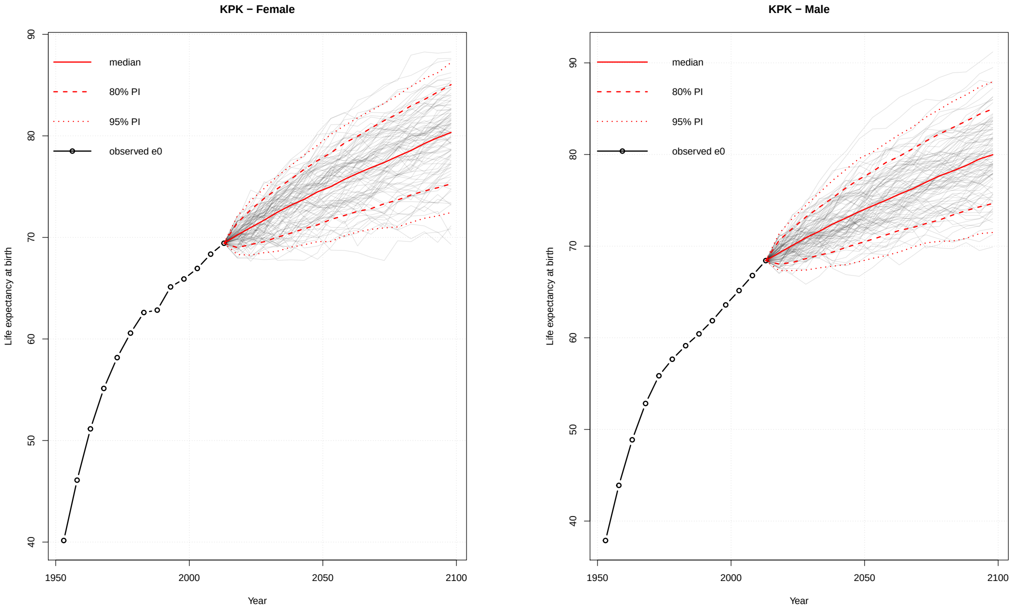


(a) Life Expectancy plot for Pakistan Female (b) Life Expectancy plot for Pakistan Male



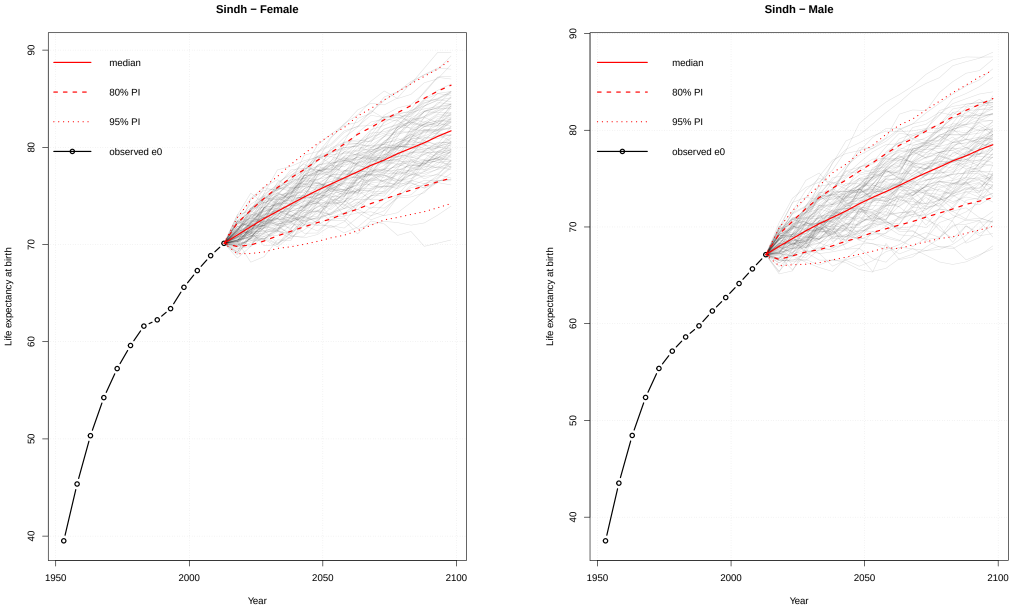
(c) Life Expectancy plot for Punjab Female (d) Life Expectancy plot for Punjab Male

FIGURE .11: Graphs showing Median trajectories of Life Expectancy at birth for both genders separately along with 80% and 95% projection intervals at national and regional levels Punjab.



(a) Life Expectancy plot for KPK Female

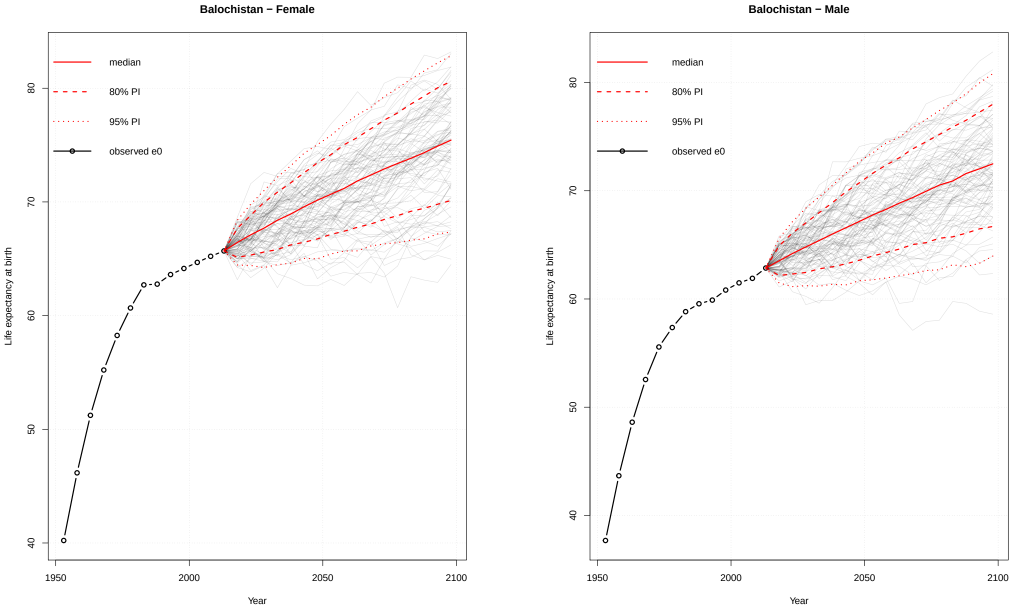
(b) Life Expectancy plot for KPK Male



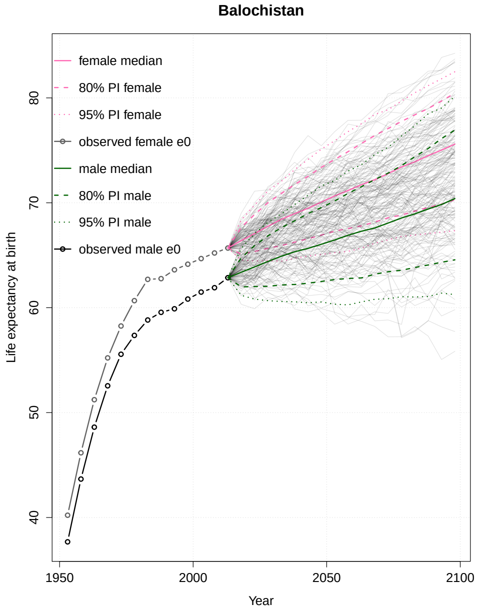
(c) Life Expectancy plot for Sindh Female

(d) Life Expectancy plot for Sindh Male

FIGURE .12: Graphs showing Median trajectories of Life Expectancy at birth for both genders separately along with 80% and 95% projection intervals for KPK and Sindh.



(a) Life Expectancy plot for Balochistan Female (b) Life Expectancy plot for Balochista Male



(c) Life Expectancy plot for Balochistan both

FIGURE .13: Graphs showing Median trajectories of Life Expectancy at birth for both genders separately and combined along with 80% and 95% projection intervals for Balochistan.

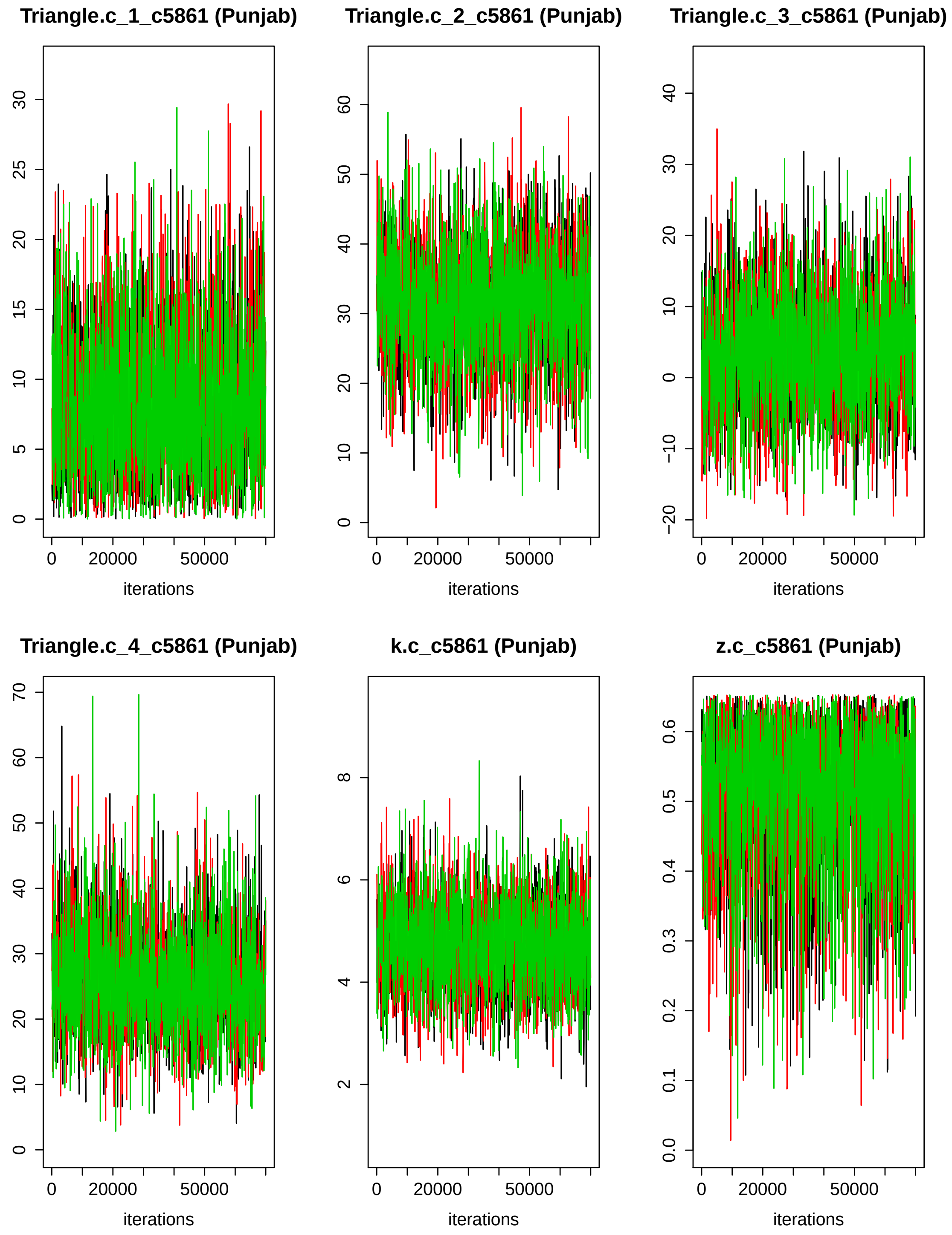


FIGURE .14: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k^c, iii)z^c for Punjab Province.

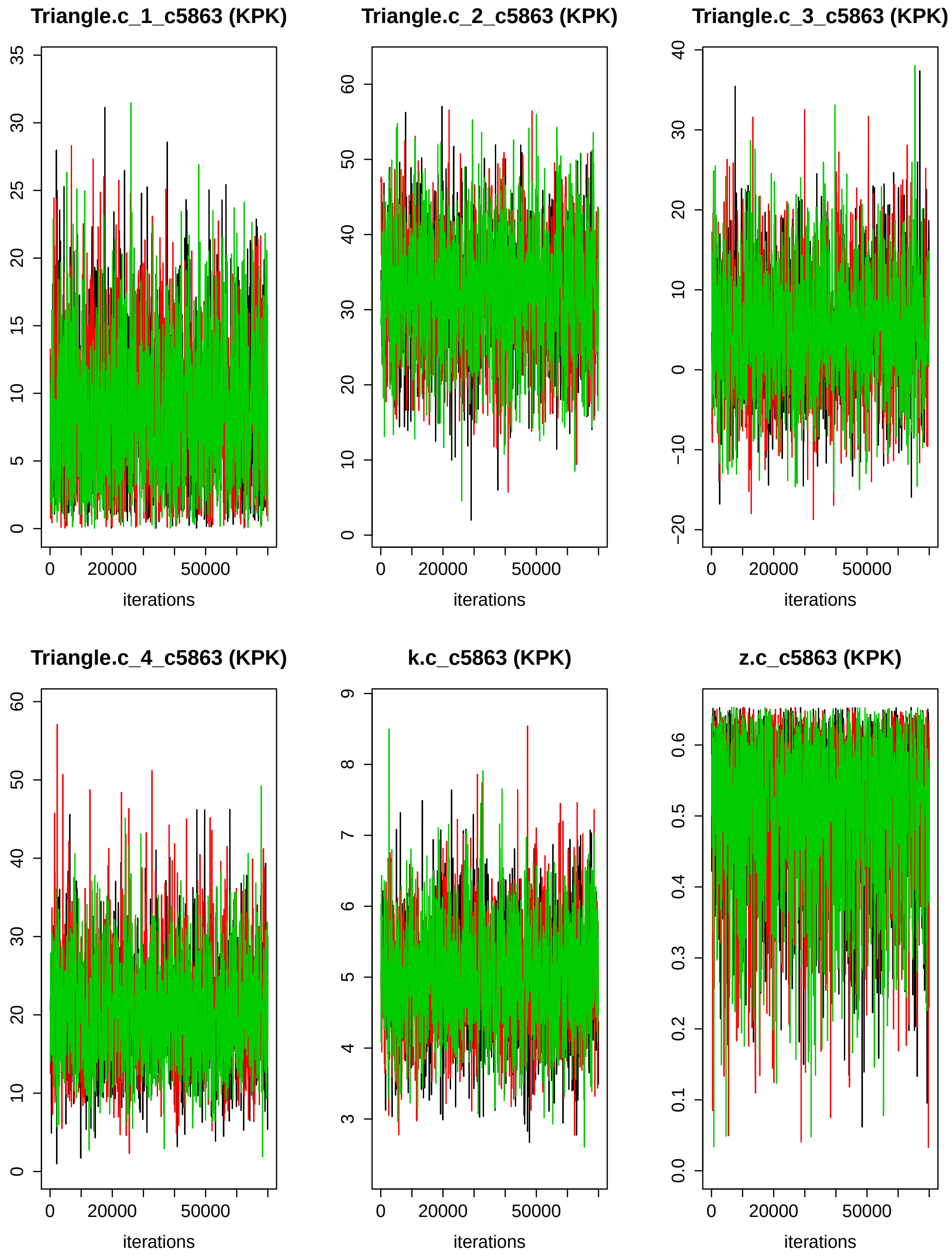


FIGURE .15: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k^c, iii)z^c for KPK Province.

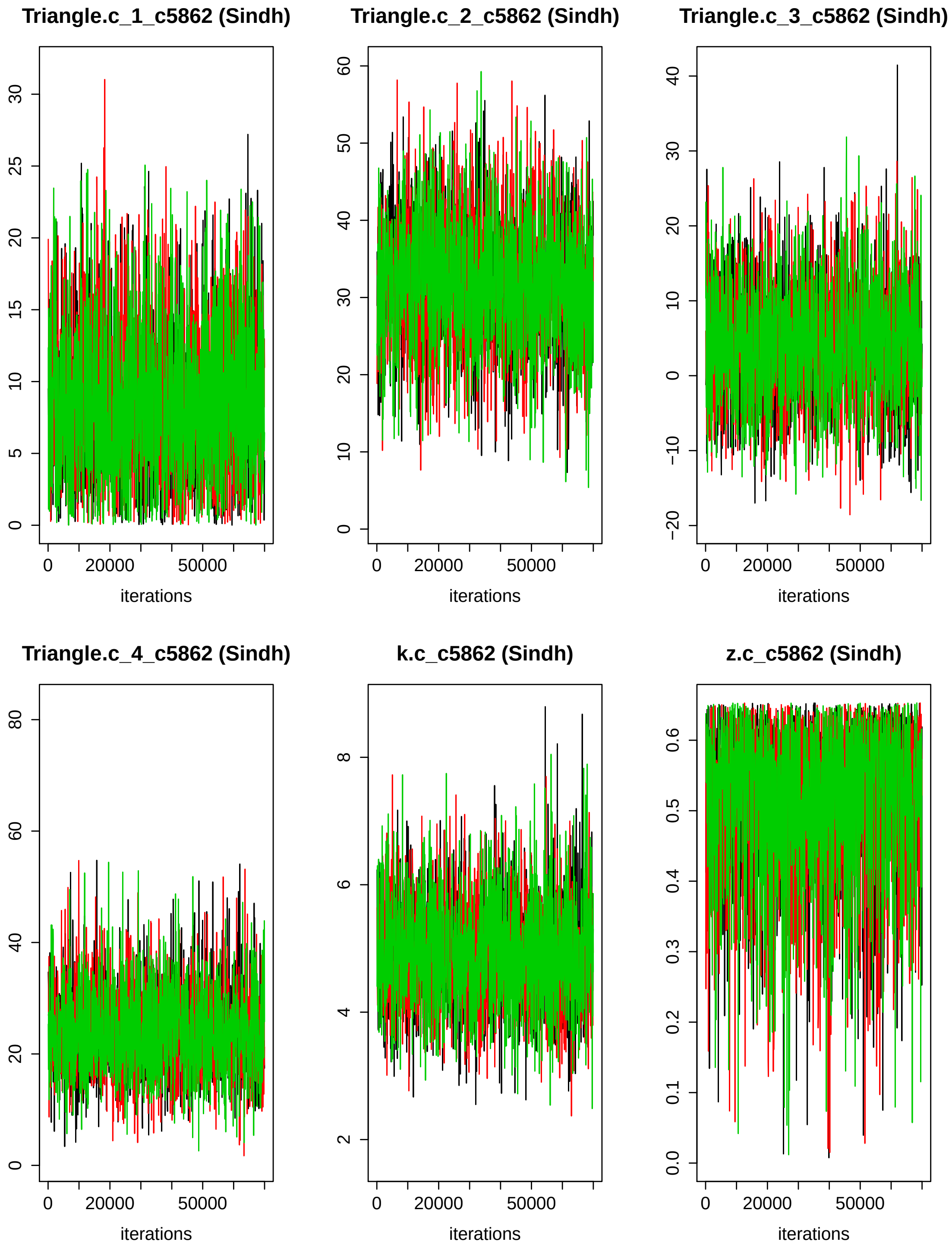


FIGURE .16: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k^c, iii)z^c for Sindh Province.

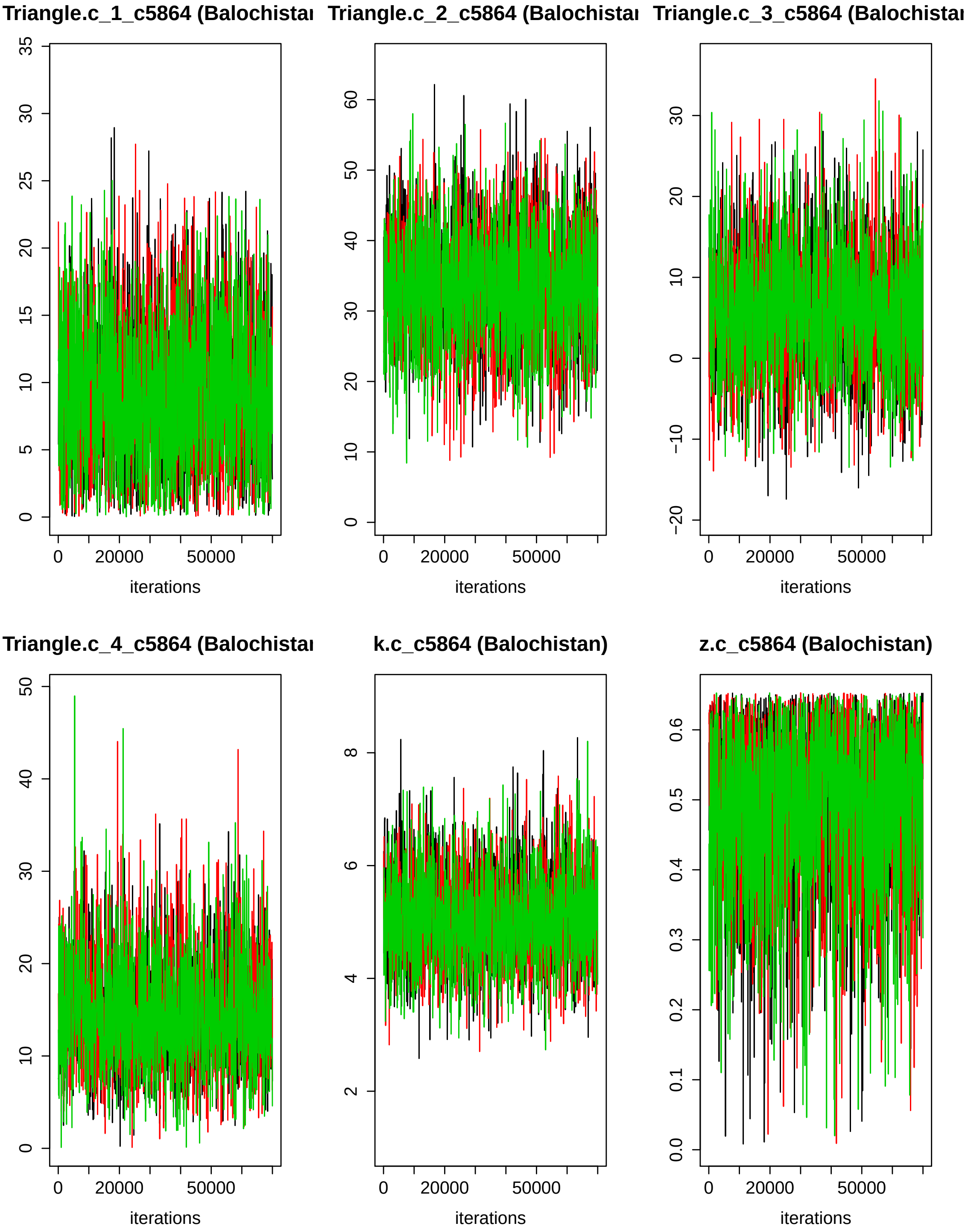


FIGURE .17: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k^c, iii)z^c for Balochistan Province.

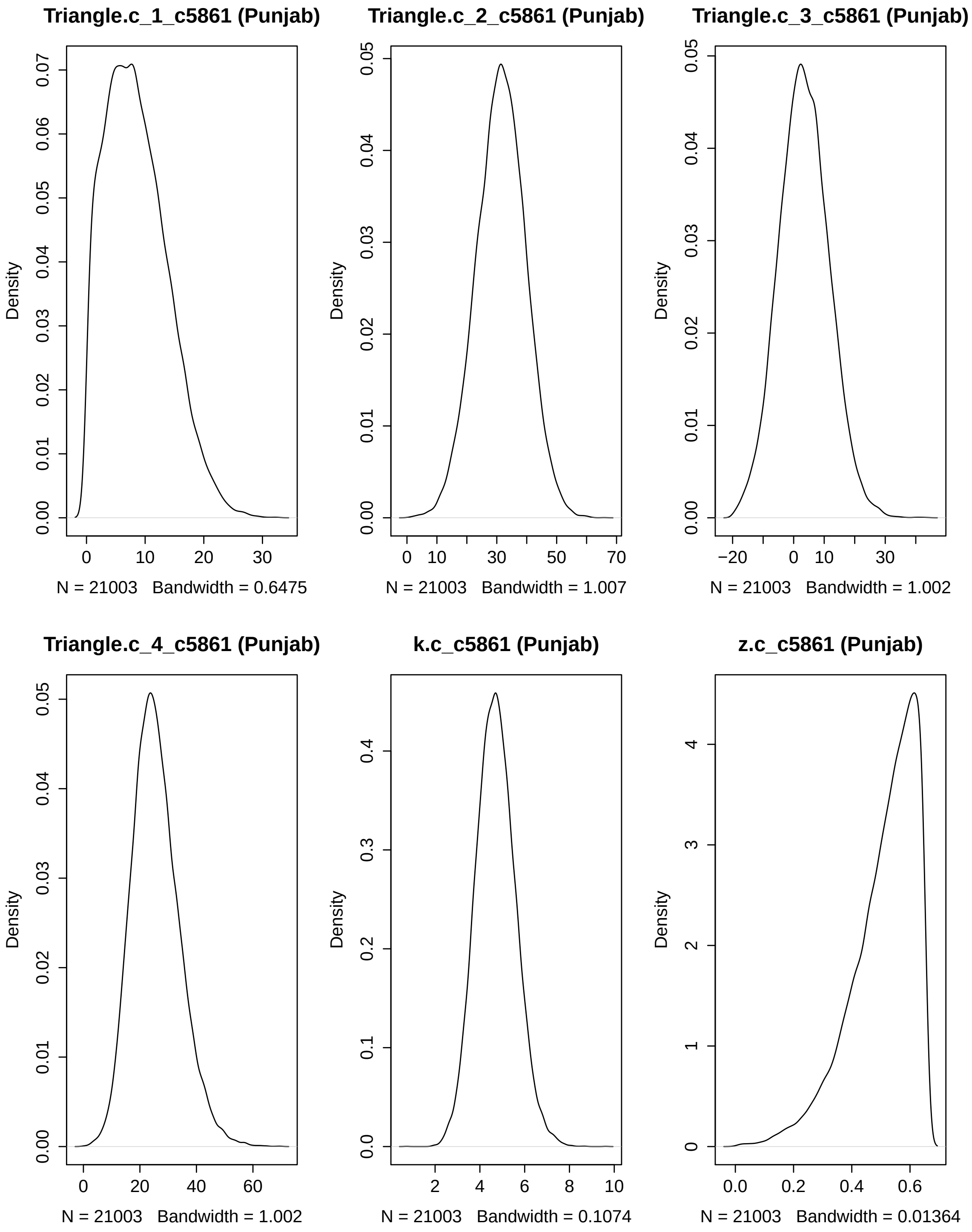


FIGURE .18: Trace Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k_c, iii)z_c for Punjab Province.

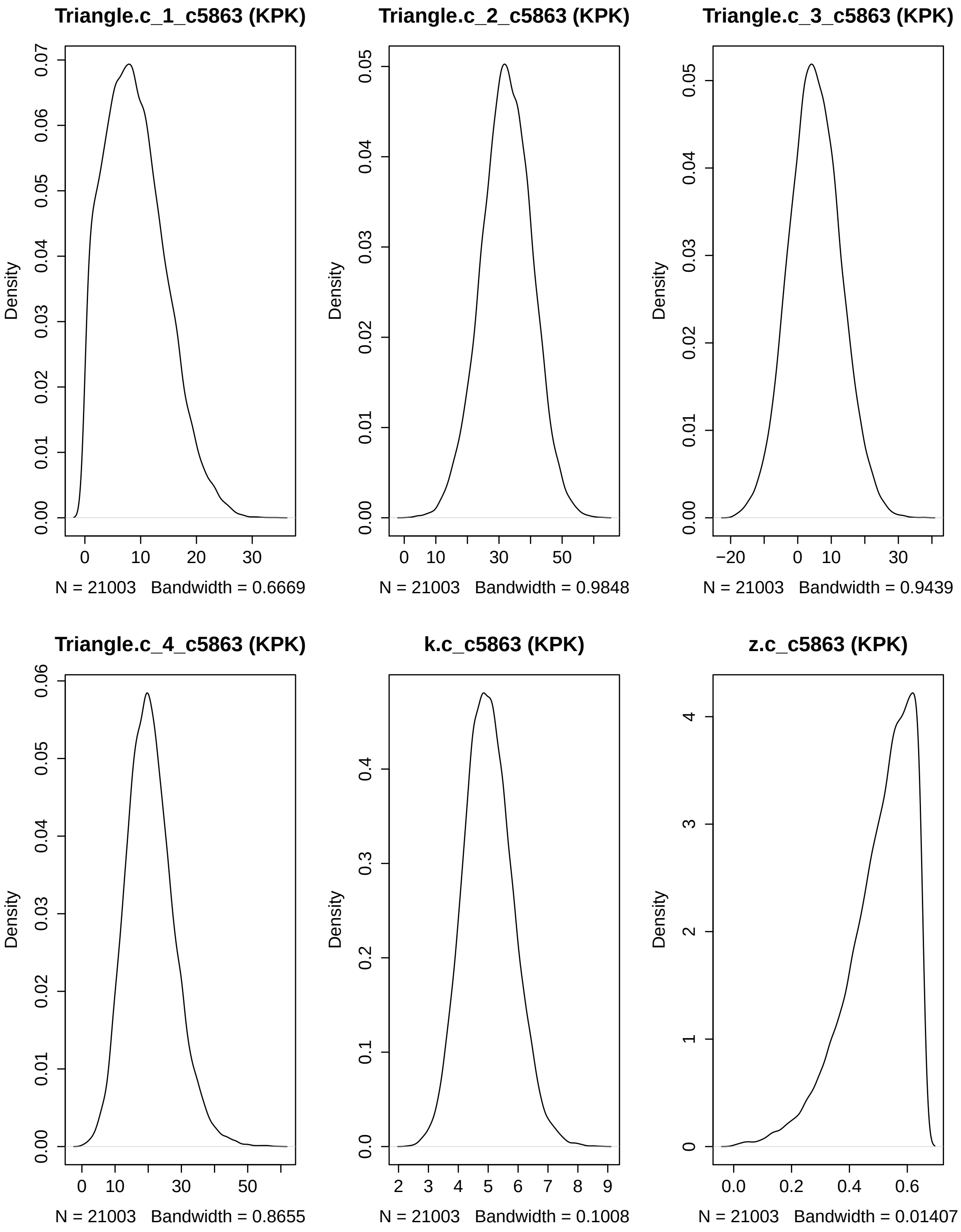


FIGURE .19: Density Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k_c, iii)z_c for KPK Province.

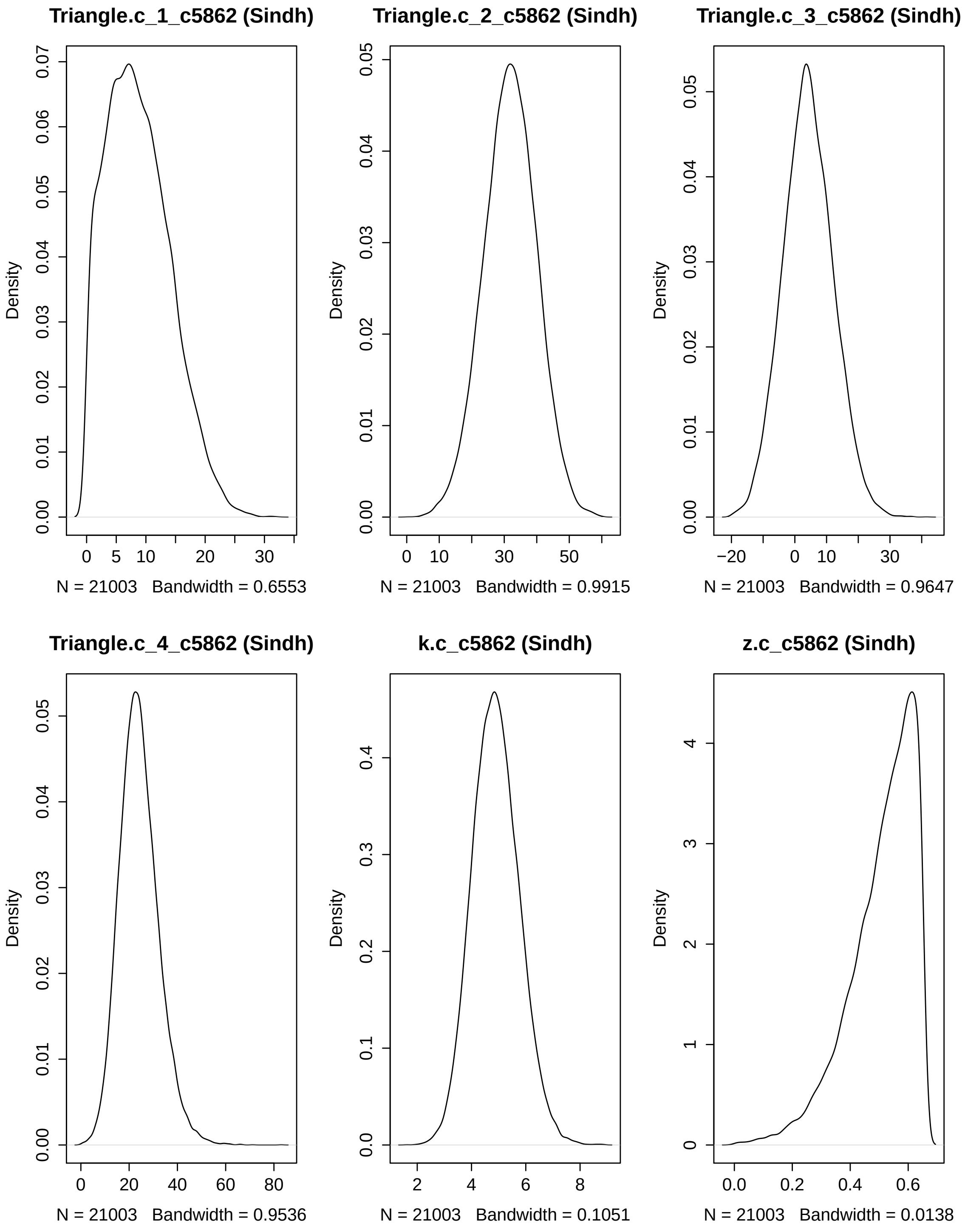


FIGURE .20: Density Plots of parameters i) $\Delta_{ci}, i = 1, 2, 3, 4$, ii) k_c, iii)z_c for Sindh Province.

Population Projections

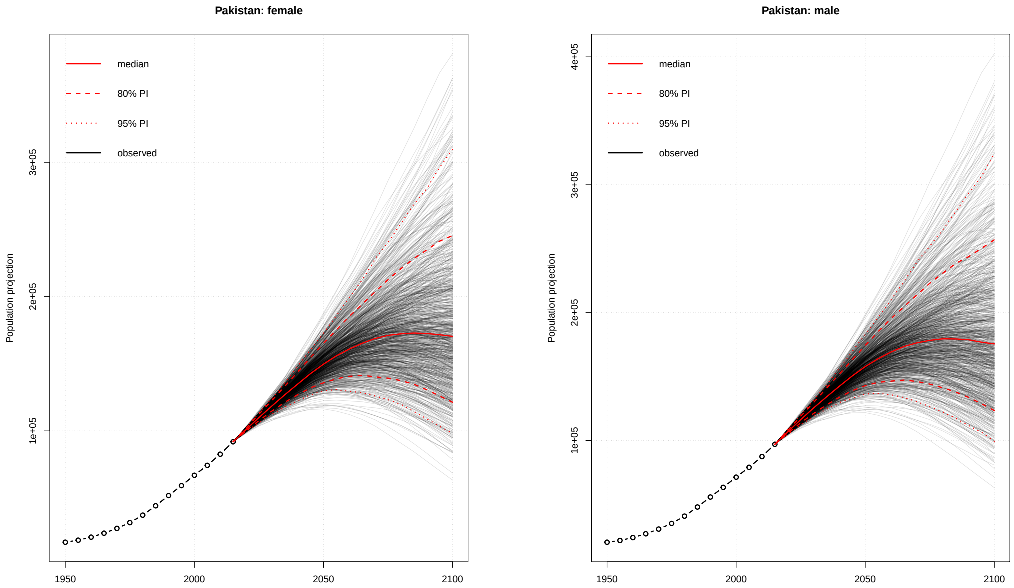
Table .24: Population projections (in thousand) at national level for male and female separately with mean, SD, 2.5%, 10.0% ,25%, 50%,75% , 90% and 97.5% projection bounds.

<table>
<tr><th colspan="10">Male population projections</th></tr>
<tr><th>Year</th><th>mean</th><th>SD</th><th>2.50%</th><th>10%</th><th>25%</th><th>50%</th><th>75%</th><th>90%</th><th>97.50%</th></tr>
<tr><td>2018</td><td>106803</td><td>908.29</td><td>105035</td><td>105632</td><td>106199</td><td>106799</td><td>107438</td><td>107943</td><td>108567</td></tr>
<tr><td>2023</td><td>116319</td><td>2015.93</td><td>112338</td><td>113796</td><td>114983</td><td>116291</td><td>117737</td><td>118952</td><td>120004</td></tr>
<tr><td>2028</td><td>125357</td><td>3356.19</td><td>118927</td><td>121142</td><td>123128</td><td>125457</td><td>127653</td><td>129744</td><td>131474</td></tr>
<tr><td>2033</td><td>134148</td><td>4996.51</td><td>124590</td><td>127825</td><td>130851</td><td>134097</td><td>137462</td><td>140913</td><td>143418</td></tr>
<tr><td>2038</td><td>142775</td><td>6953.24</td><td>129741</td><td>133816</td><td>138135</td><td>142468</td><td>147239</td><td>151975</td><td>156493</td></tr>
<tr><td>2043</td><td>150918</td><td>9335.63</td><td>133003</td><td>138962</td><td>144722</td><td>150529</td><td>156884</td><td>163261</td><td>169762</td></tr>
<tr><td>2048</td><td>158240</td><td>12100.22</td><td>136346</td><td>142872</td><td>150124</td><td>157764</td><td>165558</td><td>174319</td><td>182827</td></tr>
<tr><td>2053</td><td>164575</td><td>15261.76</td><td>136868</td><td>145488</td><td>154353</td><td>163952</td><td>173314</td><td>185111</td><td>196138</td></tr>
<tr><td colspan="10">Female Population projections</td></tr>
<tr><td>Year</td><td>mean</td><td>SD</td><td>2.50%</td><td>10%</td><td>25%</td><td>50%</td><td>75%</td><td>90%</td><td>97.50%</td></tr>
<tr><td>2018</td><td>101002</td><td>838.31</td><td>99348</td><td>99922</td><td>100441</td><td>100994</td><td>101581</td><td>102049</td><td>102610</td></tr>
<tr><td>2023</td><td>109951</td><td>1862.87</td><td>106281</td><td>107639</td><td>108733</td><td>109920</td><td>111246</td><td>112473</td><td>113287</td></tr>
<tr><td>2028</td><td>118504</td><td>3112.1</td><td>112407</td><td>114740</td><td>116474</td><td>118506</td><td>120646</td><td>122592</td><td>124171</td></tr>
<tr><td>2033</td><td>126849</td><td>4644.05</td><td>117778</td><td>121256</td><td>123769</td><td>126867</td><td>129872</td><td>133168</td><td>135584</td></tr>
<tr><td>2038</td><td>135093</td><td>6467.04</td><td>122475</td><td>126951</td><td>130905</td><td>134832</td><td>139224</td><td>143807</td><td>148056</td></tr>
<tr><td>2043</td><td>142970</td><td>8686.5</td><td>126836</td><td>132070</td><td>137292</td><td>142688</td><td>148549</td><td>154649</td><td>160580</td></tr>
<tr><td>2048</td><td>150118</td><td>11272.82</td><td>129878</td><td>135773</td><td>142531</td><td>149681</td><td>156817</td><td>165174</td><td>172616</td></tr>
<tr><td>2053</td><td>156391</td><td>14216.29</td><td>130698</td><td>138759</td><td>147026</td><td>155919</td><td>164545</td><td>175217</td><td>185976</td></tr>
</table>

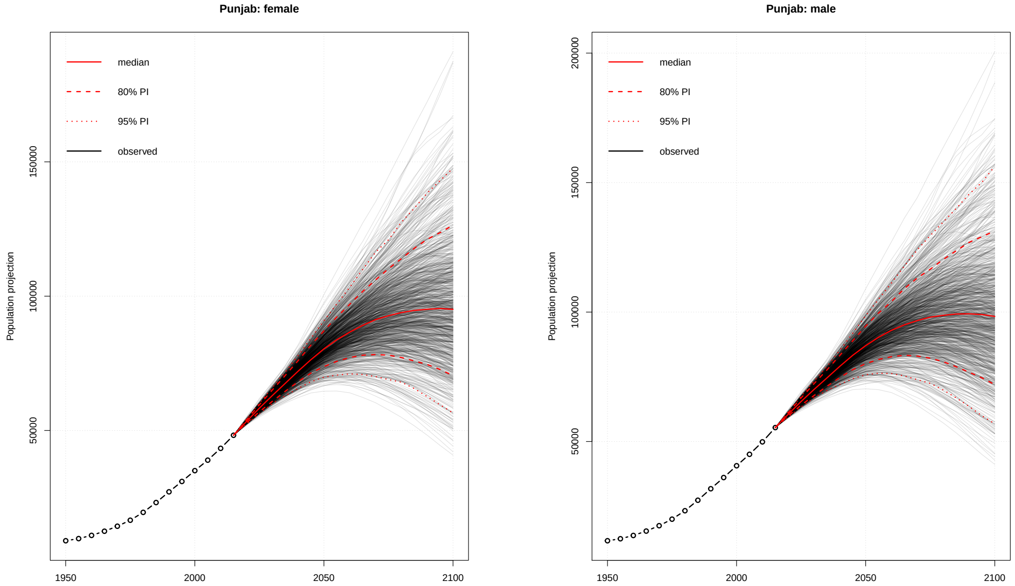
Table .25: Table showing mean population total along with Standard Deviation, Relative Standard Deviation and different confidence limits.

Year	**mean**	**SD**	**RSD**	**2.50%**	**10%**	**25%**	**75%**	**90%**	**97.50%**
2018	207806	1743.1	0.0084	204368	205559	206640	209029	210002	211142
2023	226270	3870.9	0.0171	218595	221474	223696	229001	231429	233327
2028	243861	6455.7	0.0265	231338	235889	239555	248334	252335	255617
2033	260997	9623.3	0.0369	242332	249034	254661	267349	274028	278889
2038	277867	13397.9	0.0482	252127	260704	269054	286456	295638	304545
2043	293888	17994.9	0.0612	259400	271250	282172	305420	317623	330102
2048	308358	23341.2	0.0757	266142	278716	292698	322354	338834	355670
2053	320966	29441.2	0.0917	267667	284325	300936	337792	360540	381535

population projection graphs

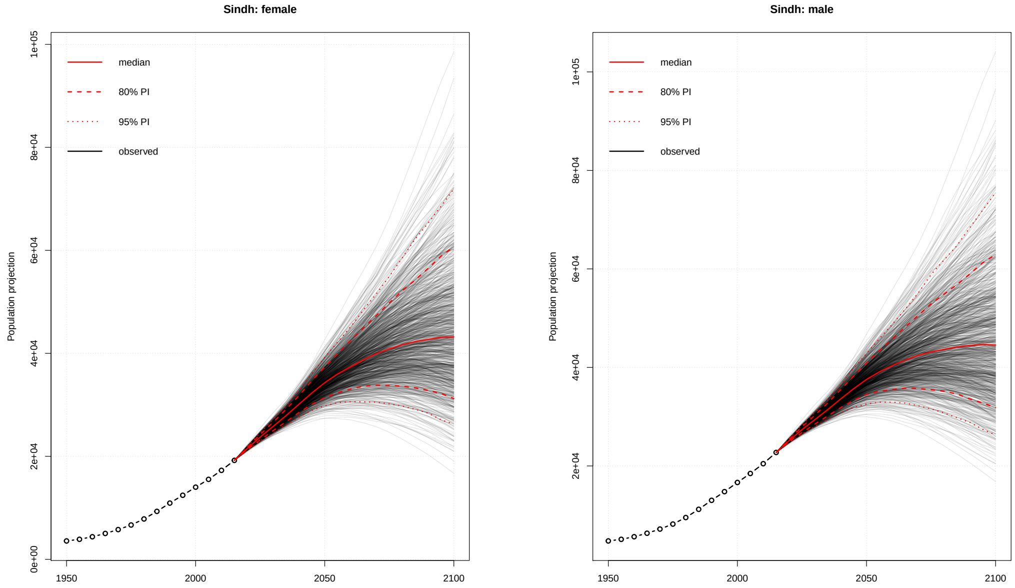


(a) Population totals projection for female (b) Population totals projection for male

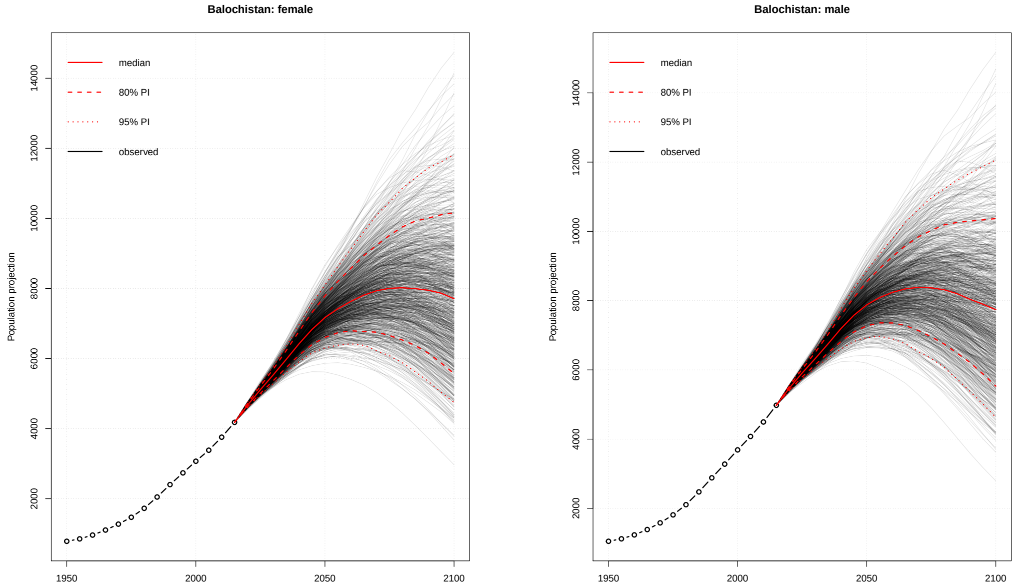


(c) Population totals projection for female (d) Population totals projection for male

FIGURE .21: Graphs showing Median trajectories of Population totals for both genders separately with 80% and 95% projection intervals for Pakistan and Punjab.



(a) Population totals projection for female (b) Population totals projection for male



(c) Population totals projection for female (d) Population totals projection for male

Figure .22: Graphs showing Median trajectories of Population totals for both genders separately with 80% and 95% projection intervals for the region of Sindh and Balochistan.

R Codes

```
############ TFR############
library(wpp2015)
library(snowFT)
library(bayesTFR)
simulation.dir <– file.path(getwd(), ”PROJECTIONS”)
my.tfr.file <– file.path(find.package(”wpp2015”), ”data”, ”my.tfr.file.txt”)
my.location.file <– file.path(find.package(”wpp2015”), ”data”, ”my.location.file.txt”)
TFR <– run.tfr.mcmc(nr.chains=3, iter=60000, output.dir=simulation.dir, my.tfr.file=my.tfr.file,
my.locations.file = my.location.file, parallel = TRUE)
TFR2 <– continue.tfr.mcmc(iter=10000, output.dir=simulation.dir,parallel= TRUE)
TFR2.1 <– run.tfr3.mcmc(sim.dir=simulation.dir, nr.chains = 3, iter = 60000, thin =
10, my.tfr.file=my.tfr.file, replace.output = TRUE, parallel = TRUE )
TFR2.2 <– continue.tfr3.mcmc(sim.dir=simulation.dir, iter=10000, parallel = TRUE)
TFR3 <– get.tfr.mcmc(sim.dir=simulation.dir)
TFR3.chain2 <– tfr.mcmc(TFR3, chain.id=2)
pred1 <– tfr.predict(sim.dir=simulation.dir, end.year=2100, burnin=1000, nr.traj=3000,
verbose=TRUE, replace.output = TRUE, use.tfr3 = TRUE, rho = NULL, sigmaAR1
= NULL)
pred2 <– get.tfr.prediction(sim.dir=simulation.dir)
# Summary functions
summary(TFR3, meta.only=TRUE)
summary(TFR3, country=”Pakistan”, par.names=NULL, thin=10, burnin=1000)
summary(pred2, country=”Pakistan”)
# Trajectories and DL curves
tfr.trajectories.plot(pred2, country=”Pakistan”, pi=c(95, 80, 75), nr.traj=100)
tfr.trajectories.table(pred2, country=”Punjab”, pi=c(95, 80, 75))
tfr.trajectories.table(pred2, country=”Sindh”, pi=c(95, 80, 75))
tfr.trajectories.table(pred2, country=”KPK”, pi=c(95, 80, 75))
```

```
tfr.trajectories.table(pred2, country=”Balochistan”, pi=c(95, 80, 75))
tfr.trajectories.table(pred2, country=”Pakistan”, pi=c(95, 80, 75))
DLcurve.plot(country=”Pakistan”, mcmc.list=m3, burnin=2000, pi=c(95, 80, 75), nr.curves=100)
# Parameter traces and density
tfr.partraces.cs.plot(country=”Pakistan”, mcmc.list=m3, nr.points=1000, par.names=
tfr.parameter.names.cs(trans = TRUE))
tfr.partraces.cs.plot(country=”Punjab”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter
= TRUE))
tfr.partraces.cs.plot(country=”Sindh”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter.n
= TRUE))
tfr.partraces.cs.plot(country=”KPK”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter.n
= TRUE))
tfr.partraces.cs.plot(country=”Balochistan”, mcmc.list=m3, nr.points=1000, par.names=
tfr.parameter.names.cs(trans = TRUE))
tfr.pardensity.plot(pred2, par.names=tfr.parameter.names(trans = TRUE), dev.ncol=3,
bw=0.05)
tfr.pardensity.cs.plot(country=”Pakistan”, mcmc.list=m3, nr.points=1000, par.names=
tfr.parameter.names.cs(trans = TRUE))
tfr.pardensity.cs.plot(country=”Punjab”, mcmc.list=m3, nr.points=1000, par.names=
tfr.parameter.names.cs(trans = TRUE),dev.ncol=2)
tfr.pardensity.cs.plot(country=”Sindh”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter.
= TRUE))
tfr.pardensity.cs.plot(country=”KPK”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter.n
= TRUE))
tfr.pardensity.cs.plot(country=”Balochistan”, mcmc.list=m3, nr.points=1000, par.names=
tfr.parameter.names.cs(trans = TRUE))
tfr.pardensity.cs.plot(country=”Asia”, mcmc.list=m3, nr.points=1000, par.names= tfr.parameter.n
= TRUE))
# Convergence
diagcheck1 <– tfr.diagnose(simulation.dir, thin=10, burnin=2000)
```

```
diagcheck2 <– get.tfr.convergence(simulation.dir, thin=10, burnin=2000)

    ############ Life Expectancy############
library(bayesLife)
sim.dir.e0 <– file.path(getwd(), ”LifeExpectancy”)
my.e0.file <– file.path(find.package(”wpp2015”), ”data”, ”my.e0F.file.txt”)
my.location.file <– file.path(find.package(”wpp2015”), ”data”, ”my.location.file.txt”)
run.e0.mcmc(sex = ”Female”, iter = 60000, nr.chains = 3, my.e0.file=my.e0.file, my.locations.file
= my.location.file, thin = 10, output.dir = sim.dir.e0, parallel = TRUE, replace.output=TRUE)
EXP <– continue.e0.mcmc(iter = 10000, output.dir =sim.dir.e0, parallel = TRUE)
EXP3 <– get.e0.mcmc(sim.dir=sim.dir.e0)
EXP3.chain2 <– e0.mcmc(EXP3, chain.id=2)
EXP.pred <– e0.predict(end.year = 2100,sim.dir = sim.dir.e0,replace.output = TRUE )
both.EXP.pred <– e0.jmale.predict(EXP.pred, estimates = NULL, gap.lim = c(0, 18),max.e0.eq1.pred
= 83, save.as.ascii = 1000, verbose = TRUE)
predEXP3 <– get.e0.prediction(sim.dir=sim.dir.e0)
summary(EXP3, meta.only=TRUE)
summary(predEXP3, country=”Pakistan”)
summary(EXP3, country=”Pakistan”, par.names=NULL, thin=10, burnin=2000)
summary(EXP3, country=”Punjab”, par.names=NULL, thin=10, burnin=2000)
summary(EXP3, country=”Sindh”, par.names=NULL, thin=10, burnin=2000)
summary(EXP3, country=”KPK”, par.names=NULL, thin=10, burnin=2000)
summary(EXP3, country=”Balochistan”, par.names=NULL, thin=10, burnin=2000)
e0.trajectories.plot(both.EXP.pred, country=”Pakistan”, pi=c(80, 95), nr.traj=100, years
= 50, both.sexes = TRUE)
e0.trajectories.plot(both.EXP.pred, country=”Punjab”, pi=c(80, 95), nr.traj=100, years
= 50, both.sexes = TRUE)
e0.trajectories.plot(both.EXP.pred, country=”Sindh”, pi=c(80, 95), nr.traj=100, years
= 50, both.sexes = TRUE)
e0.trajectories.plot(both.EXP.pred, country=”KPK”, pi=c(80, 95), nr.traj=100, years
```

```
= 50, both.sexes = TRUE)
e0.trajectories.plot(both.EXP.pred, country=”Balochistan”, pi=c(80, 95), nr.traj=100,
years = 50, both.sexes = TRUE)
e0.trajectories.table(predEXP3, country=”Pakistan”, pi=c(80, 95),both.sexes = TRUE
)
e0.trajectories.table(predEXP3, country=”Punjab”, pi=c(80, 95), both.sexes = TRUE)
e0.trajectories.table(predEXP3, country=”KPK”, pi=c(80, 95), both.sexes = TRUE)
e0.trajectories.table(predEXP3, country=”Sindh”, pi=c(80, 95), both.sexes = TRUE)
e0.trajectories.table(predEXP3, country=”Balochistan”, pi=c(80, 95), both.sexes = TRUE)
e0.trajectories.plot(predEXP3, country=”Pakistan”, pi=c(80, 95), nr.traj=100, years =
50)
e0.trajectories.plot(predEXP3, country=”Punjab”, pi=c(80, 95), nr.traj=100, years =
50)
e0.trajectories.plot(predEXP3, country=”KPK”, pi=c(80, 95), nr.traj=100, years =
50)
e0.trajectories.plot(predEXP3, country=”Sindh”, pi=c(80, 95), nr.traj=100, years =
50)
e0.trajectories.plot(predEXP3, country=”Balochistan”, pi=c(80, 95), nr.traj=100, years
= 50)
e0.trajectories.plot(predEXP3, country=”World”, pi=c(80, 95), nr.traj=100, years =
50)
e0.DLcurve.plot(country=”Pakistan”, mcmc.list=EXP3, burnin=2000, pi=c(80, 95),
nr.curves=100)
e0.DLcurve.plot(country=”Punjab”, mcmc.list=EXP3, burnin=2000, pi=c(80, 95), nr.curves=100)
e0.DLcurve.plot(country=”Sindh”, mcmc.list=EXP3, burnin=2000, pi=c(80, 95), nr.curves=100)
e0.DLcurve.plot(country=”KPK”, mcmc.list=EXP3, burnin=2000, pi=c(80, 95), nr.curves=100)
e0.DLcurve.plot(country=”Balochistan”, mcmc.list=EXP3, burnin=2000, pi=c(80, 95),
nr.curves=100)
e0.partraces.cs.plot(country= ”Pakistan”, mcmc.list = EXP3, sim.dir = sim.dir.e0,par.names
= e0.parameter.names.cs(), nr.points = 1000, dev.ncol = 3, low.memory = TRUE)
```

```
e0.partraces.cs.plot(country= ”Punjab”, mcmc.list = EXP3, sim.dir = sim.dir.e0,par.names
= e0.parameter.names.cs(), nr.points = 1000, dev.ncol = 3, low.memory = TRUE)
e0.partraces.cs.plot(country= ”Sindh”, mcmc.list = EXP3, sim.dir = sim.dir.e0,par.names
= e0.parameter.names.cs(), nr.points = 1000, dev.ncol = 3, low.memory = TRUE)
e0.partraces.cs.plot(country= ”KPK”, mcmc.list = EXP3, sim.dir = sim.dir.e0,par.names
= e0.parameter.names.cs(), nr.points = 1000, dev.ncol = 3, low.memory = TRUE)
e0.partraces.cs.plot(country= ”Balochistan”, mcmc.list = EXP3, sim.dir = sim.dir.e0,par.names
= e0.parameter.names.cs(), nr.points = 1000, dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”Pakistan”, mcmc.list = EXP3, sim.dir = sim.dir.e0,
par.names = e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”Punjab”, mcmc.list = EXP3, sim.dir = sim.dir.e0, par.names
= e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”Sindh”, mcmc.list = EXP3, sim.dir = sim.dir.e0, par.names
= e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”KPK”, mcmc.list = EXP3, sim.dir = sim.dir.e0, par.names
= e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”Balochistan”, mcmc.list = EXP3, sim.dir = sim.dir.e0,
par.names = e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE)
e0.pardensity.cs.plot(country= ”World”, mcmc.list = EXP3, sim.dir = sim.dir.e0, par.names
= e0.parameter.names.cs(),dev.ncol = 3, low.memory = TRUE) e0.pardensity.cs.plot(country=
”Asia”, mcmc.list = EXP3, sim.dir = sim.dir.e0, par.names = e0.parameter.names.cs(),dev.ncol
= 3, low.memory = TRUE)
```

```
############ Population############
library(bayesPop)
sim.dir.pop <– file.path(getwd(), ”POPULATION”)
POPULATION.pred <– pop.predict(output.dir = sim.dir.pop, inputs = list(tfr.sim.dir
= simulation.dir, e0F.sim.dir = sim.dir.e0, e0M.sim.dir = ”joint_”), keep.vital.events =
TRUE, verbose = TRUE, replace.output=TRUE)
PREDICT <– get.pop.prediction(sim.dir= sim.dir.pop)
```

```
Country <− ”Pakistan”
summary(pop.pred, country)
pop.trajectories.plot(pop.pred, country = country, sum.over.ages = TRUE)
```

For other regions is also done in a similar way.

The R packages *bayesPop*, *bayesLife*, and *bayesTFR* were implemented with some amendments. The packages are available on Cran. Further detailed information and help can be taken from Ševcıková *et al.* (2014); Ševčíková *et al.* (2011); Ševčíková and Raftery (2016).

Bibliography

Alkema, L., Gerland, P., Raftery, A. and Wilmoth, J. (2015) The united nations probabilistic population projections: an introduction to demographic forecasting with uncertainty. *Foresight (Colchester, Vt.)* **2015**(37), 19.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J. and Pelletier, F. (2008) Estimating the total fertility rate from multiple imperfect data sources and assessing its uncertainty. Technical report, Working Paper 89). Seattle: Center for Statistics and the Social Sciences, University of Washington. Retrieved from http://www. csss. washington. edu/Papers/wp89. pdf.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T. and Heilig, G. K. (2009, September) Probabilistic projections of the total fertility rate for all countries `http://iussp2009.princeton.edu/papers/93084`.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T. and Heilig, G. K. (2011) Probabilistic projections of the total fertility rate for all countries. *Demography* **48**(3), 815–839.

Armingeon, K., Isler, C., Knöpfel, L., Weisstanner, D. and Engler, S. (2013) *Codebook: Comparative political data set 1960-2013.* Bern: Institute of Political Science, University of Bern.

Azose, J. J. and Raftery, A. E. (2015) Bayesian probabilistic projection of international migration. *Demography* **52**(5), 1627–1650.

Azose, J. J., Ševčíková, H. and Raftery, A. E. (2016) Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences* **113**(23), 6460–6465.

Basten, S., Coleman, D. and Baochang, G. (2012) A re-examination of the fertility assumptions in the un's 2010 world population prospects: Intentions and fertility recovery in east asia? .

Bijak, J. (2010) *Forecasting international migration in Europe: A Bayesian view.* Volume 24. Springer Science & Business Media.

Bijak, J. (2011) Forecasting migration: Selected models and methods. In *Forecasting International Migration in Europe: A Bayesian View*, pp. 53–87. Springer.

Bijak, J. and Bryant, J. (2016) Bayesian demography 250 years after bayes. *Population studies* **70**(1), 1–19.

Bongaarts, J. (2002) The end of the fertility transition in the developed world. *Population and development review* **28**(3), 419–443.

Bongaarts, J. (2009) Trends in senescent life expectancy. *Population studies* **63**(3), 203–213.

Cairns, S., Robinson, D. and Loiselle, D. (2008) Double-sigmoid model for fitting fatigue profiles in mouse fast-and slow-twitch muscle. *Experimental physiology* **93**(7), 851–862.

Casterline, J. B. (2001) The pace of fertility transition: national patterns in the second half of the twentieth century. *Population and Development Review* **27**, 17–52.

Chib, S. and Greenberg, E. (1995) Understanding the metropolis-hastings algorithm. *The american statistician* **49**(4), 327–335.

Chunn, J. L., Raftery, A. E. and Gerland, P. (2010) Bayesian probabilistic projections of life expectancy for all countries. Technical report, Working Paper.

Coale, A. and Watkins, S. (1992) The decline of fertility in europe (princeton, nj, 1986); j. *C. Chesnais (trans. E. Kreager and P. Kreager), The demographic transition: stages, patterns, and economic implications* .

Coale, A. J. and Trussell, T. J. (1974) Model fertility schedules: variations in the age structure of childbearing in human populations. *Population index* pp. 185–258.

Craig, J. (1993) Replacement level fertility and future population growth. *Population Trends* (78), 20–22.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. (1990) Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association* **85**(412), 972–985.

Gelman, A. (1993) Iterative and non-iterative simulation algorithms. *Computing Science and Statistics* pp. 433–433.

Gelman, A., Bois, F. and Jiang, J. (1996) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* **91**(436), 1400–1412.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014) *Bayesian data analysis*. Volume 2. CRC press Boca Raton, FL.

GENERAL, A. (n.d) Demographic models `http://www.un.org/esa/population/publications/Manual_X/Manual_X_Chapter_1.pdf`.

Girosi, F. and King, G. (2008) *Demographic forecasting*. Princeton University Press.

Hastings, W. K. (1970) Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1), 97–109.

Keyfitz, N. (1972) On future population. *Journal of the American Statistical Association* **67**(338), 347–363.

Lee, R. D. and Carter, L. R. (1992) Modeling and forecasting us mortality. *Journal of the American statistical association* **87**(419), 659–671.

Lee, R. D. and Tuljapurkar, S. (1994) Stochastic population forecasts for the united states: Beyond high, medium, and low. *Journal of the American Statistical Association* **89**(428), 1175–1189.

Lipovetsky, S. (2010) Double logistic curve in regression modeling. *Journal of Applied Statistics* **37**(11), 1785–1793.

Liseo, B. (2016) Class lecture: Mc and mcmc methods:monte carlo, importance sampling, gibbs sampler and the metropolis-hastings algorithm.

Mason, K. O. (1997) Explaining fertility transitions. *Demography* **34**(4), 443–454.

Meyer, P. (1994) Bi-logistic growth. *Technological forecasting and social change* **47**(1), 89–102.

Mubarak, Z. (2017) Afghans were counted alongside pakistanis in census: chief census commissioner. *Dawn news,Accessed: 2017-10-18, https://www.dawn.com/news/1357123/* .

Müller, P. (1991) *A generic approach to posterior integration and Gibbs sampling*. Purdue University, Department of Statistics.

Neal, R. M. (2003) Slice sampling. *Annals of statistics* pp. 705–741.

NIPS, I. (2012) Pakistan demographic and health survey 2012–13. *Secondary Pakistan Demographic and Health Survey* **13**, 2013.

Oeppen, J. and Vaupel, J. W. (2002) Broken limits to life expectancy. *Science* **296**(5570), 1029–1031.

Government of Pakistan, F. B. O. S. (2005) Pakistan demographic survey - 2005. `http://www.pbs.gov.pk/content/pakistan-demographic-survey-2005`.

Government of Pakistan, F. B. O. S. (2006) Pakistan demographic survey - 2006. `http://www.pbs.gov.pk/content/pakistan-demographic-survey-2006`.

Government of Pakistan, F. B. S. (2001) Pakistan demographic survey - 2001. http://www.pbs.gov.pk/content/pakistan-demographic-survey-2001.

Government of Pakistan, F. B. S. (2003) Pakistan demographic survey - 2003. http://www.pbs.gov.pk/content/pakistan-demographic-survey-2003.

Government of Pakistan, F. B. S. (2007) Pakistan demographic survey - 2007. http://www.pbs.gov.pk/content/pakistan-demographic-survey-2007.

Government of Pakistan, P. B. O. S. (2017) Population census 2017. http://www.pbscensus.gov.pk/.

Government of Pakistan, P. B. S. (2013) Population by province/region since 1951. http://www.pbs.gov.pk/sites/default/files//tables/POPULATION%20BY%20PROVINCE%20REGION%20SINCE%201951.pdf.

Government of Pakistan, P. B. S. (2015) Pakistan labour force survey 2014-15 (33rd issue). section-ii, reliablity of estimate. http://www.pbs.gov.pk/sites/default/files//Annual%20Report%20of%20LFS%202014-15.pdf.

Government of Pakistan, P. B. S. (2016) Social indicators of pakistan. http://www.pbs.gov.pk/sites/default/files//SOCIAL%20INDICATORS%202016%20%20%28FINAL%29%20%20COLOUR%201.pdf.

Government of Pakistan, P. B. S. (August, 2017) Monthly bulletin of statistics, august 2017. http://www.pbs.gov.pk/content/monthly-bulletin-statistics-august-2017.

Government of Pakistan, P. B. S. (nd) History of census. http://www.pbs.gov.pk/content/population-census.

Preston, S., Heuveline, P. and Guillot, M. (2000) Demography: measuring and modeling population processes .

Rafi, Y. (2015) Pakistan hosts second largest refugee population globally. https://www.dawn.com/news/1188585.

Raftery, A. E., Alkema, L. and Gerland, P. (2014a) Bayesian population projections for the united nations. *Statistical science: a review journal of the Institute of Mathematical Statistics* **29**(1), 58.

Raftery, A. E., Chunn, J. L., Gerland, P. and Ševčíková, H. (2013) Bayesian probabilistic projections of life expectancy for all countries. *Demography* **50**(3), 777–801.

Raftery, A. E., Lalic, N. and Gerland, P. (2014b) Joint probabilistic projection of female and male life expectancy. *Demographic research* **30**, 795.

Raftery, A. E. and Lewis, S. (1992) How many iterations in the gibbs sampler? In *In Bayesian Statistics 4*, pp. 763–773.

Raftery, A. E. and Lewis, S. (1996) Implementing mcmc. In *In Markov Chain Monte Carlo in Practice*, pp. 115–130.

Raftery, A. E., Li, N., Ševčíková, H., Gerland, P. and Heilig, G. K. (2012) Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences* **109**(35), 13915–13921.

Roper, L. D. (2000) Using sigmoid and double-sigmoid functions for earth-statestransitions. *Personal copy not assigned to a journal, available online: http://www. roperld. com/Science/DoubleSigmoid. pdf* .

Sathar, Z. A. (2001) Fertility in pakistan: past, present and future “. *Prospects for fertility decline in high fertility countries* .

Sevcikova, H., Alkema, L., Raftery, A., Fosdick, B. and Gerland, P. (2017) Package ‘bayestfr’ .

Ševčíková, H., Alkema, L. and Raftery, A. E. (2011) bayestfr: An r package for probabilistic projections of the total fertility rate. *Journal of Statistical Software* **43**(1), 1.

Ševčíková, H. and Raftery, A. (2011) bayeslife: Bayesian projection of life expectancy r package version 3.2-0.

Ševcıková, H., Raftery, A. and Gerland, P. (2014) Bayesian probabilistic population projections: Do it yourself. In *annual meeting of Population Association of America. http://paa2014. princeton. edu/abstracts/141301.*

Ševčíková, H. and Raftery, A. E. (2016) bayespop: Probabilistic population projections. *Journal of statistical software* **75**.

Siegel, J. S. and Hamilton, C. H. (1952) Some considerations in the use of the residual method of estimating net migration. *Journal of the American Statistical Association* **47**(259), 475–500.

TARIQ, H. (2016) Census: To be or not to be? `http://www.pakistantoday.com.pk/2016/11/27/census-to-be-or-not-to-be/`.

Taylor, J. (2009) Package 'hett' .

Tierney, L. (1994) Markov chains for exploring posterior distributions. *the Annals of Statistics* pp. 1701–1728.

Trovato, F. and Odynak, D. (2011) Sex differences in life expectancy in canada: immigrant and native-born populations. *Journal of biosocial science* **43**(3), 353–367.

UNHCR (2012) Overview of refugee & idp population and unhcr operational presence,pakistan. `http://unhcrpk.org/wp-content/uploads/2012/04/Pakistan_Refugees_Overview_August16_Ver1.pdf`.

United Nations, D. o. E. and Social Affairs, P. D. (2006) Methodology of the united nations population estimates and projections, chapter vi. world population prospects: The 2004 revision. 2006. `http://www.un.org/esa/population/publications/WPP2004/WPP2004_Volume3.htm.`

United Nations, D. o. E. and Social Affairs, P. D. (2009) World population prospects: The 2008 revision `http://www.un.org/esa/population/publications/wpp2008`.

United Nations, D. o. E. and Social Affairs, P. D. (2014) World population prospects: The 2012 revision https://esa.un.org/unpd/wpp/Publications/Files/WPP2012_Methodology.pdf.

United Nations, D. o. E. and Social Affairs, P. D. (2015) World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections. https://esa.un.org/unpd/wpp/Publications/Files/WPP2015_Methodology.pdf.

United Nations, D. o. E. and Social Affairs, P. D. (2017) World Population Prospects: The 2017 Revision, Methodology of the United Nations Population Estimates and Projections. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf.

Whelpton, P. K. (1928) Population of the united states, 1925 to 1975. *American Journal of Sociology* **34**(2), 253–270.

Whelpton, P. K. (1936) An empirical method of calculating future population. *Journal of the American Statistical Association* **31**(195), 457–473.

Winkler, R., Johnson, K. M., Cheng, C., Voss, P. R. and Curtis, K. J. (2013) County-specific net migration by five-year age groups, hispanic origin, race and sex 2000-2010 .

Yusuf, F., Martins, J. M., Swanson, D. A., Martins, J. M. and Swanson, D. A. (2014) *Methods of demographic analysis*. Springer.

MUHAMMAD ADIL

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 3511867486
e-mail: adil@stat.unipd.it
adil.so.fbs@gmail.com

Current Position

Since November 2014; (Thesis submission: 31st October, 2017)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Projecting Pakistan Population with a Bayesian Hierarchical approach
Supervisor: Prof. Stefano Mazzuco.

Research interests

- Projection and analysis of demographic components (Fertility, mortality, migration, population count)
- Bayesian Hierarchical modeling

Education

2005-06 – 2007-08
Master in Statistics .
University of Peshawar, Pakistan Faculty of Statistics.
Grade. A (Division 1st)

2010-11 – 2011-12
Master in Economics .
University of Peshawar, Pakistan Faculty of Economics.
Grade. B (Division 1st)

2007-08
Bachelor degree in Education.
University of Peshawar, Pakistan (Institute of Education and Research).
Grade. B (Division 1st)

2004-05 – 2005-06
Bachelor degree in Science.
University of Peshawar, Pakistan .
Major subjects: Mathematics and Statistics.
Grade. B (Division 1st)

Work experience

August 2007 – August 2008
Employer. Peshawar Degree College of Commerece and Business Administration
Appointment as: Lecturer in Statistics

August 2008 – October 2010
Employer. Shaikh Zayed Islamic Centre, University of Peshawar
Appointment as: Lecturer in Statistics

November 2010 – February 2011
Employer. Department of Statistics, University of Peshawar
Appointment as: Lecturer in Statistics
February 2011 – November 2014
Employer. Pakistan Bureau of Statistics
Appointment as: Statistical Officer

Awards and Scholarship

2008-09
Award or scholarship details. Internship award for one year from National Internship Program, Government of Pakistan.

2014 to 2017
Award or scholarship details. University of Padova departmental scholarship for Ph.D program in Statistics.

Computer skills

- SPSS
- Minitab
- R
- Stata
- Microsoft Office

Language skills

Urdu: native; Hindko: native; Punjabi: native; ; Pushto: native; English: fluent.

Conference presentations

Adil, Muhammad.,(2017). Probabilistic Projection of Total Fertility Rate for Pakistan and its regions using Bayesian Hierarchical modeling approach. (Oral presentation) *Giornate di Studio sulla Popolazione (Popdays 2017) 12th edition*, Florence, Italy, 8-10 February, 2017.

Adil, Muhammad., Mazzuco, Stefano., (2017). Probabilistic Projection of Pakistan Population and its regions using Bayesian Hierarchical modeling approach . (Oral presentation) *British Society for Population Studies (BSPS) Annual Conference,*, Liverpool, The United Kingdom, 6-8 September, 2017.

Adil, Muhammad., Mazzuco, Stefano., (2017). Probabilistic Projection of Total Fertility Rate and Life Expectation for Pakistan and its regions using Bayesian Hierarchical modeling approach . (Poster) *28th IUSSP International Population Conference* , Cape Town, South Africa, 29 October - 4 November, 2017.

Other Interests

Watching documentries, playing/watching cricket.

References

Prof. Stefano Mazzuco
Institution: Department of Statistical Sciences, University of Padova, Italy
Address: Via Cesare Battisti, 241, 35121 Padova
Phone: +39 049 8274192
e-mail: mazzuco@stat.unipd.it