

A Bayesian Prediction for the Total Fertility Rate of Afghanistan Using the Auto-regressive Integrated Moving Average (ARIMA) Model

SAYED RAHMI KHUDA HAQBIN AND ATHAR ALI KHAN

Department of Statistics and Operations Research
Aligarh Muslim University, Aligarh-202002, India
haqbin2020@gmail.com,
atharkhan1962@gmail.com

Abstract

In this article, a simple methodology to predict the total fertility rate of Afghanistan via a Bayesian statistical analysis method has been applied. R- statistical analysis tool is used for data analysis. To forecast, the "bayesforecast" package is needed. It is a substitute package in R for the "forecast" package in the traditional (frequentist) statistical method. The Bayesian data analysis using the specific case of the general auto-regressive integrated moving average model (ARIMA) is processed as follows; As the first step, the stationarity of the given data-set is assessed, the time series has been made stationary by taking differences. After fitting several models, as the most appropriate fitted model, the ARIMA (1, 2, 1) model has been fitted to the data. The accuracy of the fitted model is examined, and thereafter, the developed model is analyzed. The posterior computation is done, using the Markov Chain Monte Carlo (MCMC) simulation method. The method ultimately focuses on drawing relevant inferences including the 16 years prediction, and the results are; in general, found to be satisfactory.

Keywords: Fertility, Total Fertility Rate, Age -Specific Fertility Rate, Auto-regressive Integrated Moving Average Model, Stationarity, Hamiltonian Monte Carlo Algorithm, Prediction.

1. INTRODUCTION

Demography is the "study of the human population" [1]. The size of families and children's numbers had a substantially decreased trend over the 20th century. Especially after 1960, the family size and the number of children (fertility) have been showing a broadly favorable pattern for sustainable development. These changes in the population resulted from a declining fertility rate. The economic development level of a nation could be related to its fertility. The developed countries have been showing a lower level of fertility trend, along with a better education system, a revolutionary change in urbanization, greater wealth, and other factors. On the other hand, undeveloped countries have been showing a higher fertility trend, the increased level of fertility associated with a desire for family, labor, and caregivers in old age, lack of education, unawareness of contraceptives, strict adherence to traditional religious beliefs, and lower rates of female employees. Population growth and dependency ratios are directly affected by fertility changes. Over the past century, fertility has been one of the most determinants, of the population growth rate. Infant and child survival will be increased to provide greater access to education and health services, especially for women. By improving the participation of women in the labor force and empowering and reducing the number of children, women have over a lifetime. Low levels of fertility also contributed by improving maternal health, reducing the mortality rate of children, poverty alleviation, and economic development. Our objective; in this article includes

the future prediction of the fertility of a population. Clearly, for this purpose, we must use some summary index of fertility. The Total Fertility Rate is a composite measure of the fertility of a population, it is the most significant component for Fertility projections.

1.1. Total Fertility Rate

The total fertility rate calculates how many children a cohort woman will have. so it represents the average **Family Size** of the population. The total fertility rate is the sum of the Age-Specific Fertility Rate (ASFR) over all ages of the childbearing period and is expressed as "per woman" i.e., the sum divided by 1000. The average number of children a woman would have over her lifetime of reproductive potential, which is between the ages of 15 and 49, is the population's total fertility rate.

Forecasting the fertility rate has a long history of being the interest of demographers. The total number of live births in Australia, which is significant for population growth has been forecasted by using an auto-regressive moving average model [2]. Another most important demo-graphical forecasted characteristic could be the mortality forecast of the U.S. through the application of a time series method [3].

Fitted a bi-variate auto-regressive model which acted as a transform function model for forecasting the total fertility rate and mean ages of childbearing in the U.S annual data [4]. Age-specific mortality in the United States is predicted for the long term, with confidence intervals, using time series methods from 1990 to 2065. [5]. In a different publication, a parameterized model used the age profiles of fertility to quickly and simply describe age-specific rates. Future vital rate profiles are projected using time series models of the parameters, which reflect the temporal patterns of the age profiles [6].

A mixed two methods; statistical time series and mathematical demography and applied to predict stochastic models of fertility rates and the mortality rates of the U.S. dataset then by using the random-matrix product theory forecasted various measures of demography [7]. In a well-done study, a range of non-stationary dynamic factor models were used to jointly estimate the total fertility rate changes within the rather homogeneous clusters of Southern European countries in order to investigate the viability of the forecast. [8]. In this research, three methods for forecasting demographic processes are used. The population forecasting for the demographic data from 1980 to 2005 is particularly intriguing. Exogenous variable structural modelling based on theory and expectation [9]. A recent work applied ARIMA model for the total fertility rate of Pakistan is similar work of the paper from a classical perspective [10]. One another work lately has been conducted by using the two-time series models (ARIMA and ARIMAX) for live birth forecasting in Nigeria, which is a measure of population changes [11]. Numerous methods for predicting fertility have been mentioned; the ones that have been discussed thus far were mostly developed for wealthy nations that have already passed the fertility transition stage and currently exhibit low fertility patterns. Another important consideration is that the developed methods are used from traditional viewpoints. Pre-transition (high fertility), the fertility transition, and post-transition are the three phases of the changes in total fertility rate in a Bayesian projection model for the total fertility rate of country-specific forecasts of all countries (low fertility). The model is based on the most recent data from the United Nations Population Division [12].

In the same way, we aim to model the TFR of Afghanistan via the ARIMA model and achieve it is 10 years forecast by harnessing the attributes of the Bayesian paradigm. The procedures of this article are as follows:

- A real dataset of the total fertility rate (TFR) of Afghanistan has been provided.
- By taking two time differences the non-stationary time series came into the form of stationary series.
- The given data set has been analyzed from the Bayesian perspective using the "Bayesforecast" package, in R statistical language and software.

- "bayesforecast" package in R is a package provides many functions for data analysis in the Bayesian prospective.
- "bayesforecast" package is a Stan-ended pack.
- Stan is probabilistic programming language, which is made easy the data analysis, when there is complex probabilistic model.
- Stan work by taking sample of the equivalent to the posterior distribution fo the model, by different algorithms.
- NUTS(MCMC) algorithm, which is Stan sampling algorithm is applied in this article for the sampling purpose of the posterior distributions.
- The model fitting procedure is done by the help of Bayesian model fitting criteria "WAIC" and "LOOIC".
- The ACF and PACF plot of the original data and differenced data has been given, it could be a proper judge property for model fitting.
- After all the fitted and proper model which is decided for the given data is; ARIMA (1,2,1) model.

2. METHODS

In this section, very briefly the method which is applied for this study are expressed. The Total Fertility Rate data-set in Afghanistan has been given. The assigned object is Bayesian prediction of Total Fertility rate in Afghanistan at least for 16 future years. For further analysis, we have been applied Bayesian method of forecasting. The Bayesian substitute **R forecast package** (in classical statistics), is **bayesforecast R package**. The stationarity of the data tested, in a Bayesian standpoint, about the model decided. ARIMA(1,2,1) model selected for further analysis. After all analysis series, we reach to a satisfactory result.

2.1. ARIMA(p,d,q) Model

A time series Y_t is said to follows an auto-regressive integrated moving average (ARIMA) model if the d^{th} difference $Z_t = \nabla^d Y_t$ is stationary and follows (ARMA) process. If $Z_t \sim ARMA(p, q)$ model, we say that Y_t is an ARIMA (p, d, q) process. Where the three parameters; "p" in the Auto-regressive terms represents the current values depending on it's $p - previous$ values (lag values), while "q" in the moving average terms represents the current deviation from the mean depending on q's previous deviations and "d" is the number of differences, by taking, it makes a non-stationary time series in the format of stationary. When stationary is not an issue, then we can define an auto-regressive moving average (ARMA) model as follows;

$$Z_t = e_t + \phi_1 Z_1 + \phi_2 Z_2 + \dots + \phi_p Z_t - \theta_1 e_1 - \theta_2 e_2 - \dots - \theta_q e_t \quad (1)$$

$$Z_t = e_t + \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=1}^q \theta_j e_{t-j} \quad (2)$$

Where; e_t is a white noise process, and ϕ_i parameter which is the coefficient of the authoritative terms Z_{t-i} , and θ_j the parameter, which is the coefficient of the moving average e_{t-j} terms. Using the backshift operator, we can write this more succinctly, and ϕ_1, \dots, ϕ_p are the auto-regressive parameters to be estimated, $\theta_1, \dots, \theta_q$ are the moving average parameters to be estimated and e_1, \dots, e_t a series of unknown random errors (residuals) that are assumed to follow a normal distribution. The (ARMA) model can be simplifies when Box-Jenkins back-shift operator applied, simplified (ARMA) model as follows:

The auto-regressive AR(p) model;

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} = e_t$$

Where;

$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ are polynomials in B of degrees of p and q respectively. For the process to be stationary the root of $\phi(B) = 0$ must lie outside the unit circle, if the root of $\theta(B) = 0$ is also outside the unit circle, it is called the invertible, there is a unique model corresponding to the likelihood in that case.

we can write Equation 1 in the form of;

$$\phi(B)Z_t = \theta(B)e_t \tag{3}$$

The Equation 3 is the Backshift form of the ARMA (p, q) and the experience suggests in data analysis models with AR and MA components often fit better data than the AR and MA pure. at the same method from Equation 2 we can write;

$$\begin{aligned} (1 - \sum_{i=1}^p \phi_i B^i)Z_t &= (1 - \sum_{j=1}^q \theta_j B^j)e_t \Rightarrow (\sum_{i=1}^p \phi_i B^i)Z_t = (\sum_{j=1}^q \theta_j B^j)e_t \\ \phi_p(B)Z_t &= \theta_q(B)e_t \end{aligned} \tag{4}$$

The above (ARMA) model can be extended and written using differences;

$$BZ_t = Z_{t-1}$$

The difference operator $\nabla = 1 - B$ is;

$$BZ_t = (1 - B)Z_t = Z_t - Z_{t-1}.$$

Power in the Backshift operator is equal to the possible difference that a time series can take;

$$B^2 Z_t = B(BZ_t) = B(Z_{t-1}) = Z_{t-2}$$

for the second difference we can write;

$$\nabla^2 Z_t = (1 - B)(1 - B)Z_t = (1 - 2B + B^2)Z_t = Z_t - 2Z_{t-1} + Z_{t-2}$$

eventually the "d" difference can be write;

$$Y_t - \sum_{k=1}^d Y_t = (1 - B)^d Z_t \tag{5}$$

"d" is the order of differencing, in that case. The formal ARIMA (p, d, q) model in the Backshift form is produced by replacing "Z_t" in the ARMA model with the differences described above [15].

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)e_t \tag{6}$$

The Equation 6 is the Backshift format of the ARIMA(p,d,q) model, with "d" difference.

3. DATA AND MODEL SPECIFICATION

There is a yearly data-set, the total fertility rate of Afghanistan, from the year of 1983-2022 reported by the; (World bank open data, free and open access to global development data), the data-set is taken from the; <https://www.macrotrends.net/countries/AFG/afghan/fertility-rate>. For a better accuracy the total fertility rate data-set in Afghanistan for the year of 1983-2022 are presented as bellow;

Table 1: Afghanistan - Historical Total Fertility Rate form 1982 to 2022.

1982-1992	7.450	7.454	7.458	7.461	7.465	7.469	7.472	7.474	7.477	7.479	7.482
1993-2003	7.516	7.551	7.585	7.620	7.654	7.560	7.465	7.371	7.276	7.182	7.041
2004-2014	6.900	6.760	6.619	6.478	6.272	6.066	5.859	5.653	5.447	5.269	5.090
2015-2022	4.912	4.733	4.555	4.414	4.273	4.133	3.992				

The allocated projection of Afghanistan’s total fertility rate for the next 16 years is the **objective** of this article. Graphical illustration is the first step in characterising a data-set for the analysis technique; thus, as a time series data-set has been provided, the most appropriate plot for the given data-set is time plot, as shown in Figure 1. From a plotted data, the below tips could be found;

- Patterns
- Unusual observations
- Change over time
- Relationship between variables

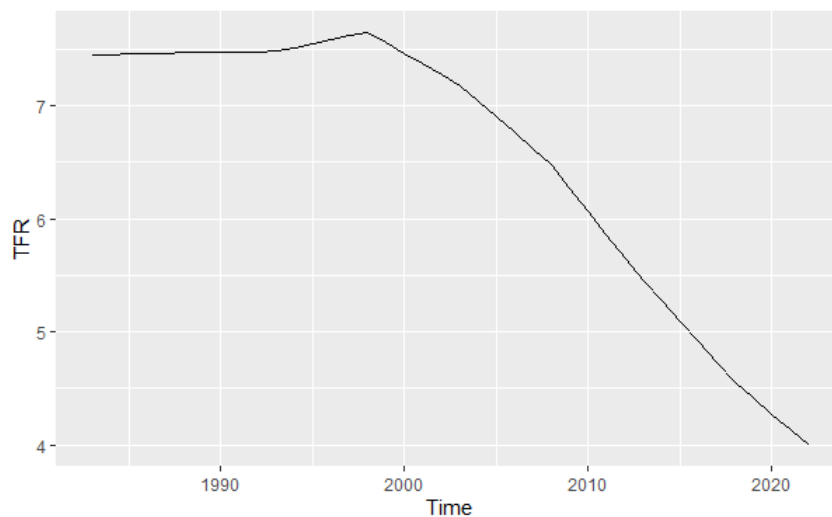


Figure 1: Displays the time series plot of the Total Fertility Rate of Afghanistan.

As we mentioned above, Figure 1 shows a constant pattern, from 1982 to 1998, a very short increase, afterwards, then the total fertility rate of the country has been showing a decreasing trend till 2022. Over all Figure 1 is visualized a decreasing trend, therefore, for fitting ARIMA model, testing the stationarity is the versatile condition. In a stationary time series, the statistical properties (mean, variance...) are independent of time, it does not depend on the time at which the series is observed. Time series with seasonality and trend patterns are not stationary, the trend and seasonality affect the value of the time series at different point of time. A series with the same pattern in any point of time is a stationary time series, (e.g: white noise series). Stationary time series will have not predictable pattern in the long- term [14].

3.1. Differencing to eliminate a trend

Differencing is the way to remove trends (non-stationarity) in a time series data. A difference operator is defined as;

$$\nabla y_t = y_t - y_{t-1}, \tag{7}$$

as well as, more broadly, for order d

$$\nabla^d y_t = (1 - \mathbf{B})^d y_t, \tag{8}$$

Where \mathbf{B} denotes the backshift operator (i.e., $\mathbf{B}^k y_t = y_{t-k}$ for $k \geq 1$) A random walk, for example, is one of the most basic and extensively used time series models, although it is not stationary. We may create a random walk model like follows:

$$y_t = y_{t-1} + w_t, w_t \sim N(0, q). \tag{9}$$

When the difference operator is applied to Equation 9, it produces a time series of Gaussian white noise errors w_t :

$$\begin{aligned} \nabla(y_t = y_{t-1} + w_t) \\ y_t - y_{t-1} = y_{t-1} - y_{t-1} + w_t \\ y_t - y_{t-1} = w_t \\ [16]. \end{aligned}$$

3.1.1 Making use of the 'diff()' function

There are many methods for making a non-stationary (time series with trend, seasonality, cyclic,...) data-set in the form of stationary. For the total fertility rate data-set, we applied two times differencing to making the time series stationary. Let's use R's diff() function to remove the trend from the total fertility rate time series. We will give some factors for identifying the stationarity of time series data; the correlation between a variable and itself at various time lags is known as the auto-correlation function (ACF). With the linear dependency of other $y_{t-1}, y_{t-2}, \dots, y_{t-k}$ eliminated, the partial autocorrelation function (PACF) assesses the linear correlation between a series y_t and a lagged version of itself y_{t+k} . In this part, for the very important point, we will

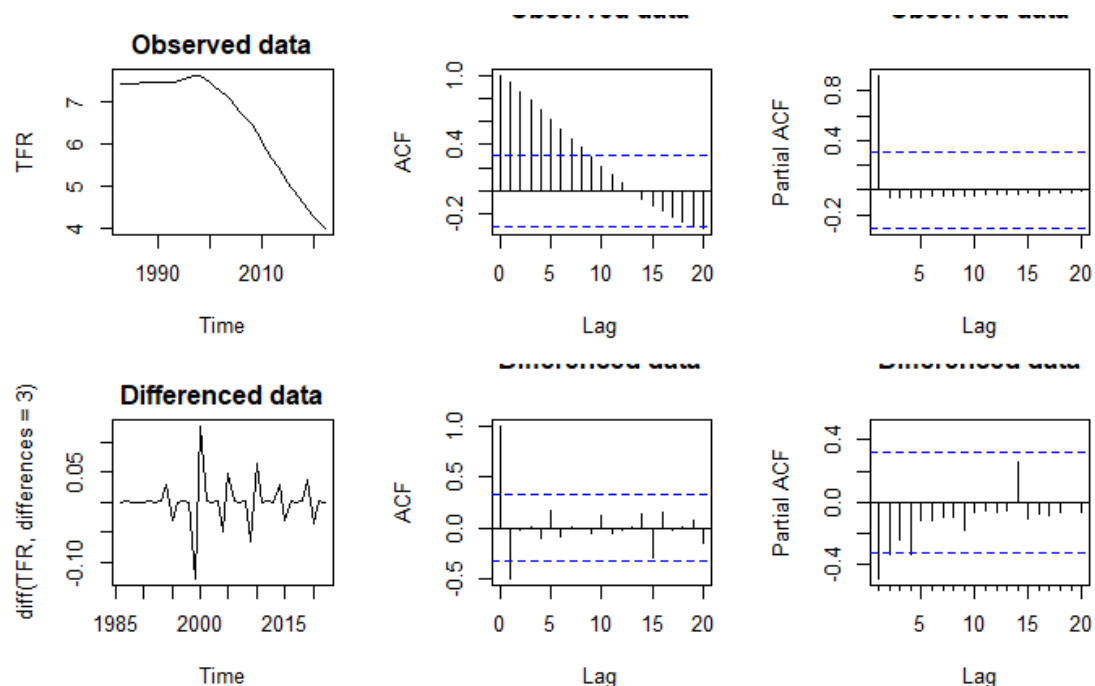


Figure 2: Displays the observed and differenced Total Fertility Rate of Afghanistan.

consider the Figure 2, and assesses it for further understanding of the process. In Figure 2 in the first row and first column is the time plot of the observed data, and the rest two other plots are the correlogram of the observed total fertility rate of Afghanistan, ACF and PACF respectively. The things we are going to learn from the correlogram are listed in bellow;

- The ACF at lag 0 is plotted as a
- reference point because it defaults to 1 (i.e., the correlation of a time series with itself).
- The blue horizontal lines represent the CIs for approximately 95%; and
- The auto-correlation is really high.

Figure 2 shows that although the partial autocorrelation at lag-1 is quite strong (equaling the ACF at $lag - 1$), the other values at $lags > 1$ are comparatively low, in contrast to what we observed for the ACF. Indices for the time lag are again real-valued in the PACF plot, but there is no value for lag-0 since it is impossible to eliminate any intermediate autocorrelation between t and $t - k$ when $k = 0$, and as a result, the PACF does not exist at lag-0. Now the second row of the Figure 2 shows a differenced and stationary of the total fertility rate of Afghanistan, for model identifying as per ACF and PACF which are given the second row of the Figure 2 second and third columns, the ACF show auto-correlation at lag-1 and PACF shows also correlation to lag-1 and decreasing correlation toward zero. Therefore, we will go to the other model fitting criteria of Bayesian approaches.

3.2. Model Specification

In a next step, after visualizing the time series data, checking stationarity, and removing trend of the data, we will approximate a model for the given data for further statistical analysis. The model used ought to call for the fewest parameters, whose values may be calculated from the observable series. Everything should be made as standardized but not simpler, according to a quotation from Albert Einstein in Parzen (1982, p. 68). For fitting a model to the data, few more steps are very important.

3.2.1 Bayesian Model Fitting Criteria

There are various methods for assessing and contrasting Bayesian models. The posterior predictive checks can be used to evaluate the model's fit to the data. It can be instructive to assess the prediction accuracy of each model under consideration, compare them, and decide what to do next if all of the models under consideration have discrepancies with the data [17].

As the total fertility rate data set is in use, we can do a WAIC and LOOIC criteria test for model selection.

Watanable-Akaike Information Criteria (WAIC): It has also been suggested to utilise the log pointwise posterior predictive density, with the posterior variance of the likelihood being used to derive the effective number of parameters.

$$lppd = \sum_{i=1}^n \log \int p(y_i|\theta)p(\theta|y)d\theta. \tag{10}$$

For more details; [18].

$$el\hat{p}pd_{waic} = l\hat{p}pd - \hat{p}_{waic} \tag{11}$$

$$WAIC = \sum_{i=1}^n \log E[p(y_i|\theta, y)] + 2 \sum_{i=1}^n var_{post}(\log p(y_i|\theta)). \tag{12}$$

For more details; [19].

Where the posterior mean of the likelihood of the i th observation is $E[p(y_i|\theta, y)]$ and $\sum_{i=1}^n var_{post}(\log p(y_i|\theta))$ the variance of each variable in the log predictive density summed over n data points. Prior knowledge is taken into account by the WAIC, and the posterior distribution is used in a non-normal way [16].

Leave-One-Out Cross Validation: When using the leave-one-out information criterion (LOOIC), $N-1$ observations are used as the validation sample, the process is repeated N times, resulting

in a different observation being predicted each time, and the prediction results are added up to produce an estimate of expected log predictive density (**elpd**) that closely resembles the outcomes that would be obtained by gathering new data and performing the validation. We put the calculations into a R package named "**loo**".

Table 2: Displays the WAIC and LOOIC.

ARIMA Models	LOOIC	WAIC
ARIMA(2,2,2)	-153.5	-155.5
ARIMA(0,2,2)	-155.8	-156.0
ARIMA(2,2,1)	-155.0	-156.8
ARIMA(2,2,0)	-156.0	-156.3
ARIMA(1,2,0)	-156.0	-155.5
ARIMA(0,2,1)	-156.6	-156.9
ARIMA(1,2,1)	-157.0	-156.9
ARIMA(1,2,2)	-154.8	-156.1

The WAIC and LOOIC are given for several ARIMA models, as per Table 2 ; ARIMA (1,2,1) Model has been showing the proper and best model among all other models, with lesser WAIC and LOOIC.

4. BAYESIAN FORMULATION

A Bayesian model is a parametric model as classical(or frequentest) model, but in Bayesian an addition prior probability distribution for the parameters of the model have to be defined. And the parameters of the model treat as a random variable rather to unknown constants. For a Bayesian model, the components are listed as bellow:

- the data, denoted by Z .
- the parameters, denoted by the Greek letters.
- the distributions of the model, given by a specification $f(Z|\theta)$ or $F(Z|\theta)$
- the prior distribution, specification of $f(\theta)$ or $F(\theta)$ or the distribution of the θ .

f is the probability distribution (cdf), F is a symbol which denotes the cumulative distribution density(cdf), θ is the prior.

4.1. Bayes' theorem

As a starting point for the Bayesian analysis, we would like to bring the Bayes' Theorem which is the important formula for Bayesian. Bayes' theorem is based on the conditional probability distribution:

$$f(\theta | Z) = \frac{f(Z | \theta)f(\theta)}{f(Z)} \tag{13}$$

$f(x)$ is the unconditional (or prior) pdf of x . The proportionality format of the Equation 13 is;

$$f(\theta | Z) \propto f(Z | \theta)f(\theta) \tag{14}$$

In the Equation 14 the $f(\theta | Z)$ is the posterior; the posterior is proportional to the likelihood times to the prior. The model is emphasized that the proportionality of the posterior is specifically related to the θ . As the demonstrator is independent of θ we easily ignore it.

4.2. Choosing the Prior Distribution of the Parameters

For the four parameters the location, scale, auto-regressive, and moving average parameters, we must describe a four-dimensional joint prior distribution in the general case. However, it is common practise to assume the parameters' prior independence, which means that before examining the data, we are unaware of whether the parameters are positively or negatively connected. We can just define four distinct priors for the four parameters with independence. We investigate weakly informative prior distributions.

$$\begin{aligned} TFR &\sim ARIMA(1, 2, 1) \\ \mu_0 = \epsilon_0 &\sim t(0, 2.5, 7) \\ \sigma_0 &\sim t(0, 1, 7) \\ ar_1 &\sim N(0, 0.5) \\ ma_1 &\sim N(0, 0.5) \end{aligned}$$

we can write the proportional prior density of the defined parameters;

$$\begin{aligned} \pi_1(\epsilon_0) &\propto (1 + \frac{\epsilon_0^2}{7})^{-4} \\ \pi_2(\phi_1) &\propto \exp^{-\frac{\phi_1^2}{0.5}} \\ \pi_3(\theta_1) &\propto \exp^{-\frac{\theta_1^2}{0.5}} \\ \pi_4(\sigma_0) &\propto (1 + \frac{\sigma_0^2}{7})^{-4} \end{aligned}$$

4.3. Likelihood Function for the Model

$Z : Z_1, Z_2, \dots, Z_{T-d}$ are the observations, from Equation 2, the conditional density of Z_t over $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$ is given by;

$$f(Z_t | Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}; \epsilon_0, \phi_i, \theta_j, \sigma^2) \propto (\frac{1}{\sigma^2}) \exp^{-\frac{1}{2\sigma^2} (Z_t - \epsilon_0 - \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j})^2} \tag{15}$$

From Equation 15 the likelihood function of the given density, can be approximated by its conditional forms;

$$L(Z | \epsilon_0, \phi_i, \theta_j, \sigma^2) \propto \prod_{t=p+1}^{T-d} f(Z_t | Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}; \epsilon_0, \phi_i, \theta_j), \tag{16}$$

For more details; [17].

For simplification we can write;

$$L(Z | \epsilon_0, \phi_i, \theta_j, \sigma^2) \propto (\frac{1}{\sigma^2})^{\frac{(T-d-p)}{2}} \exp^{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{T-d} (Z_t - \epsilon_0 - \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j})^2} \tag{17}$$

where; $\phi_i = \phi_1, \phi_2, \dots, \phi_p, \theta_j = \theta_1, \theta_2, \dots, \theta_q$. If one has a sample size of T, and would like to estimate a ARMA(p,q), there are many ways to estimate the parameters of the model: For the two times differenced Total Fertility Rate series in Afghanistan, the Likelihood can be defined from Equation 17 as;

$$f(Z | \epsilon_0, \phi_1, \theta_1, \sigma^2) \propto (\frac{1}{\sigma^2})^{\frac{(T-2-1)}{2}} \exp^{-\frac{1}{2\sigma^2} \sum_{t=2}^{T-2} (Z_t - \epsilon_0 - \phi_1 Z_{t-1} - \theta_1 \epsilon_{t-1})^2} \tag{18}$$

where Z_t obviously ∇y_t [18].

4.4. Posterior Distribution Function of the Model

As, it has been discussed in Section 4.1, Bayesian approach is the implication of the Bayes' theorem for further details Section 4.1. The prior and likelihood for the given data-set (total fertility rate) have been defined, now by applying the Bayes' theorem we can reach to the posterior distribution

of the Bayesian distribution of the data-set, once the posterior defined, it will be easy to apply any probabilistic programming language for fitting the posterior in a model and estimate the parameters. From Section 4.1, the posterior;

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

$$\log(\text{posterior}) = \log(\text{prior}) + \log(\text{likelihood})$$

$$p(\epsilon_0, \phi_1, \theta_1, \sigma^2 | Z) \propto \left(1 + \frac{\epsilon_0^2}{\sigma^2}\right)^{-4} \exp^{-\frac{\phi_1^2}{0.5}} \exp^{-\frac{\theta_1^2}{0.5}} \left(1 + \frac{\sigma^2}{\sigma^2}\right)^{-4} \times$$

$$\left(\frac{1}{\sigma^2}\right)^{\frac{(T-3)}{2}} \exp\left(-\frac{1}{2\sigma^2}\right) \sum_{t=2}^{T-2} (Z_t - \epsilon_0 - \phi_1 Z_{t-1} - \theta_1 \epsilon_{t-1})^2 \quad (19)$$

From Equation 19, it is obvious which eventually we reach to a joint joint posterior distribution of all parameters, we can find the marginal density function of each parameter and then generate sample. As we obtained a joint posterior distribution of all parameters, we can find the marginal density function of each parameter and then generate sample. In this article, which is the total fertility rate data-set, in a Bayesian approach, we are going to analyse the data-set and estimate the parameters. As we know the most important step for a Bayesian data analysis is; to create a joint posterior function of all the parameters, now we have the joint posterior function of the parameters. There are many computational tools for the Bayesian data analysis, as we will discuss in Section 5, we only run the **Code in R**, in "**bayesforecast**" **package**' **function** `stan_sarima()` **function**, which is a Bayesian **forecast**' **package** in R, based of Stan, and using the MCMC simulation method, the sampling algorithm in this **function** is NUTS(MCMC).

4.5. Markov Chain Monte Carlo (MCMC)

The modern method for approximating complex forms of the posterior distribution is the Markov Chain Monte Carlo (MCMC) simulation method. The concept is akin to treating the posterior distribution as a population and then drawing samples from it repeatedly. When you draw a large enough sample (say 1,000), the sample distribution should be extremely close to the population distribution, as you learnt in basic statistics. The samples drawn in the above comparison are correlated, thus if the first sample is high, the second one is more likely to be high as well. This is necessary since there is no direct way to take samples from the posterior distribution, which typically has a very complex structure; instead, we have various procedures that can lead us to the posterior indirectly. Correlation between samples is usually not a big deal, except that we need to draw extra samples to compensate. In Sections 5.1 and 5.2 are given the values of the parameters and in Section 5.2 plots are showing the trace plot and the sample distribution of 4,000 samples using MCMC. The shape of the MCMC sample distribution approximates the shape of the posterior distribution [19].

5. BAYESIAN INFERENCE USING NUTS(MCMC)

NUTS is an extension of Hamiltonian Monte Carlo (HMC) which is a type of MCMC. NUTS can even sample from the model with a greater number of parameters, it is a strength of NUTS sampler algorithm. Stan is a very good probabilistic programming language, the only disadvantage is that, Stan cannot sample for discrete parameters. NUTS uses a recursive algorithm to build a set of likely candidate points that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps. Hamiltonian Monte Carlo is a Markov Chain Monte Carlo (MCMC) algorithm that avoids the random walk behavior and sensitivity to correlated parameters that plague many MCMC methods by taking a series of steps informed by first order gradient information [20].

We will describe how to give the data to the model created by the R package "bayesforecast" in this part. For forecasting, the "bayesforecast" package in R is a Stan end-packed package. We must use R to give the data as a certain type to the "bayesforecast" package's function for ARMA

model fitting in order to accomplish this and obtain the outcome. In "bayesforecast" package, the model can be estimated using the guidelines below:

```
stan_sarima(
  ts,
  order = c(1, 0, 0),
  seasonal = c(0, 0, 0),
  xreg = NULL,
  period = 0,
  chains = 4,
  iter = 2000,
  warmup = floor(iter/2),
  adapt.delta = 0.9,
  tree.depth = 10,
  stepwise = TRUE,
  prior_mu0 = NULL,
  prior_sigma0 = NULL,
  prior_ar = NULL,
  prior_ma = NULL,
  prior_sar = NULL,
  prior_sma = NULL,
  prior_breg = NULL,
  series.name = NULL,
  ...)
```

Using the function *stan_sarima()*'s from "bayesforecast" package in R the default setting are; *ts* =, The ARIMA(p,d,q) order, chains is equaling 4, iterations per chain equaling 1000, and 1000 warmups. used as a warmup to eliminate dependencies in the initial settings. We fit the selected model to the total fertility rate; by the function above.

5.1. The Estimated Results

The given Table 3: shows the numerical estimated values for the parameters of the fitted model ARIMA (1,2,1) on the total fertility rate of Afghanistan.

Table 3: Displays the estimated posterior distribution of the parameters

	Mean	Se	%5	% 95	N_{eff}	Rhat
Intercept	-0.0103	0.0001	-0.0215	-0.0007	4282.407	1.0000
σ_0	0.0270	0.0001	0.0221	0.0329	3940.089	1.0015
ϕ_1	0.9209	0.0008	0.8339	0.9872	4034.682	1.0000
θ_1	0.0658	0.0024	-0.1909	0.3134	4109.169	1.0005
loglik	77.7385	0.0323	73.8071	80.3939	3808.628	1.0034

Table 3 provides a summary of the statistics for the marginalized posterior distribution as well as the MCMC convergence diagnostics. In line 3, for instance, the marginalized posterior distribution for parameter ϕ_1 is presented. Now, let's examine each column. The name of the parameter is displayed in the first column. The posterior mean, also known as the mean of the marginalized posterior distribution, is the value in the second column. The arithmetic mean of all the **draws** in this case, a total of **4000** is used to calculate this. For instance, the posterior mean of θ_1 is **0.0658**, and this finding shows that, on average, the baseline annual income grows by **0.0658** with each extra year of work experience. The Monte Carlo standard errors of the Mean are shown in the third column as Se. By dividing the SD by the square root of $N - eff$, it is calculated. The quantiles of the posterior marginalized distributions are displayed in columns

four and five. These are calculated using the draw's quantiles. Stan determines the effective number of drawings as N_{eff} in column 6 using the autocorrelation of the draws. We believe this value should be at least 100 in order to estimate distributions and perform other statistical calculations. Additionally, a small value for this parameter suggests that it has probably not yet converged; this information could be used to improve the model. An indication of MCMC convergence, $R-hat$, is computed for each parameter and is shown in column seven. Typically, it compares the sample variance of each chain using findings from many chains. According to Section 11.4 [21].

The MCMC is said to have converged if it meets the condition $R_{hat} < 1.1$ for all the parameters. We must confirm that the MCMC is converged before evaluating the results. Beginners frequently skip over this stage in favour of moving on to the analysis' next step, when they could use the posterior mean, plot the histogram of draws, or quickly interpret the findings using an MCMC that hasn't even begun to converge. Without a doubt, avoid this. The *try-and-error* procedure must be repeated numerous times before the MCMC converges. Eventually the ARIMA (1,2,1) model for the time series data as per Equation 1 can be written:

$$Y_t = -0.0103 + 0.92y_{t-1} + 0.66e_{t-1} \tag{20}$$

From Table 3, the simulation estimate for $E(\phi_1|y)$ is 0.92009, according the 90% credible interval which is (0.8339 , 0.9872) of the Table 3 statistically significant. in the same way, the simulation estimate for $E(\theta_1|y)$ is 0.06258, it is not statistically significant 90% probability interval for the variable(parameter) θ_1 is (-0.1909 , 0.3134) . Intercept and the posterior standard deviation values are given respectively. In Bayesian statistics, the credible interval is analogous to the confidence interval, but with radically distinct implications. A 90% credible interval, for example, is an interval that has a 90% chance of containing the true value of the parameter, an interpretation that is frequently and incorrectly connected with the confidence interval. Because the population parameter is fixed, frequentists cannot use probability for the parameter; instead, only the sample is probabilistic, and a 90% confidence interval must be interpreted in the sense that 90% of the intervals constructed with repeated sampling will contain the true parameter. As we can see in the Table 3 a 90% credible interval is given for each parameter of the model.

5.2. The estimated results figure

The sported plot of the estimated values are given as Figure 3;

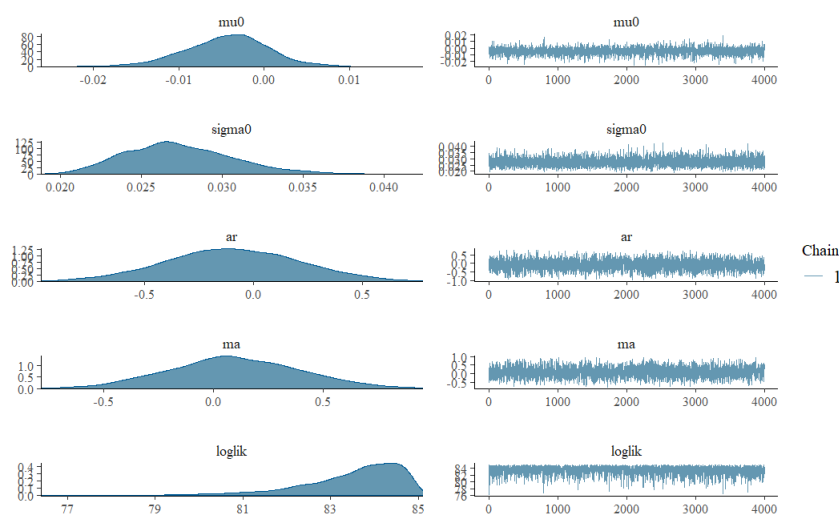


Figure 3: Displays the supported of the estimated parameters.

The probability density functions of the marginalised posterior distributions, which are

calculated from the draws after warm-up in each chain, are shown on the left column side. The right column, which additionally displays the trace plots following warm-up. These plots show that all of the chains are oscillating around a set of values after the warm-up, which indicates that the MCMC has converged.

6. MODEL CHECKING

A common procedure in Bayesian statistics for validating your model is the posterior predictive check [22]. We must first talk about the posterior predictive distribution in order to comprehend the posterior predictive check. The posterior predictive distribution in a mathematical notation.

$$p(\hat{y}|y) = \int_{\theta} p(\hat{y}|\theta, y)p(\theta|y)d\theta \tag{21}$$

From the Equation 21, θ is the parameter, in our posterior distribution which is discussed in the Section 4.4, θ is the set of all parameters $\theta=(\epsilon, \sigma, \phi_1, \theta_1)$. The formula is merely provided for your convenience because we won't be using it to make posterior predictions. As an alternative, we will employ simulations to roughly approximate the posterior predictive distribution of \hat{y} , which is really an extension of the posterior distribution of θ 's approximation. The posterior predictive check simply makes a comparison between the observed data and the model's prediction we proceed to review the the graphical posterior predictive distribution of the total fertility rate.

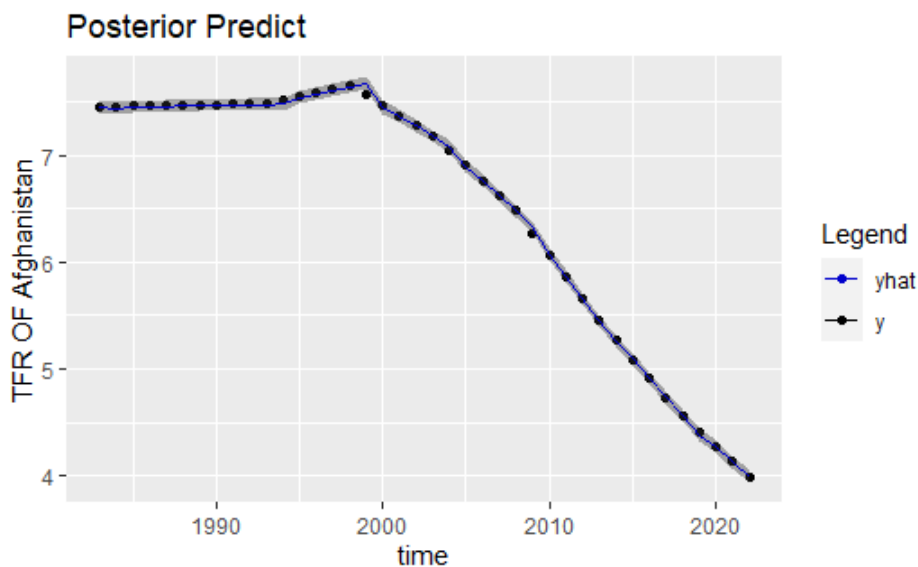


Figure 4: Displays the posterior predicted.

From Figure 4, we can see that; the fitted model on the generated draws $y - hat$ is very good fit. Therefore, we can results, the fitted model is a proper model for the given data-set.

6.1. Check Residuals

We may examine the residuals' autocorrelation using "`check_residuals()`" in the forecast package will automatically run a test and display a few common diagnosis charts. Figure 5 is the common diagnostics charts of the total fertility rate.

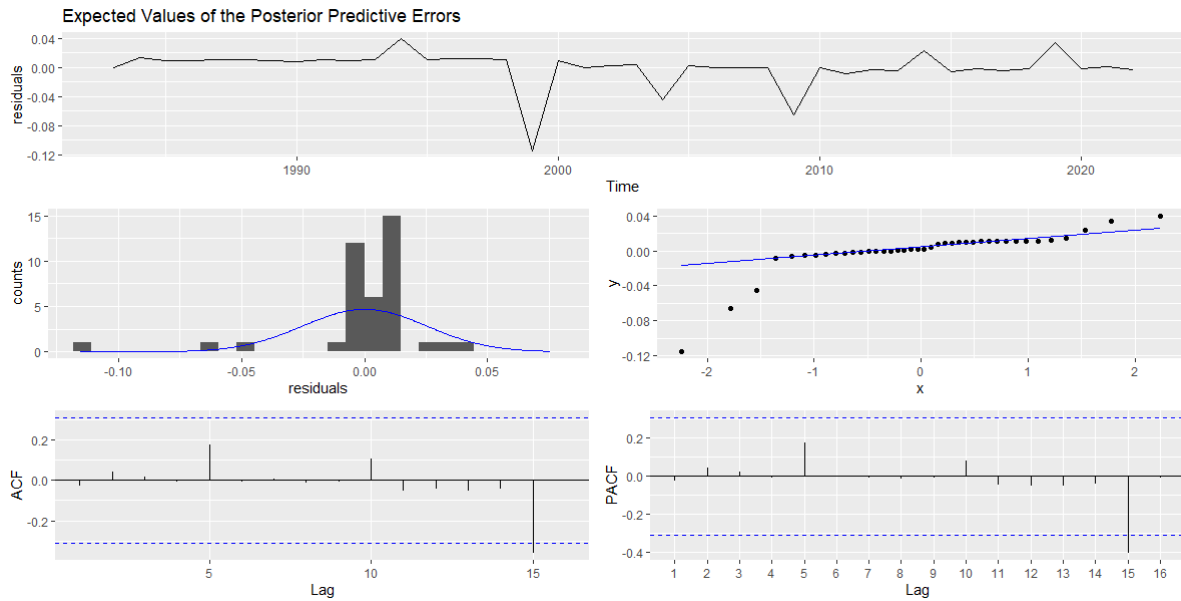


Figure 5: expected values of the posterior predictive errors.

The residuals series (upper part) shows that the given series is white noise. The dash blue lines are indicated the correlations are significantly not zero. In a white noise series, we expect each correlation to be zero. For a white noise series, we expect 95% of the spikes to lie within the $\frac{2}{\sqrt{T}}$, where T is the length of the series. If more than 5% of the spikes are not in the blue bounds, then the series is not white noise probably. The histogram and quantile graph (middle part) show that the model is around normally distributed. Finally, the residual auto-correlation presents have showed that there is no correlation between the series segments. A summary result from the above interpretation is; we can step ahead and do our prediction based, as all the evidences present a good and fair series.

7. PREDICTION

Forecasting can be used in many situations: for making a schedule to the staff of a call center for some times in the future, we need to do a forecast as per call going to be received. For an inventory or stocking there is need for forecast of stock requirement. Forecast can be for few minutes beforehand till several years in advance. For making an efficient and effective plan, need to do forecasting. Forecasting can be done very easy and sometimes very tough. predictability for an event or quantity related to several factors;

- Understanding the factors to contribute on the event.
- How much data are accessible.
- Is there any effect of forecasting to the things are going to be forecasted [23].

Forecasting fertility; as the demographic characteristics (fertility, mortality, migration) of the human population is significant for socio-economics planning. There are mentioned lots of fertility forecasting methods, the methods discussed on the Section 1.1. Mostly, they have developed for the countries which are rich today and already passed their fertility transition, and at the current time they have low fertility patterns. Another point, is the developed methods are applied from the classical perspectives. A Bayesian projection model for the total fertility rate of country-specified projections of all countries have been done, the developed model consisted of three phases for the changes in total fertility rate, pre-transition (high fertility), the fertility

transition, and post-transition (low fertility). The model has been built on the United Nations Population Division’s current [24].

In the same way, we aim to model the Total Fertility Rate of Afghanistan via the ARIMA model and achieve it is 16 years forecast, by harnessing the attributes of the Bayesian paradigm.

7.1. Forecasted Results

In the Table 4, which is numerical forecasted years, from 2022-2037 the 16 years of the total fertility rate of Afghanistan.

Table 4: Future forecasting

	Point Forecast	Lo.0.9	Hi.0.9	Lo.0.95	Hi.0.95
2023	3.853670	3.802970	3.903317	3.794032	3.915283
2024	3.718556	3.668241	3.768550	3.656223	3.776726
2025	3.586596	3.533241	3.639957	3.520394	3.650888
2026	3.455740	3.396801	3.510490	3.386231	3.520109
2027	3.332134	3.261010	3.388971	3.242005	3.400166
2028	3.207198	3.120882	3.278480	3.100771	3.290015
2029	3.085855	2.983132	3.162357	2.957278	3.173630
2030	2.967033	2.836914	3.055491	2.812259	3.068283
2031	2.851968	2.693900	2.959655	2.657657	2.972596
2032	2.739253	2.551682	2.859017	2.500849	2.878114
2033	2.626087	2.400823	2.762809	2.352485	2.776872
2034	2.519611	2.262385	2.677497	2.209586	2.693992
2035	2.411963	2.088992	2.588108	2.044437	2.604896
2036	2.311417	1.955690	2.510772	1.883435	2.533944
2037	2.210127	1.821591	2.432696	1.725132	2.452562
2038	2.114983	1.655969	2.371578	1.576980	2.388008

The variable’s values we are interested to forecast is unknown, therefore, it is a random variable. The random variable can take a range of random values. In Table 4, the first column is the years of forecast, the second column is the forecasted point or mean of the future values of the total fertility rate (which is the random variable). From columns 3-6 are the credible interval of 90% to 95%. For more details about the credible interval Section 5.1. The value for the total fertility rate of Afghanistan which is under consideration is given in Table 1, form 1982-2022. Afterwards, the value of the fertility of the country is random variable, because it is unknown and we are going to do forecast. In a Bayesian prediction we reach to the forecasting point which is the Table 4 is the numerical values of the forecasted amount.

7.2. Forecasted Results Figure

From Table 4, which is a 16 years Bayesian prediction of the total fertility rate of Afghanistan. Year of 2038 is showing a very good results. 2.1 is the Replacement fertility level: Total fertility of approximately 2.1 children per woman. This value shows the average number of children a woman would need to have in order to reproduce herself by having a daughter who reaches reproductive age. Now it is time to go to the graphical representation of the results. Figure 6, is going to support the numerical values of Table 4, which is the forecasted values of the total fertility rate of Afghanistan from 2023-2038.

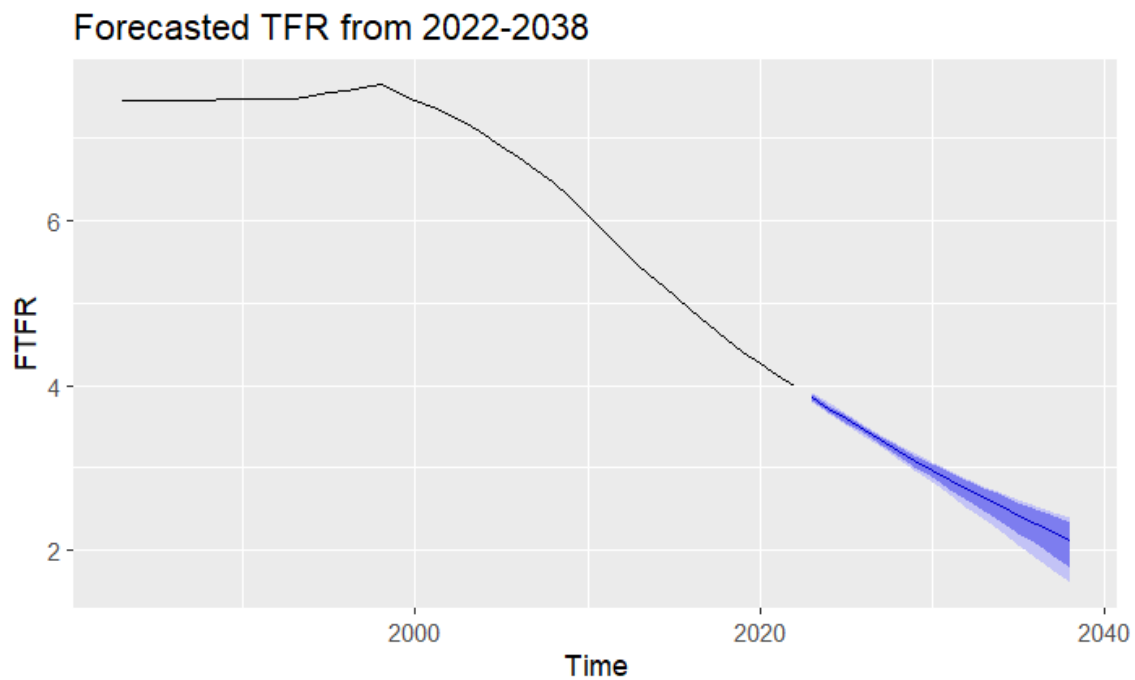


Figure 6: Future prediction

Figure 6 are consisted of two parts, from 1982-2022 the observed data, and 2023-2038 is the forecasted values of the total fertility rate of Afghanistan. The blue line in the Figure 6 shows the forecasted part of the value of the total fertility rate, which is the averages possible future value, and numerically was shown in the Table 4. In the Figure 6 it is obvious, there is two types of shadowed area, the 90% and the 95% credible intervals of the forecasted values. Forecast accuracy also can be analyzed form the shadowed areas, as much as, our forecasted time increase the credible interval or probability interval increase. And on the same way as much as years of forecast is near to the observed value the credible interval is more narrow and it shows the certainty or accuracy of the forecast.

8. CONCLUSION

In this article, we were started the work for doing Bayesian data analysis, using the total fertility rate of Afghanistan data-set. We discussed selecting a statistical model that is assumed to have generated the observed data, defining the prior and likelihood for the data, summarizing the posterior distribution, and using the posterior predictive check to assess the model's fit to the observed data. The model For the model ARIMA (1,2,1) has been chosen to the data-set. Model checking conducted, and 16 years forecasting has been done to the data. The total fertility rate of the country is projected to decline, and specifically it has been expected that approximately 2.1 (average number of children per woman) for the year 2038, this results is a very fundamental results that is the normal and standard measure of the fertility rate which is called replacement level of fertility. The importance of the demographic changes, obviously is a concern for the countries, for the effects on all affairs of human activity, economically, culturally, socially, and politically, therefore; the predicted value of the total fertility could be a good source for the government of the country for taken effective steps to control and decrease the fertility rate.

Need of the study: Afghanistan is a deve1oping country. The country has been suffering for a long time, for a long-term war perspective, awareness of this type of issue could be significant for the government and policymakers, and non-government organizations to plan in order to monitor the fertility rate of the country. There was no such study, so there is a great need for the study, it

can be a good hint for those, who are going to work on the demographic area of Afghanistan.

One important tip is needed to mention is that we have done a prediction for the total fertility rate of the country for a specific period of future time, based on the series period of past time, no prediction can be a real future, but it is a motivation work for those, who are concern in such study.

REFERENCES

- [1] 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.
- [2] Bolanle, A. (2017). Forensic accounting as a tool for fighting corruption and terrorism: Case study of nigeria. *Int. J. Soc. Sci. Econ. Res*, 2(1):2210–2225.
- [3] Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International journal of forecasting*, 22(3):547–581.
- [4] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley, Sons.
- [5] Carmichael, G. A. et al. (2016). *Fundamentals of demographic analysis: Concepts, measures, and methods*. Springer.
- [6] Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC press.
- [7] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- [8] Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016.
- [9] Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.
- [10] Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.
- [11] Holmes, E. E., Scheuerell, M. D., and Ward, E. (2020). *Applied time series analysis for fisheries and environmental data*. NOAA Fisheries, Northwest Fisheries Science Center, Seattle, WA.
- [12] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [13] Knudsen, C., McNowen, R., and Rogers, A. (1993). Forecasting fertility: An application of time series methods to parameterized model schedules. *Social Science Research*, 22(1):1–23.
- [14] Korner-Nievergelt, F., Roth, T., Von Felten, S., Gu?elat, J., Almasi, B., and Korner-Nievergelt, P. (2015). *Bayesian data analysis in ecology using linear models with R, BUGS, and Stan*. Academic Press.
- [15] Lee, R. D. (1993). Modeling and forecasting the time series of us fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9(2):187–202.
- [16] Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.
- [17] 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.
- [18] Marriott, J., Ravishanker, N., and Gelfand, A. E. (1993). *Bayesian analysis of arma processes: Complete sampling based inference under full likelihoods*. Technical report, STANFORD UNIVCA DEPT OF STATISTICS.
- [19] McDonald, J. (1979). A time series approach to forecasting australian total live births. *Demography*, 16(4):575–601.
- [20] McNowen, R. and Rogers, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography*, 26(4):645–660.

- [21] Ortega, J. A. and Poncela, P. (2005). Joint forecasts of southern european fertility rates with non-stationary dynamic factor models. *International Journal of Forecasting*, 21(3):539–550.
- [22] Raziq, A., Sheikh, N., Rafique, M., Murtaza, G., and Muhammad, A. (2021). Forecasting fertility and child mortality rates of pakistan using time series model. *Linguistica Antverpiensia*, 2021(2):1194–1213.
- [23] Tripathi, P. K., Mishra, R. K., and Upadhyay, S. (2018). Bayes and classical prediction of total fertility rate of india using autoregressive integrated moving average model. *Journal of Statistics Applications , Probability*, 7(2):233–244.
- [24] Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12).