*Annals of Pure and Applied Mathematics*
*Vol. 23, No. 1, 2021, 21-26*
*ISSN: 2279-087X (P), 2279-0888(online)*
*Published on 5 February 2021*
www.researchmathsci.org
*DOI: http://dx.doi.org/10.22457/apam.v23n1a04801*


# Linear Regression Model Using Bayesian Approach for Iraqi Unemployment Rate

***Ali Hussein Ali*[1*] and *Tasnim H. Kadhim*[2]**

[1,2]Department of Mathematics, Collage of Science, University of Baghdad
Baghdad, Iraq. [2]email: tasnim.h@sc.uobaghdad.edu.iq
[*]Corresponding author. [1]email: ahussaein287@gmail.com

*Received 13 November 2020; accepted 10 January 2021*

***Abstract.*** In this paper we used frequentist and Bayesian approaches for the linear regression model to predict future observations for unemployment rates in Iraq. Parameters are estimated using the ordinary least squares method and for the Bayesian approach using the Markov Chain Monte Carlo (MCMC) method. Calculations are done using the R program. The analysis showed that the linear regression model using the Bayesian approach is better and can be used as an alternative to the frequentist approach. Two criteria, the root mean square error (RMSE) and the median absolute deviation (MAD) were used to compare the performance of the estimates. The results obtained showed that the unemployment rates will continue to increase in the next two decades.

***Keywords:*** Linear regression, frequentist approach, Bayesian approach

***AMS Mathematics Subject Classification (2010):*** 62J02

## 1. Introduction

One of the statistical methods most used in scientific research is regression analysis. Regression analysis is applied in many scientific researches, for example medicine, economics, sociology, etc. Regression analysis is a very useful technique to answer the research question, is there a causal relationship between the variables. There are two main purposes of using regression analysis: First knowing the relationship between the response variable and independent variables. Second predict the value of the dependent variable based on the values of the independent variables. A regression model is formulated first and then estimate of the parameters of the regression model is made. There are many methods for estimating model parameters. The most common method used in research is the Ordinary Least Squares (OLS) [1,2]. The OLS method minimizes the sum of squared deviations from the estimated regression equation. Another method used to estimate the parameters of the regression model is the Bayesian approach. The difference between the frequnteist and Bayesian methods is a point of view on the parameters. In Bayesian approach, the parameters are seen as random variables that have more than one value; while in the frequnteist view the parameters have only one value (parameter is a constant). Bayesian approach has recently become an increasingly interesting and applicable statistical method in many areas of application such as social science, medicine, finance,

machine learning and economy, many studies have been carried out in applying the computational Bayesian approach in linear regression modeling. Lynch introduces basic theoretical and applied principles of Bayesian approach in many different the social sciences and illustrate incorporate information from prior research, and its ability to update estimates as new data are observed [3]. James and Scott illustrated the computational methods of the Bayesian approach and the variable selection in a linear model. This is done with the help of Markov chain Monte Carlo [5]. Tino Berger and Everaert explained unemployment rates in the United States and the non-accelerating inflation rate of unemployment (NAIRU) in a Bayesian framework. They developed a Bayesian model of unemployment based on supply and demand factors. These two factors influence unemployment rates. The model allows unemployment, which changes over time, to continue [6]. Scott and Steven gave explanation of Bayesian approach as the real options approach to assessing irreversible investment opportunities has become part of the dominant literature in financial economics because unemployment rates have costs for society more than just financial costs. Bayesian approach provides a natural framework for addressing central issues in finance. The results show an increase in European unemployment rates driven by structural factors [7]. Grzenda illustrated the benefits of scientific research using the Bayesian approach. Predictive modeling of unemployment rates was determined and represented by linear regression. They compared the effect of prior information and non-informational information on the model's accuracy. The results show that the accuracy of the models estimated at Informational distributions a prior are higher. Therefore, it becomes more accurate for regression models when additional knowledge are available about search problem [4].

The main objectives of this research is to apply each of the frequentist and Bayesian approaches to build a linear regression model for the Iraqi unemployment rates which are affected by three factors: gross domestic production, oil prices, and population growth. Then to predict the unemployment rates in Iraq for the next two decades.

## 2. Linear regression model

Multiple linear regression models are used to evaluate the relationship between dependent variables with more than one independent variable [5]. The multiple linear regression model contains one dependent variable Y and k independent variables $X_1$, $X_2$, …,$X_k$. The multiple linear regression models is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \quad (1)$$

Or

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\varepsilon} \quad (2)$$

where

**Y** is the vector of the dependent variable
**X** is the matrix of independent variables
**β** is vector of regression model parameters
**ε** is vector of errors

So the multiple linear regression model can be written as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{bmatrix} 1 & x_{11} \cdots & x_{1k} \\ . & & \\ . & x_{21} & x_{2k} \\ . & \vdots & \vdots \\ 1 & x_{n1} \cdots & x_{nk} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

The OLS formula for estimating the parameters of the model is:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad (3)$$

**3. Bayesian linear regression**

Bayesian linear regression is one in which the estimation of parameters is made using Bayesian approach. The Bayesian regression provides updates of our beliefs. The main idea of using the Bayesian approach is when the regression model contains variables that need to be estimated and these variables are undergoing changes over time, and need updating data or the model contains many independent variables. This feature is present in the Bayesian approach. The Bayesian approach uses prior information and the observed data. Parameters are estimated using Markov chain Monte Carlo (MCMC) method. Bayesian models are these days used to describe real life problems observed under uncertainty. A Bayesian model is a collection of probabilistic statements which describe and interpret present or predict future interpretation. Model provides a theoretical framework for better understanding of phenomena of interest.The response in Bayesian regression model sampled from a normal distribution [8]. In normal linear regression model there are n independent observations $y_1$,$y_2$,.. $y_n$. Each observation $y_i$ has its own mean $\mu_i$ and all observations have the same variance $\sigma^2$. Linear function of the predictor variables $x_1$,$x_2$...,$x_k$ are the unknown means. The Bayesian approach contains a prior distribution, likelihood distribution and posterior distribution. The variables $(Y|X,\beta,\sigma^2)$ are normally distributed. Thus the variables $(Y|X,\beta,\sigma^2) \sim N(X\beta,\sigma^2)$ and the likelihood of these variables are as follows:

$$P(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta},\sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{\beta X})^T(\boldsymbol{Y}-\boldsymbol{\beta X})\right] \quad (4)$$

There are several prior distributions that can be used in Bayesian approach to a linear regression model; one of them is the conjugate prior distribution. A prior is called conjugate if prior and the posterior density function belongs to the same distribution family. The conjugate prior is usually used because this makes the mathematical computation of the likelihood function as simple in model. Using the normal distribution for the parameters and an inverse gamma distribution for the prior distribution [8]:

$\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\rho},\boldsymbol{\gamma})$ and $\sigma^2 \sim IG(\varphi,\omega)$

The prior for **β** is a multivariate normal distribution with mean $\rho$ and covariance$\gamma$. Prior for $\sigma^2$ is an inverse gamma distribution with shape $\varphi$ and scale $\omega$.In other words we can abbreviate the prior for $(\boldsymbol{\beta},\sigma^2)$ by normal inverse gamma $(\boldsymbol{\rho},\boldsymbol{\gamma},\varphi,\omega)$.

**3.1. Measuring performance quality**

Two measures of the quality of performance of estimates were used:

1. Root mean square error (RMSE)

It is the root mean of squares difference between the observed and fitted values and is given by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (5)$$

2. Mean Absolute Deviation

It is the average distance between the observed and fitted values and is given by

$$MAD = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (6)$$

**4. Methodology**

This research uses data which obtained from Republic of Iraq Ministry of Planning [9]. In our study, there are three independent variables that affect unemployment rates: gross domestic production (GDP), oil prices, and population growth. Data is processed using linear regression model with frequentist method using the OLS. The data is also modeled by Bayesian approach. Computations of the OLS and the Bayesian approach uing MCMC method are both done by the R package.

**5. Results and discussions**

Table 1 is the output of OLS estimates of the linear regression coefficients. From Table 1, the population growth, $\beta_3$ gave a positive effect which can be explained: if the population growth rates increase, unemployment rates will also increase. Otherwise, the other two variables give negative effect; the GDP variable and the oil prices variable. If these variables increase, then unemployment rates will decrease. Based on the p value the less important variables are rejected. The variable is rejected when P - value> 0.1, element is the least important and must be removed.

**Table 1:** Estimated linear regression coefficients using OLS method

| Estimate | T value | P-value |
|---|---|---|
| $\beta_0 = 3.42273$ | 3.267 | 0.004845 |
| $\beta_1 =$ -0.00783 | -0.659 | 0.00519 |
| $\beta_2 =$ -0.03325 | - 4.439 | 0.000413 |
| $\beta_3 = 0.26780$ | 7.621 | 1.03e-06 |

The results indicate that all variables affect the unemployment rates are significant. So the regression model becomes as follows:

$$y = 3.42273 - 0.00783X_1 - 0.03325X_2 + 0.26780X_3 \quad (7)$$

To estimate parameters using Bayesian approach as a direct relation, we have a normal prior distribution for β and inverse gamma distribution for the parameter $\sigma^2$. The estimated regression coefficients $\beta_0$ to $\beta_3$ using Bayesian approach are given in Table 2.

Table 2. Estimated regression coefficients using Bayesian approach

| Estimates | *Quantile (2.5%)* | Quantile (97.5%) |
|---|---|---|
| $\beta_0 =$ 3.54382 | -0.0413 | 0.6852 |

Linear Regression Model Using Bayesian Approach for Iraqi Unemployment Rate

| | | |
|---|---|---|
| $\beta_1 = -0.00794$ | -0.0137 | *0.5962* |
| $\beta_2 = -0.04253$ | 0.01436 | 1.45e-05 |
| $\beta_3 = 0.34681$ | -0.0038 | 0.2875 |

The model of linear regression with Bayesian approach given as:

$$y = 3.54382 - 0.00794X_1 - 0.04253\,X_2 + 0.34681X_3 \quad (8)$$

After we found the parameter values in two methods, Table 3 shows the comparison using two criteria. The smaller value of RMSE and MAD indicate a better model. The two criteria used in this research showed that the values of the linear regression model with a Bayesian approach are smaller than the value from the OLS method. So in this case it is better to use linear regression model with Bayesian approach.

**Table 3:** Comparison between OLS and Bayesian approach in linear regression model

| **Estimation Method** | ***RMSE*** | **MAD** |
|---|---|---|
| OLS | 0.695121 | 0.81509 |
| Bayesian approach | 0.684385 | 0.79956 |

Table 4 shows the forecasted unemployment rates in Iraq for the next two decades using the frequentist and Bayesian approaches.

**Table 4:** Forecasted unemployment rates in Iraq

| Year | Unemployment Rate using Frequentist Approach | Unemployment Rate using Bayesian Approach |
|---|---|---|
| 2020 | 8.81% | 8% |
| 2021 | 9.22% | 8.28% |
| 2022 | 8.73% | 8.95% |
| 2023 | 9.26% | 8.55% |
| 2024 | 8.41% | 8.51% |
| 2025 | 8.40% | 8.20% |
| 2026 | 8.65% | 8.22% |
| 2027 | 7.82% | 9.30% |
| 2028 | 9.67% | 11.25% |
| 2029 | 8.76% | 11.81% |
| 2030 | 8.57% | 10.43% |
| 2031 | 8.57% | 12.43% |
| 2032 | 9.03% | 12.38% |
| 2033 | 9.17% | 12.42% |
| 2035 | 11.01% | 11.02% |
| 2036 | 11.84% | 11.65% |

| | | |
|---|---|---|
| 2037 | 11.47% | 11.51% |
| 2038 | 11.68% | 12% |
| 2039 | 12.20% | 12.30% |

**6. Conclusions**

We studied the factors affecting the increase in unemployment rates in Iraq, and we found that all the independent variables are significant. We used two criteria for non-aligned in the Bayesian versus the frequentist debate to find out the best approach to use. The results showed that the regression model with Bayesian approach is better than the regression model with classical approach by using the RMSE and MAD criteria. Therefore, the Bayesian approach directly constitutes the best decision-making process

**REFERENCES**

1. F.Liang, C.Liu and R.J.Carroll, *Advanced Markov Chain Monte Carlo Methods Learning from Past Samples*, John Wiley and Sons Ltd. (2010).
2. G.Casella and E.George, *Explaining the Gibbs Sampler*, The American Statistician (1992).
3. S.M.Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists* (2007).
4. S.Weisberg, *Applied Linear Regression*, 2nd Ed., J. Wiley & Sons, Inc., New York (1985).
5. M.J.Heaton James and G. Scott, *Bayesian Computation and the Linear Model* (2009).
6. T.Berger and G.Everaert, Unemployment persistence and the NAIRU: A Bayesian approach, *Scottish Journal of Political Economy*, 55(3) (2008) 281-299.
7. W.Grzenda, Informative Versus non- Informative prior distributions and their impact on the accuracy of Bayesian approach. Statistics in transition new series, December 2016.
8. W.M.Bolstad, *Understanding Computational Bayesian Statistics*, John Wiley & Sons. (2010).
9. https://www.mop.gov.iq/