FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, (Special Issue) 2024, pp 19 - 33
DOI: https://doi.org/10.33003/fjs-2024-0803-1988

A BAYESIAN SIMULATION APPROACH TO MODELING THE RELATIONSHIP BETWEEN NARCOTIC DRUG USE PREVALENCE AND UNEMPLOYMENT RATE USING AGGREGATE DATA

***[1]Udoumoh, E. F., [1]Emaikwu P. O., [1]Agada, P. O. and [2]Subeno, T.**

[1]Department of Statistics, Federal University of Agriculture, Makurdi, Nigeria
[2]Delta State Polytechnic Ogwashi-Uku, Delta State, Nigeria

*Corresponding authors' email: uenobong@gmail.com
ORCID: https://orcid.org/0000000171988819

ABSTRACT

Fitting a Binary Logistic Model relating narcotic drug use prevalence and unemployment rate can be a challenge in the face of aggregate data. This aggregation limits the use of the Classical Binary Logistic Regression Model. This limitation informed our development of a Bayesian Statistical Simulation Modeling Procedure. The modeling procedure embeds the Markov Chain Monte Carlo (MCMC) algorithm which is implemented on an Open Source Software Platform-Windows; Bayesian Inference Using the Gibbs Sampler (WINBUG). Part of the modeling activity is the mathematical analysis on the response of the success probability of narcotic drug use prevalence to changes in unemployment rate. This was done under conditions of positive and negative values of the regression parameters (constant and covariate coefficient). The extent to which unemployment rate is a risk factor of the success probability was investigated and the zones within Nigeria where unemployment rate is a risk factor of narcotic drug use prevalence where also identified. Results revealed that unemployment rate is a risk factor to narcotic drug use prevalence in four (4) zones (North-Central, South-East, South-West and North-East) with risk levels of 53.65%, 51.59%, 49.42% and 46.02% respectively. While, factors latent to study impact negatively on five (5) zones (North-East, North-West, North-Central, South-East and South-West). It is recommended that attention should be drawn to the South-South zone where unemployment rate is not a risk factor to narcotic drug use prevalence, but other factors latent to the study are impacting positively on it.

Keywords: Aggregate Data, Narcotic Drug, MCMC, Bayesian Inference, WINBUG

INTRODUCTION

The term "narcotic" comes from the Greek word "stupor" and originally referred to a variety of substance that dulled the senses and relieves pain (DEA, 2020). Though people still refer to all drug as "narcotics", today "narcotic" refers to opium, opium derivatives and their semi-synthetic substitutes. The basic therapeutic use of narcotic is for pain relief and hence they are often called narcotic analgesic. However, this substance has suffered so much abuse in recent times owing to some identified causes which include unemployment. Hayes (2022) defined unemployment as a situation when a person who is actively searching for employment is unable to find work. Unemployment is considered to be a key measure of the health of the economy. The most frequent measure of unemployment is the unemployment rate, which is the number of unemployed people divided by the number of people in the labor force. Anderson (2022) also defined unemployment rate as the percent of the labor force that is jobless. He referred to unemployment rate as a lagging indicator which generally rises or falls in the wake of changing economic conditions. Rastegari *et al.* (2013) conducted a comparison between classification and regression trees and logistic regression analysis to determine the factors influencing drug injection history among prisoners. a sample of 2720 Iranian prisoners was studied to determine the factors influencing drug injection. The collected data was divided into two groups of training and testing. A logistic regression model and a classification Regression Tree (CARTs) were applied on training data. The result of their finding shows that the regression model and the CARTs had 8 and 4 significant variables, respectively. The overall result shows that, heroin use, history of imprisonment, age at first drug use, and marital status were important factors in determining the history of drug injection.

Substance abuse and unemployment are said to be associated with each other. The National Survey on Drug Use and Health from 2012 found that 17 percent of unemployed people had a substance abuse disorder, compared to 9 percent of full-time workers. The question is, does unemployment eventually lead to substance abuse? Although conclusive evidence is yet to be discovered, research suggests that it's actually the state of unemployment that prompts substance abuse. Pourallahvirdi*et al.* (2016) conducted a prospective study to ascertain the major cause of drug abuse from the viewpoint of addicted persons. In this study the researchers mainly focused on the major causes of drug addiction in adults. Their findings showed that failures in life and escape from problems were the major causes of drug addiction. Amiri *et al.* (2016) conducted a quantitative content analysis to ascertain factors affecting tendency for drug abuse in people attending addiction treatment centre's. Their study outlined the most common causes of tendency for drug use as; addicted friends, unemployment, lack of attention to religious tendencies, economic problems, marital discords, lack of recreational facilities, availability of drugs and failures to say no to others' demand were the major causes of drug abuse. Obot (2019) conducted a prospective study to explore the availability use, consequence and policy implications of drug in the Nigerian population. The result showed that illicit drug use among adults is 14.4 percent, which is significantly higher than the global prevalence of 5.5 percent; their finding further identified cannabis as the most common drug used in Nigeria, followed by opioids, cough syrups containing codeine and tranquillizers and sedatives. Jatau *et al.* (2021) conducted a review of epidemiological studies and drug laws to ascertain the burden of drug abuse in Nigeria. Their study did a systematic search on 253 articles. The result revealed that commonly abuse drugs include cannabis, cocaine, amphetamine, heroin, diazepam, codeine, cough syrup and tramadol. Other studies on narcotic drug use and related issues include the works by Azagba *et al*. (2021), Mohammed (2018), and Adebimpe and Folashade (2018).

Aggregate data refers to data on individuals that have been averaged by year, by geographical area, by service agency or in some other way (Jacob, 2016). Holderness (2016) studied the problems involved in using of aggregate data to infer individual behavior as evidence from law, finance and ownership concentration. His research uses cross-country comparisons in finance using two distinct approaches. The result shows that when one uses data from individual firms, the coefficient on each of three key legal measures either changes sign or loses statistical significance based on similar specifications and the same underlying ownership data. The data for this work is termed aggregate because, the distribution of individual cases of the number of narcotic drug users and their corresponding number of unemployed persons over time are not available. What was available is the aggregate data on the number of narcotic drug users at a given time and the corresponding total number of unemployed person at that time. The incompleteness of this data has limited the use of Classical Logistic Regression Model. In an attempt to address this challenge of incomplete data, researchers have adopted methods like the expected maximum (EM) algorithm and errors-in-variable (EIV) regression approach (McLachlan, *et al*., 2004). Elsewhere, Choi, *et al.* (2008) demonstrated the use of a Bayesian statistical modeling procedure via the Markov Chain Monte Carlo (MCMC) algorithm to estimate the Binary Logistic Regression model with incomplete data, where missing individual responses were treated as additional parameters to be estimated. Further on the list of works on Bayesian Simulation approach include Agada *et al.* (2019) on a Bayesian Simulation Modeling Approach to Predicting Maternal Age-Specific Infants Survival Outcomes with Incomplete Data and Agada *et al*. (2021) on A Bayesian Simulation Approach to Modeling the Relationship between Rice Mill Loss Rate and Moisture Content with Incomplete Data.

In this paper, the relationship between narcotic drug use prevalence and unemployment rate using aggregate data is considered. The Binary Logistic Model relating zonal narcotic drug use prevalence and unemployment rate is presented using the Bayesian simulation approach. The essence is to determine whether a zone's unemployment rate is a risk factor to its narcotic drug use prevalence and also to assess the level of risk in other to determine the impact (positive/negative) of other factors latent to the study on the zone's narcotic drug use prevalence.

MATERIALS AND METHODS

In this section, we present the method of data collection, Binary Logistic Model relating narcotic drug use prevalence and unemployment rate. We centred the zonal unemployment rate at the over-all zonal mean, in order to determine the impact of other factors latent to the study. The effects of the zonal unemployment rate on the zonal narcotic drug use prevalence rate were determined. Furthermore, the Bayesian Binary Logistic Simulation Model relating Narcotic drug use prevalence and unemployment rate as well as the Bayesian Statistical Simulation Modeling procedure were presented. Finally, Model convergence diagnostic checks were also presented in other to determine the accuracy of the model.

Source of Data

The data sets for this work are secondary data which consist of the total number of narcotic drug users and the number of unemployed persons in each of the six (6) geo-political zones in Nigeria as at 2018. The geo-political zones are North-East, North-West, North-Central, South-East, South-West and South-South. The data on the number of narcotic drug users was sourced from the National Survey on Drug Use and Health 2018, (conducted by the Nigerian National Bureau of Statistics (NBS)) and the Centre for Research and Information on Substance Abuse (CRISA). This body receives technical support from the United Nation Office on Drugs and Crime (UNODC). The data on the number of unemployed persons were sourced from the year 2019 publication of the Nigerian National Bureau of Statistics on unemployment and underemployment by states

Table 1: Data of actual Narcotic Drugs use and Unemployed[1]

Zones	No. of Narcotic Drug User	No. of Unemployed
North-East	2113000	3715711
North-West	3052000	7226157
North-Central	1563000	5190419
South-East	1591000	4572754
South-West	4424000	5626168
South-South	2171000	7695908

(Source: National Bureau of Statistics Bulletin, 2018)

The Binary Logistic Model Relating Zonal Narcotic Drug use Prevalence and Unemployment Rate

We state the model as;

$$Logit\ (p_i) = \alpha_i + \beta_i(x_i - \bar{x}) \qquad (1)$$

where,

$Logit\ (p_i) = \log(\frac{p_i}{1-p_i})$,

$\alpha_i\ and\ \beta_i$ are the logistic regression parameters, p_i(the success rate) is the Zonal narcotic drug use prevalence, the covariate x_i is the Zonal unemployment rate and $i = 1, 2, \dots, 6$ index the number of zones in a geo-political zone in Nigeria.

From Equation (1) we have that;

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha_i + \beta_i(x_i - \bar{x})$$

$$\frac{p}{1-p} = e^{\alpha_i + \beta_i(x_i - \bar{x})}$$

$$\frac{1-p_i}{p_i} = \frac{1}{e^{\alpha_i + \beta_i(x_i - \bar{x})}}$$

$$\frac{1-p_i}{p_i} = e^{-[\alpha_i + \beta_i(x_i - \bar{x})]}$$

$$1 - p_i = p_i e^{-[\alpha_i + \beta_i(x_i - \bar{x})]}$$

$$1 = p_i + p_i e^{-[\alpha_i + \beta_i(x_i - \bar{x})]}$$

$$1 = p_i(1 + e^{-[\alpha_i + \beta_i(x_i - \bar{x})]})$$

Therefore,

$$p_i = \frac{1}{1 + e^{-[\alpha_i + \beta_i(x_i - \bar{x})]}} \qquad (2)$$

Centring the Zonal Unemployment Rate (x_i) at the Over-all Zonal Mean ($\bar{x}$)

We centred the zonal unemployment rate (x_i) at the zonal mean ($\bar{x}$). This is because none centring means zero unemployment rate which is not realistic in practice, but centring at the over-all zonal mean unemployment rate will help determine the impact of other factors latent to the study on the zone's narcotic drug use prevalence.

Centring implies $x_i = \bar{x}$

When this holds, Equation (2) becomes,

$$p_i = \frac{1}{1+e^{-\alpha_i}} \quad (3)$$

If $\alpha_i < 0$ then;

$$p_i = \frac{1}{1+e^{\alpha_i}} \quad (4)$$

This has a negative impact on p_i

If $\alpha_i > 0$ then;

$$p_i = \frac{1}{1+e^{-\alpha_i}} \quad (5)$$

This has a positive impact on p_i

Effect of the Zonal Unemployment Rate, x_i on Zonal Narcotic Drug Use Prevalence Rate, p_i

We establish mathematically, the effect of the Zonal unemployment rate, x_i on the

Zonal narcotic drug use prevalence rate, p_i of the i[th] zone. Recall Equation (2);

$$p_i = \frac{1}{1+e^{-[\alpha_i+\beta_i(x_i-\bar{x})]}}$$

Observe that, if $\beta_i < 0$ then;

$p_i \to 0$ as$x_i \to \infty$ This shows that increased and decreased levels of x_i has the effect decreasing and increasing p_i respectively. Hence, we term the Zonal unemployment rate (x_i) a Non-risk factor of the Zonal narcotic drug use prevalence rate (p_i).

Observe also from Equation (2) that if $\beta_i > 0$ then;

$p_i \to 1$ as$x_i \to \infty$. This shows that increased and decreased levels of x_i has the effect increasing and decreasing p_i respectively. Hence, we term the Zonal unemployment rate (x_i) a Risk factor of the Zonal narcotic drug use prevalence rate (p_i)

Level of Risk Factor

We will model the level of the risk factor (L_{Risk}) as:

$$L_{Risk} = P(B_i > 0) * 100\% \quad (6)$$

Where $P(\beta_i > 0)$ is the probability of having positive value of β_i. That is, the proportion of times (%) that reduced levels of Zonal unemployment rate of a zone, will have negative effect on the prevalence of narcotic drug use. $P(B_i > 0)$ also helps to determine the significance of the β_i in the model. $P(B_i > 0) \geq 0.5$ indicates that β_i stays in the model.

Bayesian Theorem

Let s be a sample space and let A and B be two event in S. We denote the probability of A by P(A) and B by P(B). Where $P(A) \neq 0$ and $P(B) \neq 0$.

The probability of an event A occurring given B is defined by $P(A/B)$. The conditional distribution formula is defined as

$$P(A/B) = \frac{P(A\cap B)}{P(B)} \quad (7)$$

And

$$P(B/A) = \frac{P(A\cap B)}{P(A)} \quad (8)$$

are helpful in deriving Bayes' rule. Substituting P(A∩B)=P(B/A)P(A) gives us

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad (9)$$

we refer to Equation (8) as Bayes' rule. In the case where A is a set of j mutually exclusive event, A_j we use the law of total probability in the discrete case to calculate $P(B)$:

$$P(B) = \sum_j P(B/A_j)P(A_j) \quad (10)$$

Given that event B has occurred, the posterior probability of any of these j events occurring is

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_j P(B/A_j)P(A_j)} \quad (11)$$

Where $P(A/B)$ is the posterior probability P(A) is the prior probability, P(B/A) the likelihood and P(B) the marginal likelihood. Thus, the Bayes' rule can be written as

$$Posterior = \frac{likelihood*Prior}{Marginal\ likelihood} \quad (12)$$

To find the posterior probability density, we use the parameter estimation form of Bayes' rule

$$posterior \propto prior * likelihood \quad (13)$$

Bayesian Prior, Likelihood and Posterior

We define $Y_i, i = 1 \ldots n$ to be a random variable associated with a sample of size n, $L(y_1, y_2, \ldots, y_n|\theta)$ the likelihood parameter θ is viewed to be a random variable with a probability distribution $g(\theta)$ called the prior distribution of θ. The prior was criticized for introducing subjective information which was purely an educated guess and can vary from one scientist to another. The prior distribution is specified before any data are collected and provides a theoretical description of information about θ that was available before any data were obtained. We will assume that the parameter θ has a continuous distribution with density $g(\theta)$. Using the likelihood of the data and the prior on θ, it follows that the joint likelihood of

$$f(y_1 y_2 \ldots y_n|\theta) = L(y_1 y_2 \ldots y_n|\theta)g(\theta) \quad (14)$$

With marginal density or mass function of $Y_1, Y_2, \ldots, Y_n$ as

$$m(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(y_1 y_2 \ldots y_n|\theta)g(\theta)d\theta \quad (15)$$

Have the posterior density of $\theta|y_1, y_2, \ldots, y_n$:

$$g^*(\theta|y_1 y_2 \ldots y_n) = \frac{L(y_1 y_2 \ldots y_n|\theta)g(\theta)}{\int_{-\infty}^{\infty} L(y_1 y_2 \ldots y_n|\theta)g(\theta)d\theta} \quad (16)$$

The posterior density summarizes all of the important information about the parameter θ by making use of the information contained in the prior for θ and the information in the data.

The Bayesian Binary Logistic Simulation Model relating Zonal Narcotic Drug Use Prevalence and Unemployment Rate

We present the Bayesian Statistical Modelling Procedure for modeling the relationship between the zonal narcotic drug use prevalence and unemployment rate. The modeling procedure embeds the Markov Chain Monte Carlo (MCMC) algorithm implemented on an Open Source Software Platform-Windows Bayesian Inferences Using the Gibbs Sampler (WINBUG).

The bayesian statistical simulation modelling procedure

Given two faces of the coin; the narcotic drug use prevalence (p_i) for a zone i and the non- narcotic drug use prevalence ($1 - p_i$), we proposed the Binomial Likelihood such that;

$y_i/p_i \sim$ Binomial (p_i, n)

Where, y_i is the number of narcotic drug users in zone$i$ while the success rate $p_i = {y_i}/{n_z}$ is the zonal prevalence of narcotic drug use in zone$i$. n_z is the number of narcotic drug users in a zone. We state that the computation of p_i per $n(= 1000)$ persons was done in order to determine the observed values of y_i per 1000 persons and for computational ease.

Logistically, p_i is the transformation of the regression mean,

$\alpha_i + \beta_i(x_i - \bar{x})$ and we state that;
$Logit(p_i) = \alpha_i + \beta_i(x_i - \bar{x})$.
We propose that the regression parameters $\alpha_i and \beta_i$ have the priors;
$\alpha_i \sim Normal(0, 0.01)$, $\beta_i \sim Normal(0, 0.01)$.
Using the parameter version of the Bayes Theorem (Scott, 2007), the posterior distribution of the model parameter α_i and β_i relating their respective prior densities $f(\alpha_i)$ and $f(\beta_i)$ and their data likelihoods $f(y_i \backslash \alpha_i)$ and $f(y_i \backslash \beta_i)$ are

$$f(\alpha_i \backslash y_i) = \frac{f(\alpha_i) f(y_i \backslash \alpha_i)}{\int f(\alpha_i) f(y_i \backslash \alpha_i)\, d\alpha_i} \quad (17)$$

$$f(\beta_i \backslash y_i) = \frac{f(\beta_i) f(y_i \backslash \beta_i)}{\int f(\beta_i) f(y_i \backslash \beta_i)\, d\beta_i} \quad (18)$$

Algorithm for the Bayesian logistic regression model fitting in WinBUG

In WINBUG syntax, we fit the Bayesian Logistic Regression Model with centred covariate as follows;
Model {
For (i in 1: k) {
$y[i] \sim dbin(p[i], n)$
$logit(p[i] < -alpha[i] + beta[i] * (x[i] - mean(x))$
$prob[i] < -step(beta[i] - 0.5)$
$alpha[i] \sim dnormal(0, 0.01)$
$beta[i] \sim dnormal(0, 0.01)$ } }
where the data list of number of narcotic drug users $y[0.142, 0.205, \ldots, 0.146]$and unemployment rate (in percentage) $x[10.920, 21.237, \ldots, 22.617]$ as well as the initialization list for the model parameter arrays; $alpha$ [ ] $and\ beta$[ ] are the prior for narcotic drug use prevalence and unemployment rate respectively and distributed normally with mean Zero(0) and Standard deviation 0.1, defined for each zone i in the nation. k is set as the number of zones in the nation while n is set at 1000. The simulation will be run for 300,000 burn-in periods after which samples of 300,000 iterations were collected. A thinning of 32 was maintained throughout the simulations and the overlay check box in WINBUG was checked to reduce autocorrelation. Other modeling requirements are as stated by the WINBUG Software documentation.
Recall that WINBUG uses Equation (2) in computing the Zonal narcotic drug use prevalence (p_i) for a zone i within the nation Nigeria. This gives the simulated value of p_i. We state here that posterior median values of α_i and β_i are used in this work when the standard deviation is high. Otherwise, the mean values are used.

Model convergence diagnostic check

Model convergence diagnostic was done using history plots, density plots and autocorrelation plots. The plots will be produced when the model parameters and measures are monitored on WinBUG. Our approach for investigating convergence issues is by inspecting the mixing and time trends within the chains of individual parameters. The history plots are the most accessible convergence diagnostic and are easy to inspect visually. The history plot of a parameter, plot the simulated values for the parameter against the iteration number. The history plot of a well-mixed parameter should traverse the posterior domain rapidly and should have nearly constant mean and variance. The density plots of the model parameters were checked against their actual probability distribution to see whether the right distribution is simulated. This was done for the alpha and beta distribution for each zone i.
Samples simulated using Markov Chain Monte Carlo (MCMC) methods are correlated. The smaller the correlation, the more efficient the sampling process. Though, the Gibbs, MCMC algorithm will typically generates less-correlated draws, there will be a need to monitor the autocorrelation of each parameter to ensure samples are independent. The autocorrelation plot that will come from a well mixing chain becomes negligible fairly quickly after a few lags. This was achieved for the model parameters and measures.

RESULTS AND DISCUSSION

The result of this study include the actual values of the narcotic drug use prevalence and unemployment rates in each of the six (6) geo-political zones in the Nation. The results also include the model parameter and measure values for each geopolitical zone. Results on model diagnostic checks for the six zones which include history plots, density and auto-correlation plots were presented in figure 8. In addition, the study results also include a distribution of unemployment rate factor status and a distribution of unemployment rate risk level across zones.

Table 2: Distribution of actual Narcotic Drug Use Prevalence and Unemployment Rate

Zones	No. of Narcotic Drug User	No. of Unemployed	Narcotic Drug Use Prevalence	Unemployment Rate (%)
North-East	2113000	3715711	0.142	10.920
North-West	3052000	7226157	0.205	21.237
North-Central	1563000	5190419	0.105	15.254
South-East	1591000	4572754	0.107	13.439
South-West	4424000	5626168	0.297	16.534
South-South	2171000	7695908	0.146	22.617

Source**:** National Survey on Drug Use and Health 2018, (conducted by the Nigerian National Bureau of Statistics (NBS)) and the Centre for Research and Information on Substance Abuse (CRISA). The data on the number of unemployed persons were sourced from the year 2019 publication of the Nigerian National Bureau of Statistics on unemployment and underemployment by states.

Table 3: Distribution of Model Measures across Zones in the Nation

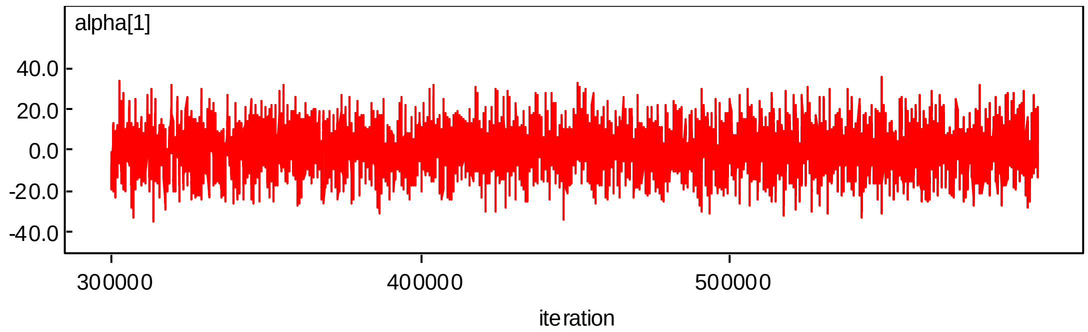
Zones	Sign of Model parameter(α)	Zonal Impact on Zonal Narcotic Drug Use Prevalence	Sign of Model Parameter(β)	Unemployment Rate Factor Status	Unemployment Rate Level of Risk (%)
North-East	-	Negative	+	Risk-Factor	46.02
North-West	-	Negative	-	Non-risk Factor	-
North-Central	-	Negative	+	Risk-Factor	53.65
South-East	-	Negative	+	Risk-Factor	51.59
South-West	-	Negative	+	Risk-Factor	49.42
South-South	+	Positive	-	Non-risk Factor	-

The signs (+/-) are that of the median values of parameter (α) and parameter(β).

Table 4: Model parameters and measure values for each zone within the Nation

Node	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
alpha[1]	-0.04397	10.06	0.2101	-19.91	0.1119	19.65	300000	9375
alpha[2]	-0.1252	9.725	0.2641	-19.07	-0.2419	19.06	300000	9375
alpha[3]	-0.6631	7.992	0.1312	-16.28	-0.5874	14.9	300000	9375
alpha[4]	-0.1525	9.486	0.1613	-18.75	-0.09899	18.24	300000	9375
alpha[5]	-0.8375	1.328	0.01316	-3.421	-0.818	1.755	300000	9375
alpha[6]	0.5364	9.584	0.2029	-18.46	0.551	19.45	300000	9375
beta[1]	0.3059	1.751	0.03657	-3.153	0.3336	3.734	300000	9375
beta[2]	-0.2697	2.128	0.05781	-4.462	-0.2422	3.874	300000	9375
beta[3]	1.05	5.656	0.09288	-10.02	1.099	12.05	300000	9375
beta[4]	0.6112	2.938	0.04992	-5.158	0.6235	6.321	300000	9375
beta[5]	0.1858	10.05	0.1	-19.49	0.3896	19.76	300000	9375
beta[6]	-0.3877	1.611	0.03411	-3.56	-0.3906	2.801	300000	9375
p[1]	0.142	0.0111	1.233E-4	0.1213	0.1418	0.1643	300000	9375
p[2]	0.2049	0.01278	1.461E-4	0.1805	0.2046	0.2301	300000	9375
p[3]	0.1051	0.009787	1.054E-4	0.08672	0.1048	0.125	300000	9375
p[4]	0.107	0.009768	9.398E-5	0.08895	0.1066	0.1272	300000	9375
p[5]	0.2971	0.01442	1.615E-4	0.2694	0.2969	0.3255	300000	9375
p[6]	0.1459	0.01121	1.23E-4	0.1246	0.1457	0.1689	300000	9375
prob[1]	0.4602	0.4984	0.008913	0.0	0.0	1.0	300000	9375
prob[2]	0.3657	0.4816	0.01125	0.0	0.0	1.0	300000	9375
prob[3]	0.5365	0.4987	0.006963	0.0	1.0	1.0	300000	9375
prob[4]	0.5159	0.4997	0.007833	0.0	1.0	1.0	300000	9375
prob[5]	0.4942	0.5	0.004362	0.0	0.0	1.0	300000	9375
prob[6]	0.2878	0.4527	0.008744	0.0	0.0	1.0	300000	9375

Note: (i) Table 3 indicate the zones within the Nation
(ii) The model parameters are alpha and beta
(iii) The model measures are narcotic prevalence (p) and risk level of unemployment rate (prob)



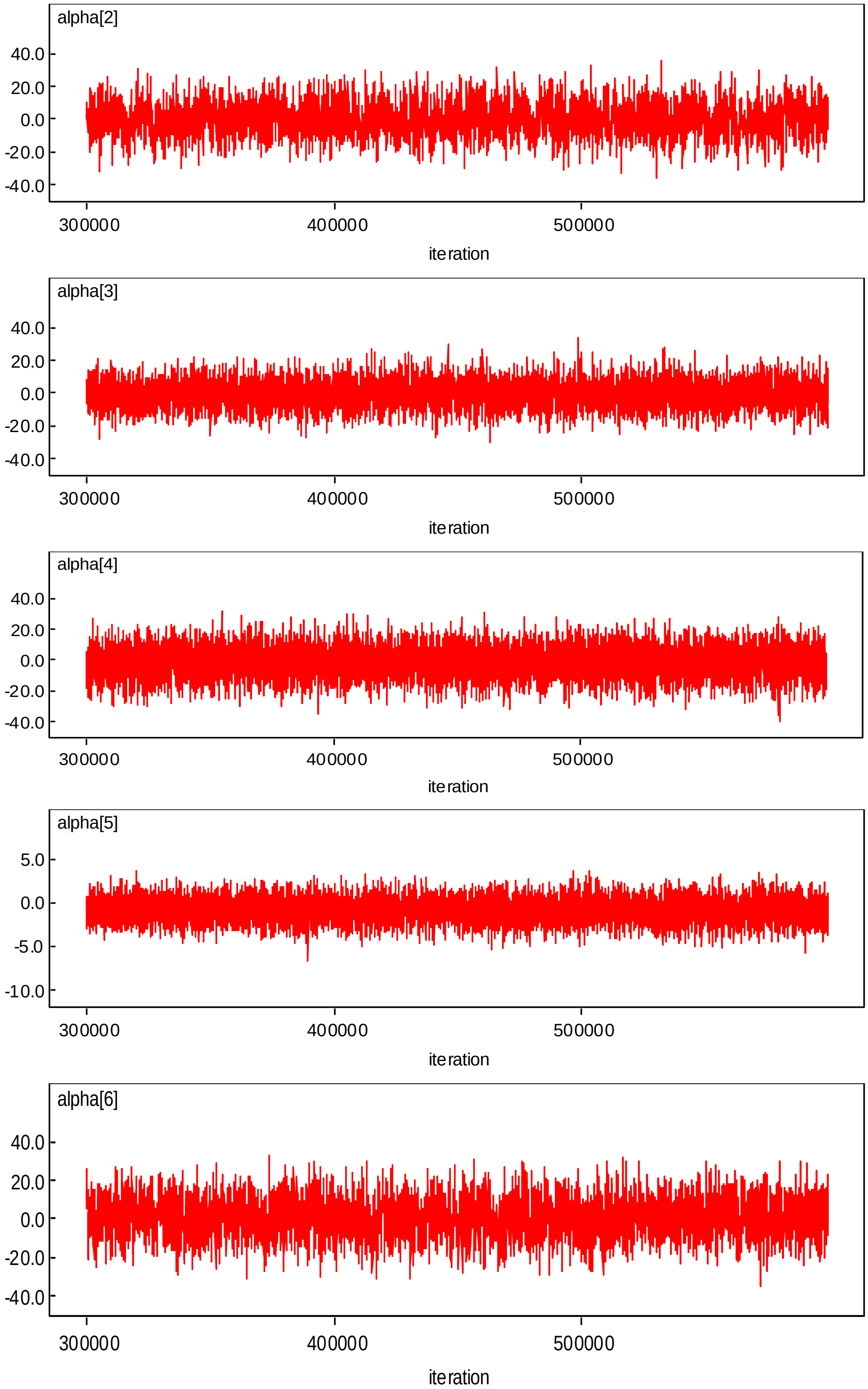


Figure 1: History Plots of Model parameters alpha for 1-6 zones within the nation

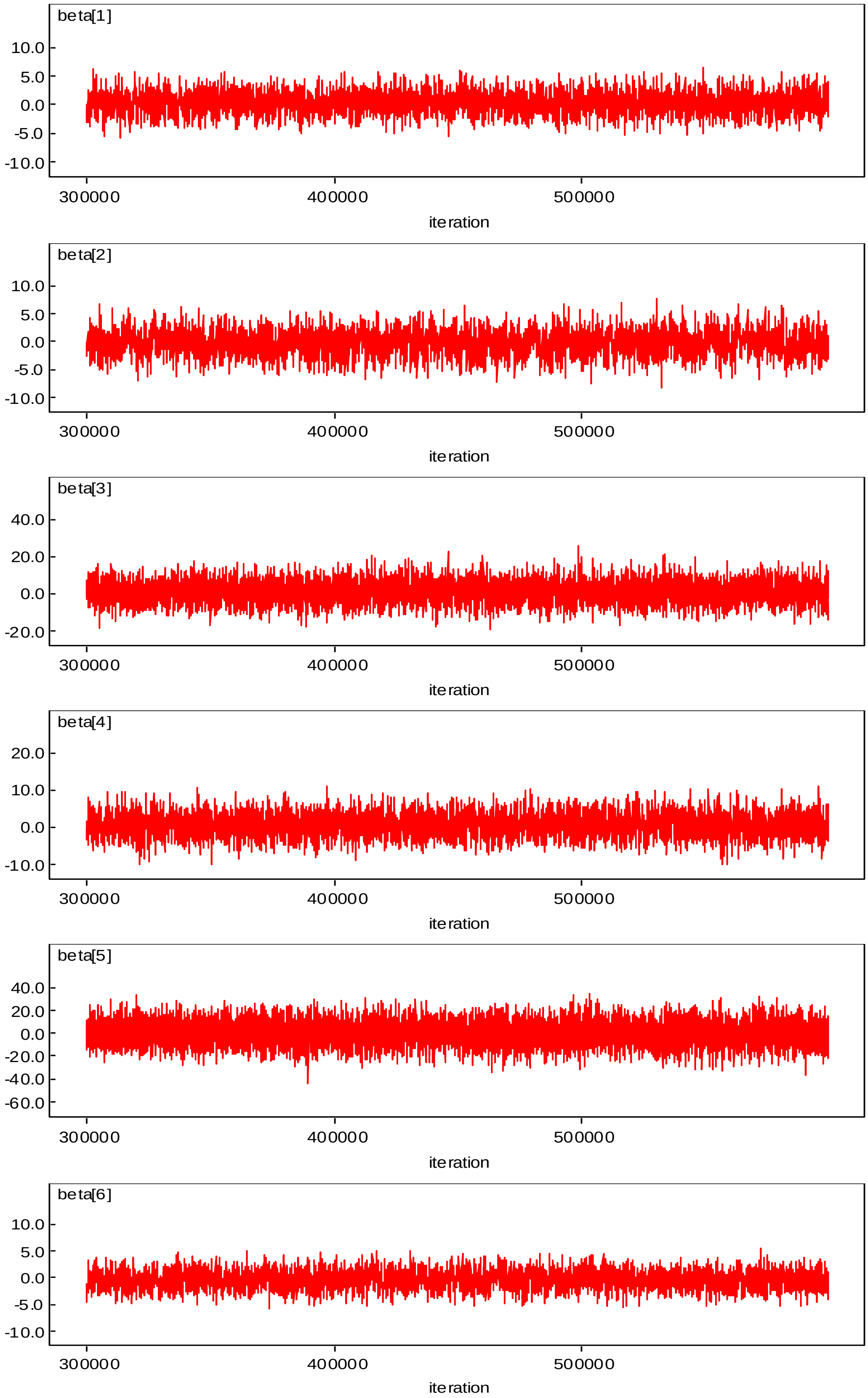


Figure 2: History Plots of Model parameters beta for 1-6 zones within the nation

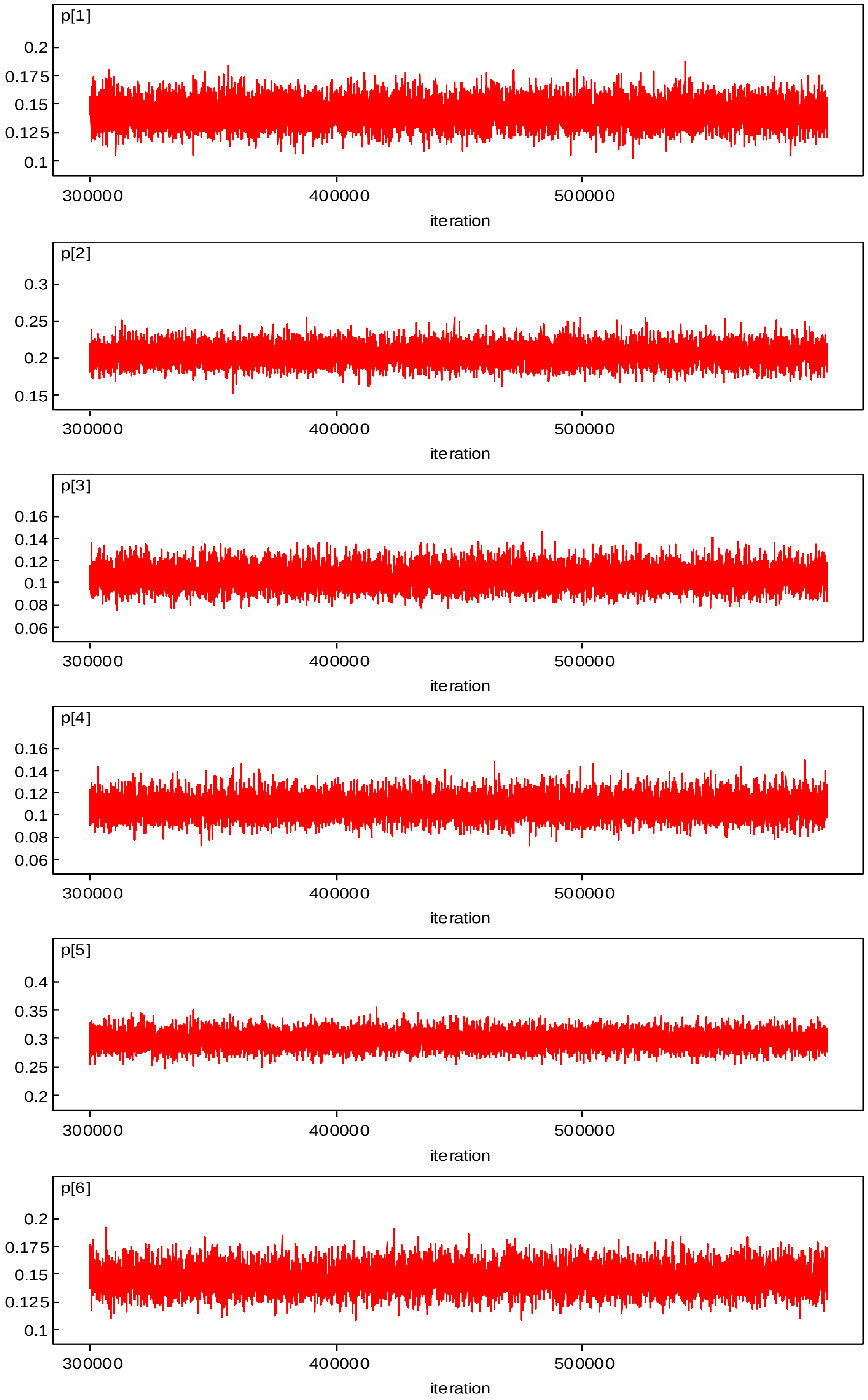


Figure 3: History Plots of Narcotic Drug Use Prevalence (p) for zones within the Nation.

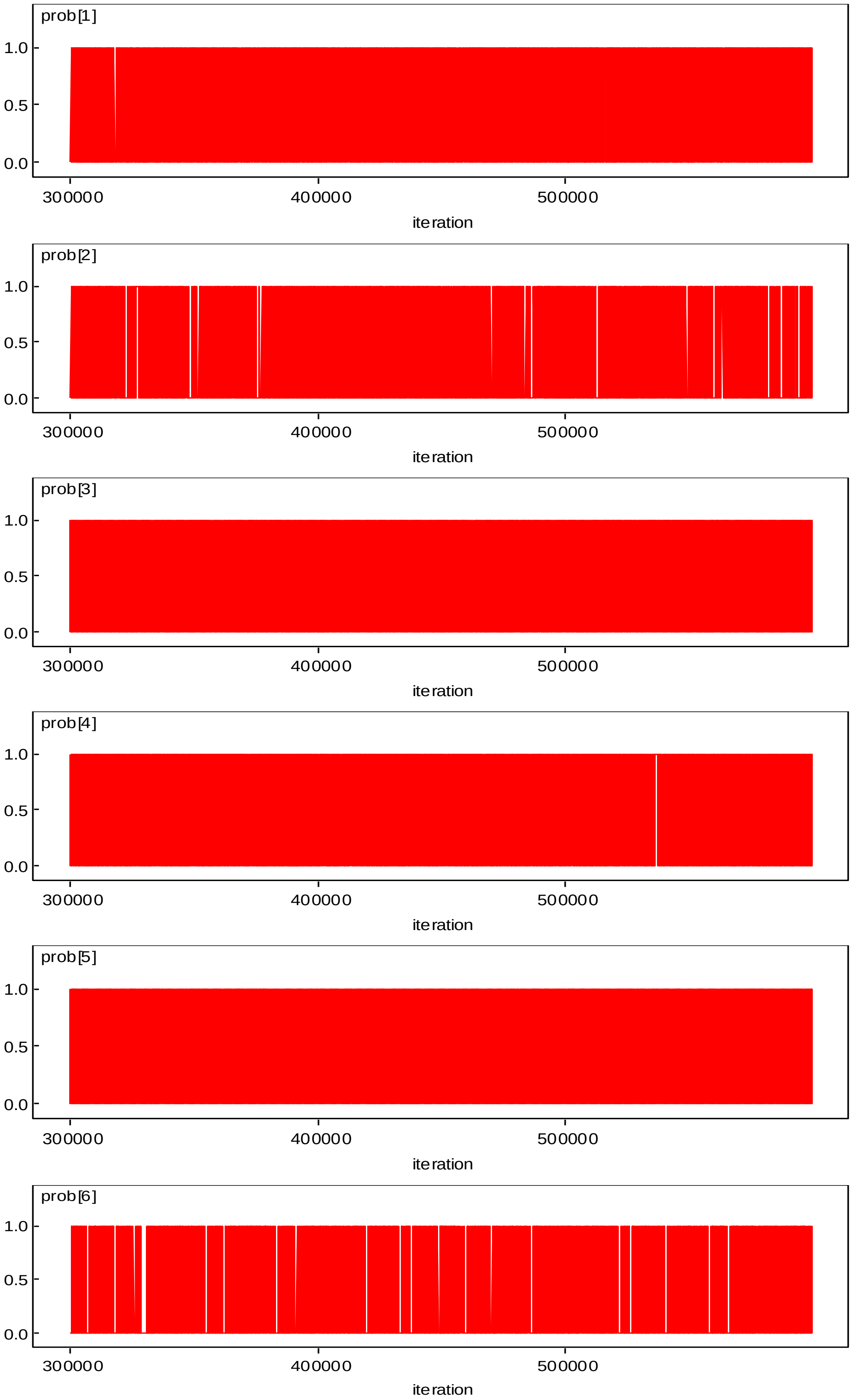


Figure 4: History Plots of risk level of Unemployment Rate of zone within the nation

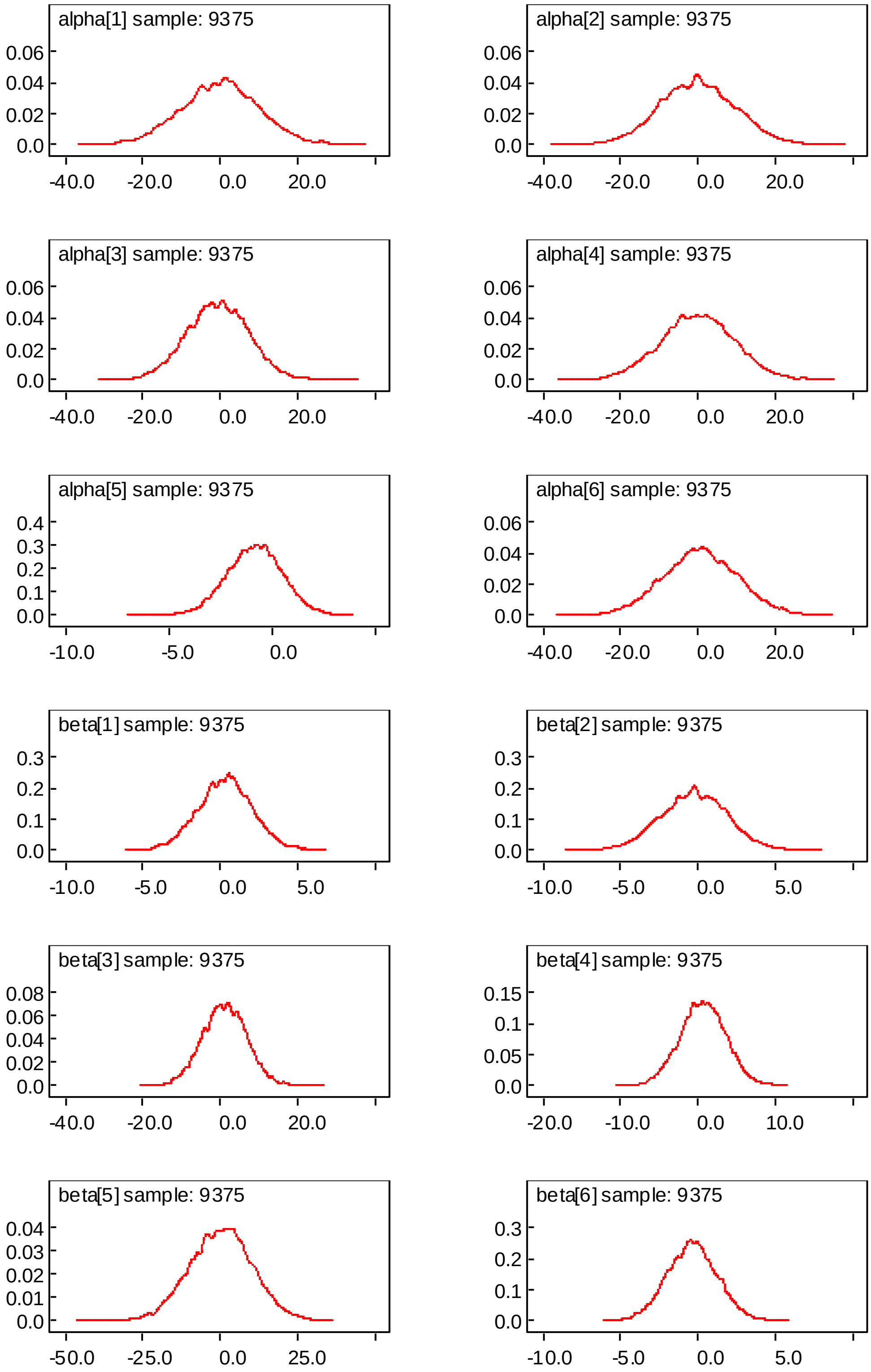


Figure 5: Density plots of model parameters alpha and beta of zones with the nation

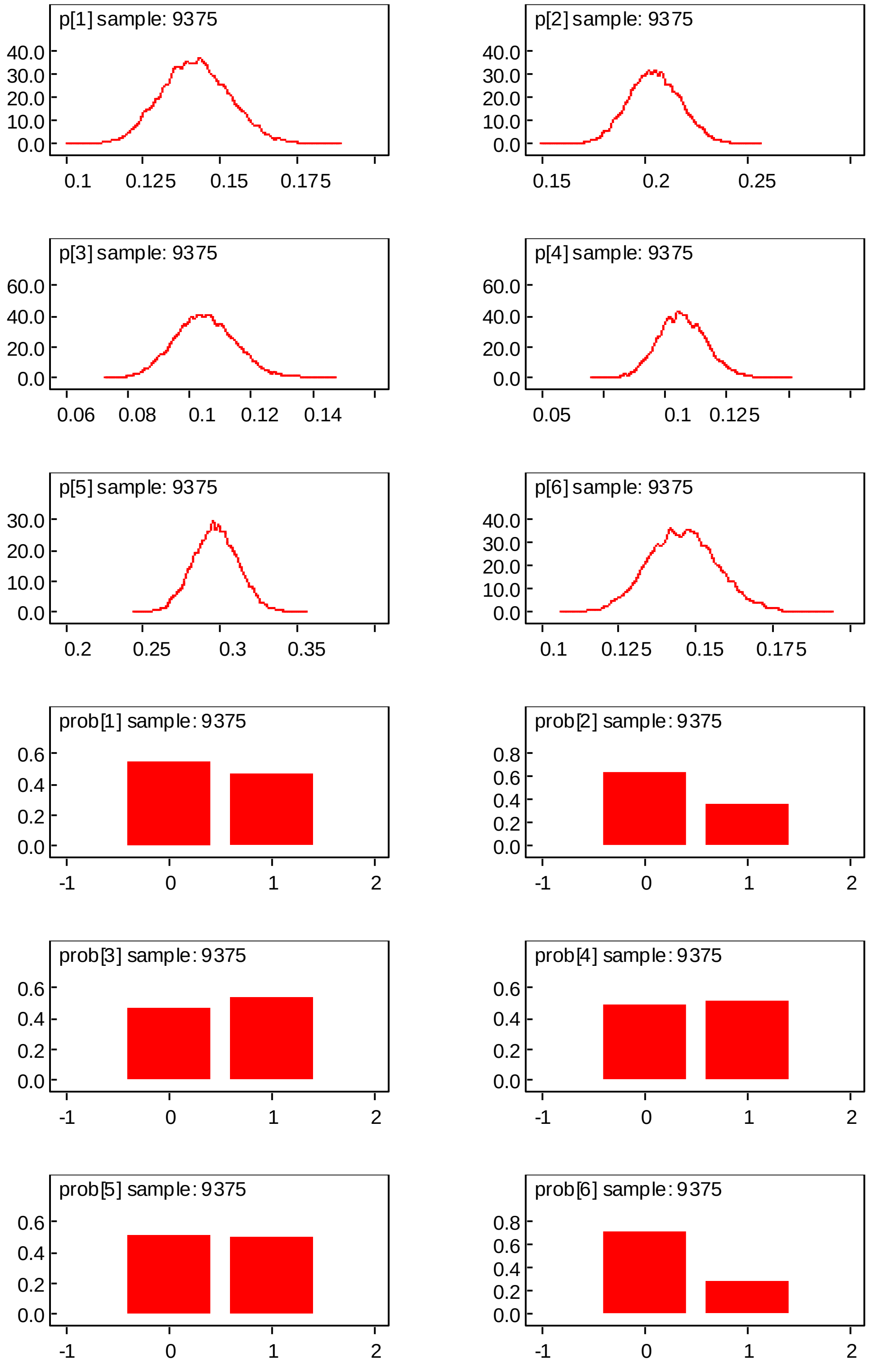


Figure 6: Density plots for Narcotic Drug use Prevalence (p) and Risk Levels of Unemployment Rate (prob) of zones within the Nation

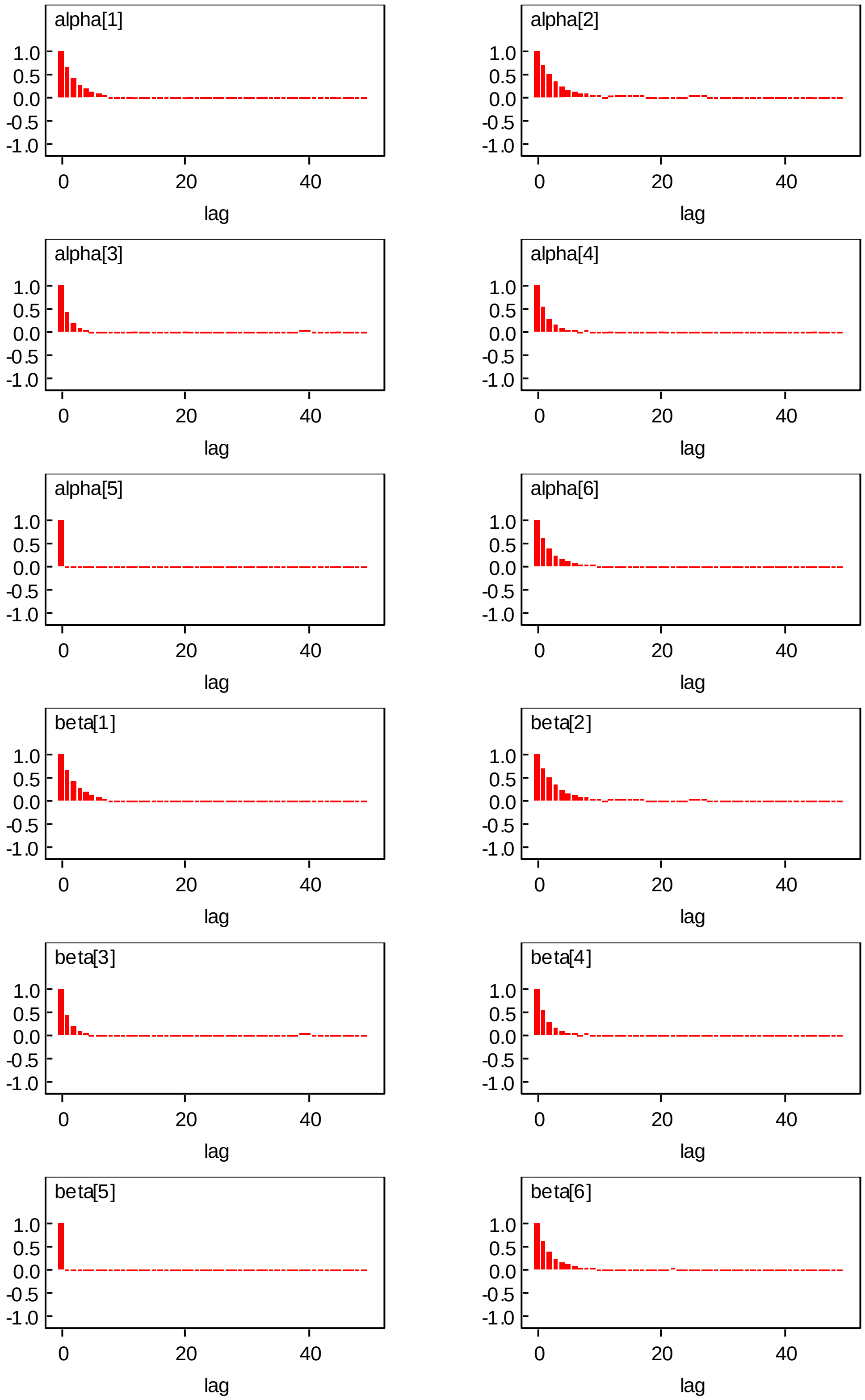


Figure 7: Autocorrelation plots of model parameters alpha and beta of zones within the nation

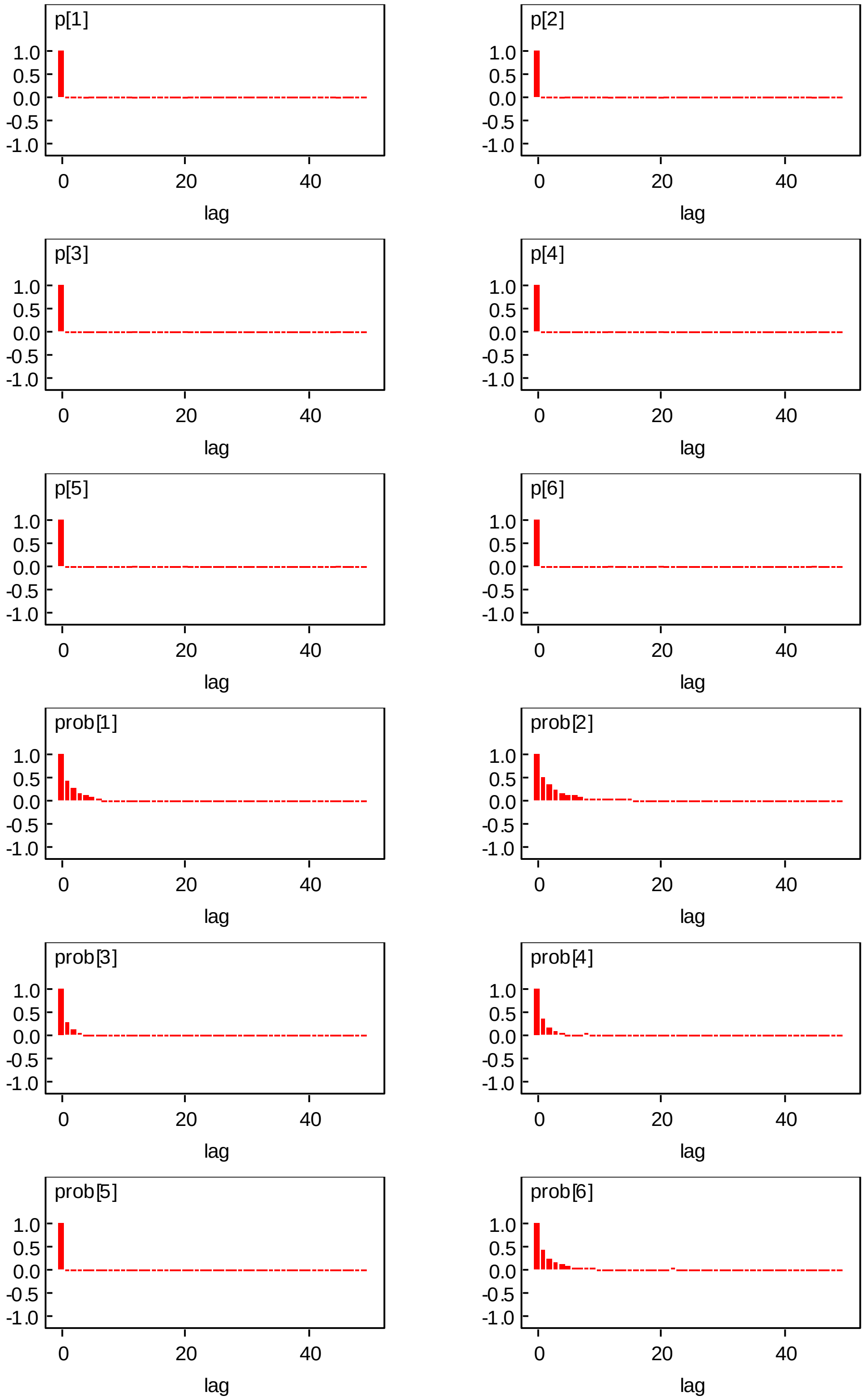


Figure 8: Autocorrelation plots of Narcotic Drug use Prevalence (p) and Risk Levels of Unemployment Rate (prob) of zones in the Nation

Discussion
The actual values of the Narcotic Drug use prevalence and unemployment rates across the zones in each of the six (6) geo-political zones in Nigeria is captured in Table 2. As relayed on this table, the zones include the North – East, North – West, North – Central, South – East, South – West and South – South Zones. These values were computed from actual aggregate data (in the year 2018) on the number of narcotic drug users and number of unemployed persons across the zones in Nigeria. The number of narcotic drug users per 1000 persons were determined for each zone and used as number of trials for the Binomial distribution in the course of the Bayesian statistical modelling. The unemployment rate data sets for each zone in Nigeria also serve as input to the

model.

As earlier mentioned, the incompleteness of this data (aggregates) limits the use of the Classical Logistic Regression Model (Taeryon et al., 2008). This limitation calls for the development of a Bayesian Statistical Modelling Procedure using the MCMC – Gibbs algorithm on the WINBUG platform.

After the development of the model, convergence diagnostic checks were conducted for each model parameter and measure in order to ascertain model adequacy. The history plot, density plots and auto-correlation plots were used for this purpose. See Figures 1– 8 for results of the zones. Observe that the history plots show that the model parameters and measures are well – mixed. This is because they traverse the posterior domain rapidly with nearly constant mean and variance. The model prior distribution for alpha and beta is normal (0, 0.01). The density plots of these priors reflect this distribution which further validates the model. The auto-correlation plots of each parameter and measure depict the independence of the samples generated. This is because the auto-correlation plots become negligible fairly quickly, after a few lags.

Model parameters and measure values for each zone in each geo-political zone were captured on Table 4. Details on Table 4 include the mean, standard deviation, Monte Carlo Simulation Error, median and 95 % credible interval. Estimates of the model parameters (alpha and beta) and their measures (prevalence rate (p)) and level of risk (prob (beta > 0)) and their respective mean values. The values of alpha and beta for each zone in the nation helps to determine the relationship between the narcotic drug use prevalence and unemployment rate. This is achieved when they are plugged into equation (1). The closeness of values of the actual Narcotic Drug Use Prevalence Rate (p) on Table 2 for each zone and their respective simulated values ($p[i]$)on Tables 4 further validates the model.

As established in the mathematical analysis, the sign of the model parameter; alpha assists in determining the impact of other factors latent to the study on the narcotic drug use prevalence in a zone while, the sign of beta assists in determining whether the unemployment rate in a zone is a risk factor of its narcotic drug use prevalence. The level of this risk is computed using Equation (6). Which is, the proportion of times (%) that reduced levels of Zonal unemployment rate of a zone, will have negative effect on the prevalence of narcotic drug use. The overall results show that factors latent to the study negatively impact on the narcotic drug use prevalence in four (4) zones (North-West, North-Central, South-East and South-West) while, unemployment rate is a risk factor to narcotic drug use prevalence in four (4) zones (North-East, North-Central, South-East and South-West). Details of these results are on Table 3. Note that posterior median values of α_i and β_i were used in this work when the standard deviation is high. Otherwise, the mean values were used.

The result obtained in this paper also conform to that of other authors such as (Taeryon et *al.*, 2008) who demonstrated success in Bayesian Regression Modeling Approach in handling cases of incomplete data. Besides the success of the Bayesian Statistical Simulation Modeling approach in handling cases of incomplete data, its robustness has been proven and demonstrated in handling; Semi Parametric Binary Response data that places restrictions on the link function with the exception of monotonicity and known location and scale (Michael *et al.*, 1996), Large – Scale Bayesian Logistic Regression for Text Categorization (Alexander *et al.*, 2007) and a Bayesian Ordinal Logistic Regression Model to correct for inter-observer measurement error in a geographical oral health study (Samuel *et al.*, 2005). Also on the list are; A Bayesian Approach to Robust Binary Non Parametric Regression that assumes the argument of the link function is an additive function of the explanatory variables and their multiplicative interactions, a Bayesian Goodness of Fit Test and Semi Parametric Generalization of Logistic Regression with Measurement Data (Angela *et al.*, 2013) and the Hierarchical Bayesian Logistic Regression to forecast metabolic Control in type 2 Diabetes Mellitus Patients *(*Arianna *et al.,* 2016). The aforementioned researchers made use of incomplete data in their modeling activity. In the same vein, this work has demonstrated success in the use of the Bayesian Regression Modelling Approach in handling incomplete data challenge while modeling the relationship between Zonal Narcotic Drug Use Prevalence and Unemployment rate.

CONCLUSION

A Bayesian Binary Logistic Modelling Procedure was developed for relating Narcotic Drug Use Prevalence and Unemployment Rate using incomplete data. It was established that unemployment rate is a risk factor to narcotic drug use prevalence in four (4) zones of Nigeria namely: North-East, North-Central, South-East and South-West. Apart from unemployment, some factors latent to the study were found to impact negatively on four (4) zones (North-West, North-Central, South-East and South-West) but positively impact on the North-East and South-South zone. The study recommends that the modelling procedure should be applied to similar problems with the challenge of incomplete data. The Nigerian government should make frantic efforts in reducing unemployment rates in the four (4) zones (North-East, North-Central, South-East and South-West) where it is a risk factor to narcotic drug use prevalence. The attention of the Nigerian government should be drawn to the South-South zone where unemployment rate is not a risk factor to narcotic drug use prevalence but other factors latent to the study are impacting positively on it. The North-East zone where unemployment rate is a risk factor and other factors latent to the study are positively impacting should be noted also.

REFERENCES

Adebimpe, A. A. and Folashade, B. O. (2018). Substance Abuse among Females in Nigeria. *International Journal of Business and Social Science*, 9(5). Retrieved from https://doi.org/10.30845/ijbss.v9n5p12. Accessed 29/08/2019.

Agada, P.O., Udoumoh, E. F. and Gboga, H.P. (2019). A Bayesian Simulation Modeling Approach to Predicting Maternal Age-Specific Infant Survival Outcomes with Incomplete Data. *FUDMA Journal of Sciences (FJS),* 3(4):328-344.

Agada, P.O., Ogwuche, I.O. and Eneojo, J.O. (2021). A Bayesian Simulation Approach to Modeling the Relationship Between Rice Mill Loss Rate and Moisture Content with Incomplete Data: A Case Study. *FUW Trends in Science & Technology Journal*, 6(2):381-390.

Alexander, G., David, D.L and David, M. (2007). Large – Scale Bayesian logistic regression for text categorization. *Technometrics special issue on statistics in information technology*. 49(1):291- 304.

Amiri, M., Taheri, Z., Hosseini, M., Mohsenpour, M., and Davidson, P. M. (2016). Factors Affecting Tendency for Drug

Abuse in people Attending Addiction Treatment centres: A Quantitative Content Analysis. Journal of Addiction Research and Therapy, 7(270). https://doi.org/10.4172/2155-6105.1000270.

Anderson, S. (2022). What is the Unemployment Rate? Rates by State. Retrieved from https://www.investopedia.com/terms/u/unemploymentrate.asp Accessed 11/11/2022.

Angela, S., Adam, J.B. and Timothy, E.H. (2013). A bayesian goodness of fit test and semiparametric generalization of logistic regression with measurement data. *Biometrics*. Pp. 508 - 519.

Arianna, D., Alberto, M., Posquala, D., Giulia, C., Marsida, T., Lucia, S., Carlo, C., Luca, C., and Riccardoi, B. (2016). Hierarchical Bayesian Logistic Regression to forecast metabolic Control in type 2 Diabetes Mellitus Patients. AMUA Annu Symp. Proc., Pp. 470 – 479.

Azagba, S., Shan, L., Qeadan, F. and Wolfson, M. (2021). Unemployment rate, Opioids Misuse and other Substance abuse: quasi-experiment evidence from treatment admissions data. *Journal of BMC Psychiatry vol 9, no 1.*

Choi, T., Schervish, M. J., Schmitt, K. A., & Small, M. J. (2008). A Bayesian Approach to a Logistic Regression Model with Incomplete Information. Biometrics, 64(2), 424–430. http://www.jstor.org/stable/25502078

Drug Enforcement Administration (2020): *Drug Fact Sheet*. Retrieved from https://www.dea.gov/sites/default/files/2020-06/Narcotics-2020.pdf Accessed: 10/07/2022.

Drug Use in Nigeria (2018). United Nations Office on Drugs and Crime. Retrieved from https://www.unodc.org/documents/data-and-analysis/statistics/Drugs Accessed 23/08/2019

Hayes, A. (2022). What is unemployment? Understanding causes, types, measurement. Retrieved from https://www.investopedia.com/term/u/unemployment.asp Accessed 10/11/2022.

Holderness, C. G. (2016). Problems Using Aggregate Data to infer individual Behavior: Evidence from Law, Finance and Ownership concentration, *Journal of Critical Financial Review*,5:1-40.

Jacob, R. (2016). Using Aggregate Administrative Data in social policy Research. OPRE Report, University of Michigan.

Jatau, A. I., Sha'aban, A., Gulma, K. A., Shitu, Z., Khalid, G. M., Isa, A., Wada, A. S and Mustapha, M. (2021). The Burden of Drug Abuse in Nigeria: A Scoping Review of Epidemiological Studies and Drug Laws. Journal of Public Health Rev. https://doi.org/10.3389/phrs.2021.1603960

Mclachlan, G. J., Krishnan, T., and Ng, K. S. (2004). The EM algorithm. (2004),Humboldt-Universitatzu Berlin, Center for Applied Statistics and Economics(CASE), Berlin.

Michael, A., Newton,C. C. and Rick, C. (1996). Bayesian inference for semi parametric binary regression. *Journal of the American Statistical Association*. Pp. 142 – 153.

Mohammed, R. A. and Mohamed, G. I. (2018). On Estimation Methodes for Binary Logistic Regression Model with Missing Values. *International Journal of Mathematics and Computational Sciences*. 4(3):79-85.

Obot, I. S. (2019). Drug in the Nigerian population: Availability, use, consequences and policy implications. *Journal of Bulletin on Narcotics*, 62(13).

Pourallahvirdi, M., Rahmani, F., Ranjbar, F., Ebrahimi, B. H., and Ettehadi, A. (2016). Major causes of Drug Abuse from the viewpoint of Addicted persons Referred to Addiction Treatment Centers of Tabriz city, Iran. Arch Neurosci; 3(3).https://doi.org/10.5812/archneurosci.37653

Rastegari, A., Haghdoost, A.A. and Baneshi, M. R. (2013). Factors Influencing Drug Injection History among Prisoners: A Comparison between classification and Regression Trees and Logistic Regression Analysis. Journal of Addict Health, 5(2):7-15.

Samuel, M., Mwalili, M., Emmanuel, L. and Dominique, D. (2005). A Bayesian Ordinal Logistic Regression Model to correct for interobserver measurement error in a geographical oral health study. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. Pp. 77-93.

Scott, L. M.(2007). Introduction to Applied Bayesian Statistics and Estimation for Social Scientists. New York:Springer.

Taeryon, C., Mark, J.S., Ketra, A.S. and Mitchell, J.S. (2008). A Bayesian approach to a Logistic Regression Model with Incomplete data. Biometrics. Pp. 424 - 430.