International Invention of Scientific Journal

Available Online at http://www.iisj.in

- eISSN: 2457-0958

Volume 04 | Issue 09 | September, 2020 |

BAYESIAN LOGISTIC REGRESSION ANALYSIS OF FACTORS AFFECTING JOB SATISFACTION AMONG SCHOOL TEACHERS: A CASE STUDY AT PRIMARY AND SECONDARY SCHOOLS IN SHASHEMENE TOWN, ETHIOPIA

Corresponding Author : Sano kedir Mohammed Department of Statistics,

Madda Walabu University, Robe, Ethiopia

P.O.Box:247: E-mail Address:sanokedir@yahoo.com

Article Received 30-08-2020, Accepted 26-09-2020 , Published 30-09-2020

Abstract:

This study has dealt with factors affecting the level of job satisfaction among school teachers at Shashemene town in 2010/11 academic year. Data were obtained from primary and secondary sources. A cross sectional survey was conducted on a total of 405 teachers from different government and non-government schools using stratified random sampling technique. A designed questionnaire was used to obtain data on socio-economic and demographic information, overall levels of job satisfaction using Minnesota satisfaction questionnaire and teachers feeling in many aspects of the job. It was found that eight unobserved variables had significant effects on job satisfaction. These were motivation, working condition, payment condition, co-workers condition (colleagues), school administration, curriculum condition, personal life and student behaviors. Other factors such as school facility and growth opportunity were not found to have significant effects at 5% significance level.

1. INTRODUCTION

Teachers are the pillars of the society, who help students to grow to shoulder the responsibility of taking their nation ahead of others. They play a very crucial role for the transfer of knowledge in schools and in achieving the objectives of Ethiopia in development. More than one half of the additional teachers are needed in sub Saharan Africa. On top of that, approximately one million teachers have to be replaced every year to balance out the attrition of teachers (UNESCO, 2010). Without teachers, there can be no education, and without education, sustained economic, political and social development is not possible. To evaluate education quality, the most obvious indicator is learning achievement. Moreover to analyze its determinants, background information is needed on students, teachers and schools. For the analysis under taken here, information on teachers is particularly relevant. Most importantly, the data must include information to draw conclusions about teachers' job satisfaction.

Porter and Steers (1973) argued that the extent of employee job satisfaction reflects the cumulative level of met worker's expectations. Employees expect their job to provide a mix of features like pay, promotion, autonomy for which the employee has certain preferential values. The range and importance of these preferences vary across individuals, but when the accumulation of unmet expectation becomes sufficiently large there is less job satisfaction and greater probability of withdrawal behavior. Many countries in sub-Saharan Africa are currently reporting teacher attrition rates which are very low, and unlikely to be sustained. If there are no voluntary resignations, an education system should expect an attrition rate of between 3% and 4% annually arising simply from retirement, illness and death (Amina*et al*., 2010).

Teachers desire security, recognition, new experience and independence economically. When these needs are not fulfilled they become tense. Disgruntled teachers, who are not satisfied with their job cannot be committed and productive, hence they would not perform at the best of their capabilities. Dissatisfaction among workers is undesirable and dangerous in any profession; it is suicidal if it occurs in teaching profession. If factors responsible for dissatisfaction can be differentiated, attempts can be made either to change those conditions or to reduce their intensity so as to increase the holding power of the profession. Teaching profession is facing problems related to teachers' job satisfaction. According to Kim and Loadman (1994), teachers' job satisfaction is an effective response to one's situation at work. Thus, teacher job satisfaction refers to a teacher's effective relation to his or her teaching role. More often than ever before, teachers are under tremendous pressure from politicians, parents, and local communities to deliver quality education to all children. Teachers have a critical role to play in the schools along with supporting development activities in the wider community. They are central to the realization of national and international educational goals. Most governments and other key stakeholders recognize the crucial importance of improving the living and working conditions of teachers in order to achieve the desired improvements in quality and access to basic education.

1.2 Statement of the Problem

The determinant factors of teachers job satisfaction is a major problem of many researchers, policy makers and education leaders, because it is one of a vital factor that affects student achievement. Job satisfaction is dynamic; it is a condition where individuals are contented or discontented with their jobs. Lack of job satisfaction may lead to strikes, work to rule, absenteeism, resignation, low performance and disciplinary problems.

Teaching profession is facing problems due to teachers' job satisfaction. The general perception is that teachers' in the government school are dissatisfied with their profession. They are said to be dissatisfied with teaching in the government schools. If the problem is true that the government school teachers are dissatisfied, what then is this dissatisfaction? In what aspects, are they satisfied? Therefore, it is necessary to probe into this matter through a careful study.

- Is there an association between teachers' overall job satisfaction and socio-demographic variables?
- What are the most significant factors that influence overall job satisfaction of teachers with respect to demographic variables?

1.3 Objectives

- The general objective of the study has been to determine or asses factors affecting overall job satisfaction among teachers at Shashemene town.

The specific objectives have been to:-

- Identify factors that influence most, the overall level of job satisfaction among teachers.
- Compare statistical model of job satisfaction using Bayesian approach of different prior parameters.

2. Methodology

The study was conducted in Shashemene town, a rapidly growing town in Oromia region of Ethiopia. Shashemene is centrally located, which is 250 km far from the capital city of Ethiopia, i.e. Addis Ababa.

Geographically, it lies between Latitude of 7, 2000 (712'0.000"N) and Longitude of 38, 6000 (3836'0.000"E). The altitude of this town ranges from 1500metres to 2300 meters above sea level.According to CSA (2007) report, the estimated population size of the town was 102,062, of which 51,477 were males and 50,585 were females. Educational institutes in the town, include one governmental TVET college, two non-governmental university college and four private colleges, three governmental high schools and two governmental preparatory schools.

In Shashemene town, there were 48 primary schools of which 10 were governmental and 38 were non-governmental schools, having 278 and 501 teachers, respectively. There were also 10 high schools of which sevenwere non-governmental and three were governmental and there were six preparatory schools of which four were non-governmental and two were governmental. The total numbers of teachers employed in preparatory schools andhigh schools were 69 and 254, respectively (Education Bureau of Shashemene Town, 2010).

The target populations for the study have been teachers employed in primary and secondary schools of governmental and non-governmental during 2010/11 at Shashemene town. These teachers were full timer paying tax from their salary.A cross-sectional survey design with stratified random sampling technique was used in this study.

3.5.2 Bayesian Logistic Regression Model

The beginning of the 21st century, found Bayesian statistics to be fashionable in science. Bayesian statistics provides a very different approach to the problem of unknown model parameters. The basis for Bayesian inference derived from simple probability theory (Baye`s theorem). The main difference between the classical and the Bayesian approach is that the latter considers the unknown parameters as random variables instead of considering just a single value that are characterized by a prior distribution. This prior distribution is combined with the traditional likelihood to obtain the posterior distribution of the parameter of interest on which the statistical inference is based. Moreover classical methods to analyze binomial regression data relying on asymptotic inferences while Bayesian methods performed using simple computations apply for any sample size (Casella and Berger, 2002).

Bayesian inference differs from classical inference in treating parameters as random variables and using the data to update prior knowledge about parameters and functional of those parameters. One is also likely to need model predictions and these are provided as part of the updating process. Prior knowledge about parameters and updated (or posterior) knowledge about them, as well as implications for functional and predictions, are expressed in terms of densities.

Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior distribution for the unknown parameters. The foundation of Bayesian statistics is Baye's theorem. Suppose one observes a random variable y and wish to make inferences about another random Variable β, where β is drawn from some distribution f (β). The conditional probability distribution can be written as:

$$f(\beta \mid y) = \frac{f(y \mid \beta) f(\beta)}{f(y)}$$

Where, $f(y) = \int f(y, \beta) d\beta$ as β's are continuous variables.

$\beta = (\beta_0,\ \beta_1, \ldots, \beta_k)$.

$f(\beta)$ is assumed to be prior distribution of the unknown parameters β.

f (β|y) is the posterior distribution of β given the data y.

The advantages of Bayesian approach is, it provides a fairly explicit solution to common problems of statistical inference, new problems of high dimensional data analysis that are coming up because of emergence of high-dimensional data sets, as well as complex decision problems of real life. It can handle presence of prior knowledge or partial prior knowledge. A common problem with pure frequentist models is their inability to incorporate prior knowledge about model parameters. This information may range from simple interval constraints to detailed knowledge of distributional properties revealed in previous or related studies. Increased flexibility in model building can be achieved using a Bayesian framework. Bayesian analysis is more accurate in small samples since it depends on priors.

3.5.2.1 Prior Distribution

Specification of the prior distribution is important in Bayesian analysis since it influences the posterior inference. Existing evidence about the parameters of interest may be available from earlier studies or from experts' opinions, and can be formalized as prior distribution of the parameter of interest. A prior distribution can be non informative, informative, or very informative.

Non informative prior distributions are used in cases in which no extra-sample information is available on the value of the parameters of interest (Mila et al., 2003 and Clark et al., 2002). In statistical terms, this lack of knowledge is represented with a distribution that attributes approximately the same probability to each possible parameter value. Informative prior distributions are used when some prior knowledge about the parameters of interest is available, such as when existing belief or evidence indicates that a parameter should take a value within a range. Formally, this knowledge is represented with a distribution that has a known mean and large variance. Very informative prior distributions are used when very strong prior knowledge about the parameters of interest is available, such as when existing belief or evidence indicates that a parameter of interest should be a specific value. In statistical terms, this knowledge can be represented with a distribution that has a known mean and small variance.

The choice of a non informative, prior distribution typically involves a certain amount of subjectivity; historically, this has been one of the reasons for disagreement between Bayesian and classical statisticians. However, as the size of the data increase, the likelihood becomes increasingly influential relative to the prior where in the limit, the prior is immaterial or one needs to specify a prior that will not influence the posterior distribution such as normal or uniform distribution. For this study, the most common prior for logistic regression parameters was used, which was a multivariate normal distribution with mean vector μ and covariance matrix Σ. That is;

$$\beta \sim N_k(\mu, \Sigma) \qquad (3.7)$$

The most common choice for μ is zero vector, and Σ is a diagonal matrix $\Sigma = diag(\sigma_0^2, \sigma_1^2, ..., \sigma_k^2)$, Variances are often considered to be large to make the priors of p.d.f' s non-informative. A common choice for the variances σ_j^2 is in the range from 10 to 100.

Alternatively, the prior distribution can be expressed as:

$$f(\beta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{\beta_j - \mu}{\sigma_j}\right)^2 \right) \qquad (3.8)$$

where j=0,1,2, . . . ,k.

Prior knowledge is summarized in a density f (β), here taken as non-informative priors.

Bayesian estimation of the model parameters requires the specification of a prior distribution for all the unknown parameters. With small samples this choice can be critical, but with larger samples the choice is less crucial, since information in the data heavily outweighs information in the prior.

If the posterior is highly dependent on the prior, then the data (likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contain significant information. In general, any prior distributions can be used, depending on the available prior information.

3.5.2.2. Likelihood Function

While the joint distribution of n independent Bernoulli trials is still the product of each Bernoulli distribution, the sum of independent and identically distributed Bernoulli trials has a Binomial distribution. Specifically, let $Y_1, Y_2, \ldots, Y_n$ be independent Bernoulli trials with success probabilities $P_1, P_2, P_3, \ldots, P_n$, that is Y_i = 1 with probability P_i or Y_i=0 with probability 1- P_i, for i= 1,2, . . . ,n.

As described above, since the trials are independent, the joint distribution of $Y_1, \ldots, Y_n$ is the product of n Bernoulli probabilities. As usual, the likelihood function used by Bayesians matches that from frequents inference. When one has the probability of success, which in logistic regression, varies from one subject to another, depending on their covariates, the likelihood contribution from the i[th] subject is binomial:

$$L(\beta \mid Y) = \prod_{i=1}^{n} \left[P_i^{Y_i} (1-P_i)^{(1-Y_i)} \right] \ldots\ldots(3.9)$$

Where, P_i represents the probability of the event for subject i who has covariate vector X_i, Y_i indicates the presence, Y_i=1, or absence Y_i=0 , i=1,2,…,n of the event for that subject. We know that, in equation (1) logistic regression the probability of success written as:

$$P_i = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \ldots \beta_k x_{ik})}}{1 + e^{\beta_o + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}}}$$

Since individual subjects are assumed independent from each other, the likelihoods function over a data set of subjects is:

$L(\beta|\mathbf{Y})$

$$=\prod_{i=1}^{n}\left(\frac{e^{\beta_o+\beta_1x_{i1}+....\beta_kx_{ik}}}{1+e^{\beta_o+\beta_1x_{i1}+....\beta_kx_{ik}}}\right)^{y_i}\left(1-\frac{e^{\beta_o+\beta_1x_{i1}+....+\beta_kx_{ik}}}{1+e^{\beta_o+\beta_1x_{i1}+.....\beta_kx_{ik}}}\right)^{(1-y_i)} \quad ....(3.10)$$

3.5.2.3 Posterior Distribution

Given the likelihood of logistic regression and the prior distribution given above, the posterior distribution of the Bayesian logistic regression model is defined as the product of likelihood unction and prior distribution function. It is given as:

$f(\beta|\mathbf{Y,X})$

$$=\frac{\prod_{i=1}^{n}p_i^{\,y_i}(1-p_i)^{1-y_i}f(\beta)}{f(y)} \propto \prod_{i=1}^{n}p_i^{\,y_i}(1-p_i)^{1-y_i}f(\beta)$$

$$=\prod_{i=1}^{n}\left[\left(\frac{e^{\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}}}{1+e^{\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}}}\right)^{(1-y_i)}\right]X$$

$$\prod_{j=0}^{k}\frac{1}{\sqrt{2\pi}\sigma_j}\exp\left(-\frac{1}{2}\left(\frac{\beta_j-\mu_j}{\sigma_j}\right)^2\right) \quad ....(3.11)$$

The mean of the posterior distribution can be used as a point estimate of β. However, computing the estimate of β of the posterior distribution may be difficult. Hence one needs to use in this case simulation techniques. The most popular method of simulation techniques is the Marcov Chain Monte Carlo (MCMC) methods, which has been applied in this study.

3.5.2.4 Marcov Chain Monte Carlo Methods

Markov chain Monte Carlo(MCMC) method is a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a large number of steps is then used as a sample from the desired distribution. The quality of the sample improves as a function of the number of steps. Usually it is not hard to construct a Markov Chain with the desired properties. The more difficult problem is to determine how many steps are needed to converge to the stationary distribution within an acceptable error. A good chain will have rapid mixing and the stationary distribution is reached quickly starting from an arbitrary position described further under Markov chain mixing time (Albert, 2008).

A major difficulty towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high dimensional functions. MCMC methods are attempted to simulate direct draws from some complex distribution of interest such as posterior distribution. Also, it is used to generate an irreducible Markov Chain with stationary probabilities, which is the posterior distribution.

In quantitative sciences, the problem of evaluation of integrals of the type given below is often required.

$$\mathrm{I} = \int_x g(x)dx$$

One of the solutions finding of the interest is based on generating random samples and then obtaining the integral shown above by its statistical unbiased estimate, the sample mean. Assuming that the density function f(x) of a random variable enables us to easily generate random values, this can be expressed as:

$$\mathrm{I} = \int_{\mathrm{x}} \left[\frac{\mathrm{g(x)}}{\mathrm{f(x)}}\right] \mathrm{f(x)dx} = \int_{\mathrm{x}} \mathrm{g^*(x)f(x)dx},$$

where, g* (x) = g(x)/f(x). Hence the integral I can be efficiently estimated by generating $x^{(1)}, x^{(2)}, ..., x^{(n)}$ from the target distribution with probability density function f(x) and calculating the sample mean $\hat{I} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}$.

This concept was known from the early days of the electronic computers and was originally adopted by the research team of Metropolis in Los Alamos (Anderson, 2007; Metropolis and Ulam, 2001). The main advantage of this approach is its simplicity. Even if integrals are tractable, nowadays it is much easier to generate samples than calculate high-dimensional integrals. The method described above is directly applicable to many problems in Bayesian inference. Hence for every function of the parameter of interest f(ß|y), one can calculate the posterior mean and variance by:

1. Generating a sample $\beta^{(1)}, \beta^{(2)}, ..., \beta^{(n)}$ from the posterior distribution f (β |y).
2. Calculating the sample mean of f(β |y) by simply calculating the quantity
$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \beta^{(i)}$$

Simulation can also be used to estimate and visualize the posterior distribution of f (β |y*)* itself. The main problem in the above mentioned procedure is how to generate from the posterior density f (β |y*)* random sample. Generally, the most commonly used MCMC techniques are Metropolis-Hasting and Gibbs sampler algorithm. Here in this study Gibbs sampler algorithm was used.

3.5.2.4.1 Gibbs Sampler Algorithm

Gibbs sampling or Gibbs sampler is an algorithm to generate a sequence of samples from the joint probability distribution of two or more random variables. The purpose of such a sequence is to approximate the joint distribution; to approximate the marginal distribution of one of the variables, or some subset of the variables (for example, the unknown parameters or latent variables); or to compute an integral (such as the expected value of one of the variables). Typically, some of the variables correspond to observations whose values are known, and hence need not to be sampled. Gibbs sampling is commonly used as a means of statistical inference, especially Bayesian inference. It is a random algorithm (i.e. an algorithm that makes use of random numbers, and hence produces different results each time it is run), and is an alternative to deterministic algorithms for statistical inference such as variational Bayes or the expectation-maximization algorithm (EM).

The Gibbs sampler (Geman and Geman, 1986) is the most widely used MCMC technique. It is a transition kernel created by a series of full conditional distributions that is a Markovian updating scheme based on conditional probability statements. If the limiting distribution of interest is $f(\beta)$ where β is a k length vector of coefficients to be estimated, then the objective is to produce a Markov chain that cycles through these conditional statements moving toward and then around this distribution. The set of full conditional distributions for β are denoted $\beta$and defined by f **(β)** = $f(\beta \mid \beta_{-i})$ for i = 0,1, 2… k, where the notation β_{-i} indicates a vector of all parameters in βwithout the β_i coefficient.

It is essential that there be a definable conditional statement for each coefficient in the β vector and that these probability statements be completely articulated so that it is possible to draw samples from the described distribution.

These requirement facilities the iterative nature of the Gibbs sampling algorithm which is described as follows:

1. Choose starting values: $\beta^{[0]} = [\beta_0^{[0]}, \beta_1^{[0]}, \ldots, \beta_k^{[0]}]$

2. At the j^th^ step (j = 1,2,3, . . .,n), complete the single cycle by drawing values from the conditional distributions as follows the k distributions given by:

$$\beta_0^{[j]} \sim f(\beta_0 \mid \beta_1^{[j-1]}, \beta_2^{[j-1]}, \ldots, \beta_{k-1}^{[j-1]}, \beta_k^{[j-1]})$$
$$\beta_1^{[j]} \sim f(\beta_1 \mid \beta_0^{[j]}, \beta_2^{[j-1]}, \ldots, \beta_{k-1}^{[j-1]}, \beta_k^{[j-1]})$$
$$\vdots$$
$$\beta_k^{[j]} \sim f(\beta_k \mid \beta_0^{[j]}, \beta_1^{[j]}, \ldots, \beta_{k-1}^{[j]})$$

3. Return $\beta^{[1]}, \ldots, \beta^{[n]}$.

Once convergence is reached, all simulated values are considered from the target posterior distribution and a sufficient number of samples should then be drawn so that all areas of the posterior are explored. Notice that the important feature during each iteration of the cycling through the β vectors, conditioning occur on βvalues that have already been sampled for that cycle; otherwise the β values are taken from the last cycle. So in the last step for a given *j* cycle, the sampled value for the k^{th}parameter gets to condition on all j-step values.

The statements above clearly demonstrate that it is required to have the full set of conditional distributions to run the Gibbs Sampling algorithm. The Gibbs sampleras implemented in WinBUGS to approximate the posterior distributions for each parameter is used in the Bayesian analysis.

3.5.2.5. Convergence of the Algorithm

Before one summarizes simulated parameters, one must ensure that the chains have converged. The convergence of the algorithm is essential for producing results from the posterior distribution of interest and the term convergenceof an MCMC algorithm refers to whether the algorithm has reached its equilibrium (target) distribution. The practical results from a given MCMC analysis are not to be considered reliable until the chain has reached its stationary distribution and had time to sufficiently mix throughout. However, convergence diagnosis will be adopted to determine whether the sampler has reached its stationary distribution.

3.5.2.6 Tests for Convergence Diagnostics

Several diagnostic tests have been developed to monitor the convergence of the algorithm. The simplest way is to monitor the MC error, since small values of this error indicate that one has calculated the quantity of interest with precision.

Monitoring autocorrelations is also very useful since low or high values indicate fast or slow convergence, respectively. A second way is to monitor the trace plots:the plots of the iterations versus the generated values. If all values are within a zone without strong periodicities and (especially) tendencies, then one can assume convergence.

A poor choice of starting values and/or proposal distribution can greatly increase the required burn-in time (trending section). For showing evidence of poor mixing, time series trace is seen for a minimum burn in period for some starting value. Various convergence tests have been used to assess whether stationarity has indeed been reached (Walsh, 2004).

Among several ways of convergence assessment in diagnosing summary statistics from in-progress models, the most popular and straightforward convergence methods are discussed as follows.

Autocorrelation

High correlation between the parameters of a chain tends to give slow convergence, whereas high autocorrelation within a single parameter chain leads to slow mixing and possibly individual non convergence to the limiting distribution, because the chain will tend to explore less space in finite time. In analyzing Markov chain autocorrelation, it is helpful to identify lags in the series in order to calculate the longer-run trends in correlation, and in particular whether they decrease with increasing lags. Diagnostically, though, it is not necessary to look beyond 30 to 50 lags (Merkle et al., 2005 and Gill, 2002). These can be accessed in WinBUGS using the "autcor" button on the Sample Monitor Tool.

Time Series Plot

Iteration numbers on x-axis and parameter value on y-axis are commonly used to assess convergence (Merkle et al., 2005 and Gill, 2002). If the plot looks like a horizontal band, with no long upward or downward trends, then one has evidence that the chain has converged. These can be accessed in WinBUGS using the "history" button on the Sample Monitor Tool.

Gelman-Rubin Statistic

For a given parameter, this statistic assesses the variability within parallel chains as compared to variability between parallel chains (Merkle et al., 2005, and Gill, 2002). The model is judged to have converged if the ratio of between to within variability is close to 1. Plots of this statistic can be obtained in WinBUGS by using the "bgr diag" button. The green line represents the between variability, the blue line represents the within variability, and the red line represents the ratio. Evidence for convergence comes from the red line being close to 1 on the y-axis and from the blue and green lines being stable (horizontal) across the width of the plot.

Density Plot

A classic sign of non convergence is multimodality of the density estimate (Gill, 2002). If the density plot become uni-modal, it indicates convergence of the chain.

Following sufficient period to mix and converge, the chain approaches or converges to its stationary distribution, from which one throws away all the data; this is what one calls the burn-in period. After the burn-in each iterates, MCMC posterior summary statistics obtained can be used for inference.

3.5.2.7. Assessing Accuracy of the Bayesian Logistic Regression Fitting

Once one is sure that the convergence has been achieved, one needs to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference. The more samples one saves, the more accurate is the posterior estimates. One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter. This is an estimate of the difference between the mean of the sampled values (which one is using as the estimate of the posterior mean for each parameter) and the true posterior mean.

As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5 percent of the sample standard deviation. The Monte Carlo error (MC error) and sample standard deviation (SD) have been computed along with the summary statistics while using WinBUGS software.

4. RESULTS AND DISCUSSION

The data comprised of a sample of 405 teachers working at elementary and high school level in 2010/11 at Shashemene town. The response variable in this study was the level of teacher's overall job satisfaction which was computed from 20 items of MSQ.

It can be seen from Table 4.1 that, among the sampled respondent teachers, 78% were males and 22% were females; 51.9% were from non-government schools and 48.1% were from government schools; and 67.9% teachers were from elementary schools and 32.1% were from secondary school level. The teachers' level of education ranged from certificate to second degree, with the proportion of 9.4%, 48.1%, 42.2% and 0.2% for certificate, diploma, first degree and second degree, respectively. Among the respondents 47.2% were single, 51.4% were married and 1.5% were divorced. Respondent teachers from natural science, languages, mathematics teachers and social science were 31.4%, 31.1%, 19% and 18.5%, respectively. Most of the teachers, 86.2%, were relatively young who were below the age of 40 years. The most frequent age group of teachers was 26-30 years, having 37.3%. The least, 6.2% were in the age group of 41-45 years. 7.6% were in the age group of 46 years and above. The weekly work load of 16.3% respondent teachers was that below 20 periods, whereas 6.7% had 31 periods and above. Majority of the respondents had periods ranging from 21-30. The data showed that, 31.6% teachers had low level of satisfaction and 68.4% teachers were in high level of satisfaction in the job. Among the respondents, 35.1% were interested to proceed in teaching profession for the future, 42.5% were not interested to proceed in teaching profession, and 22.4% were undecided.

Table 4.1: Summary of Teachers' Level of Overall Job Satisfaction with Respect to the Categorical Variables (Shashemene, 2011).

Categorical Variables	Categories	Not satisfied		Satisfied		Total	
		N	%	N	%	N	%
Sex of the respondents	Female	32	36%	57	64%	89	22%
	Male	96	30.4%	220	69.6%	316	78%
	Total	128	31.6%	277	68.4%	405	100.0%
Type of school	Non-government	54	25.7%	156	74.3%	210	51.9%
	Governmental	74	38%	121	62%	195	48.1%
	Total	128	31.6%	277	68.4%	405	100.0%
Level of school	Elementary	84	30.6%	191	69.4%	275	67.9%
	secondary	44	33.9%	86	66.1%	130	32.1%
	Total	128	31.6%	277	68.4%	405	100.0%
Educational level	Certificate	12	31.6%	26	68.4%	38	9.4%
	Diploma	59	30.3%	136	69.7%	195	48.2%
	First degree	57	33.3%	114	66.7%	171	42.2%
	Second degree	0	0%	1	100%	1	.2%
	Total	128	31.6%	277	68.4%	405	100.0%

Marital status	Single	59	30.9%	132	69.1%	191	47.2%
	Married	65	31.3%	143	68.7%	208	51.4%
	Divorced	4	66.6%	2	33.4%	6	1.4%
	Total	128	31.6%	277	68.4%	405	100.0%
Subject of teacher	Natural science	44	34.6%	83	65.4%	127	31.4%
	Mathematics	21	27.3%	56	72.7%	77	19%
	Language	32	25.4%	94	74.6%	126	31.1%
	Social science	31	41.3%	44	58.7%	75	18.5%
	Total	128	31.6%	277	68.4%	405	100.0%
Age	20-25years	36	28.3%	91	71.7%	127	31.4%
	26-30years	51	33.8%	100	66.2%	151	37.3%
	31-35years	18	32.7%	37	67.3%	55	13.6%
	36-40years	7	43.8%	9	56.2%	16	3.9%
	41-45years	9	36.0%	16	64.0%	25	6.2%
	46 and above	7	22.6%	24	77.4%	31	7.6%
	Total	128	31.6%	277	68.4%	405	100.0%
Weekly work load	Below 20 periods	21	31.8%	45	68.2%	66	16.3%
	21-25 periods	57	34.1%	110	65.9%	167	41.2%
	26-30 periods	47	32.4%	98	67.6%	145	35.8%
	31 periods & above	3	11.1%	24	88.9%	27	6.7%
	Total	128	31.6%	277	68.4%	405	100.0%
To proceed with the same profession	Yes	32	22.5%	110	77.5%	142	35.1%
	No	73	42.4%	99	57.6%	172	42.5%
	Undecided	23	25.3%	68	74.7%	91	22.4%
	Total	128	31.6%	277	68.4%	405	100.0%

4.4. Results of Bayesian Logistic Regression Analysis

The explanatory variables used in the Bayesian logistic regression were factors retained from the factor analysis and one variable from socio-demographic which showed a significant association with the outcome variable in the Bivariate analysis. For each subject factor scores were obtained as a data, correspondingly for eight extracted common factors of psychometric properties and type of school was dichotomized as government and non-government. The Gibbs sampler was defined with 50,001 iterations in three different chains and 20,000 burn-in terms were discarded. Hence 30,001 iterations in three different chains (90,003 samples) were obtained from the full posterior distribution. Moreover, the eight covariates were found to have statistically significant influence using different prior variances of 10 and 100 from the Bayesian logistic regression of nine candidate variables.

Running the Gibbs sampler more than one chain simultaneously provides time series and autocorrelation plots of each chain in different colors that helps to check convergence. One can say confidently that convergence has been achieved if all the chains appear to be overlapping. The Gillman-Rubin statistic can also be used for checking convergence. Different assessment techniques for convergence are presented in the following section for both different priors.

4.4.1. Convergence Assessment

Time Series Plot

A time series plot is used for diagnosing the convergence of parameter estimates in Bayesian analysis. The WinBUGS package gives the plot by giving the iteration number on the x-axis and parameter value on the y-axis for each significant parameter. For all simulated parameters, the plots of the last 30,001 iterations for three independently generated chains demonstrated good “chain mixture” which was a clear indication of convergence shown in Figure 1 and 2 and in Appendix (Figures A1.1 and A2.1). The time series plots in Figure 1 and 2 and in Appendix (Figures A1.1 and A2.1) showed that the chains with three different colors overlapped. Hence, reasonably convergence was achieved.

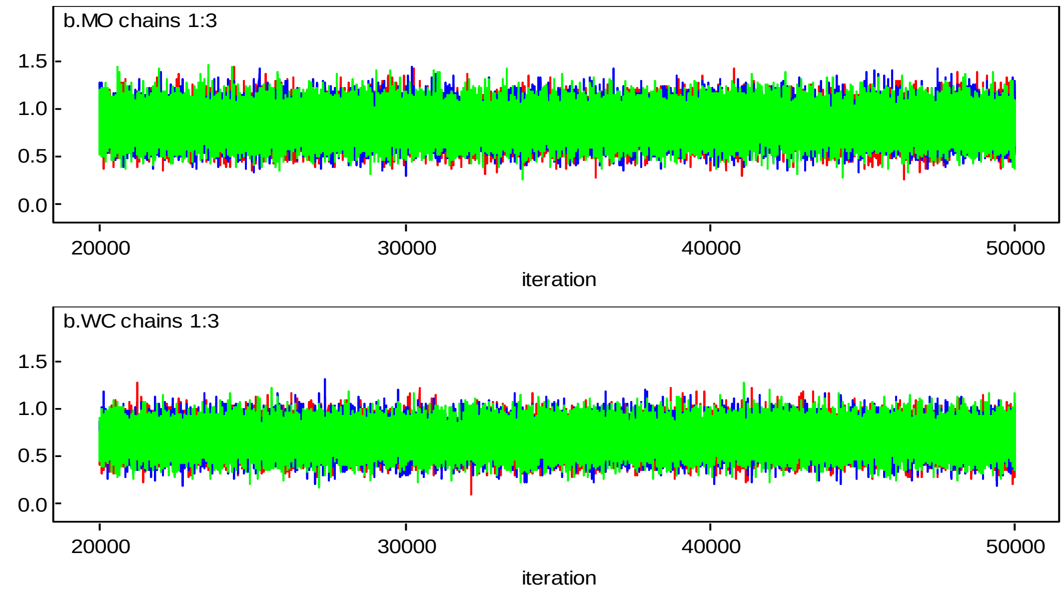

Figure 1: Convergence of Time Series Plots for the Coefficients of Motivation and Working Condition with Prior Variance 100.

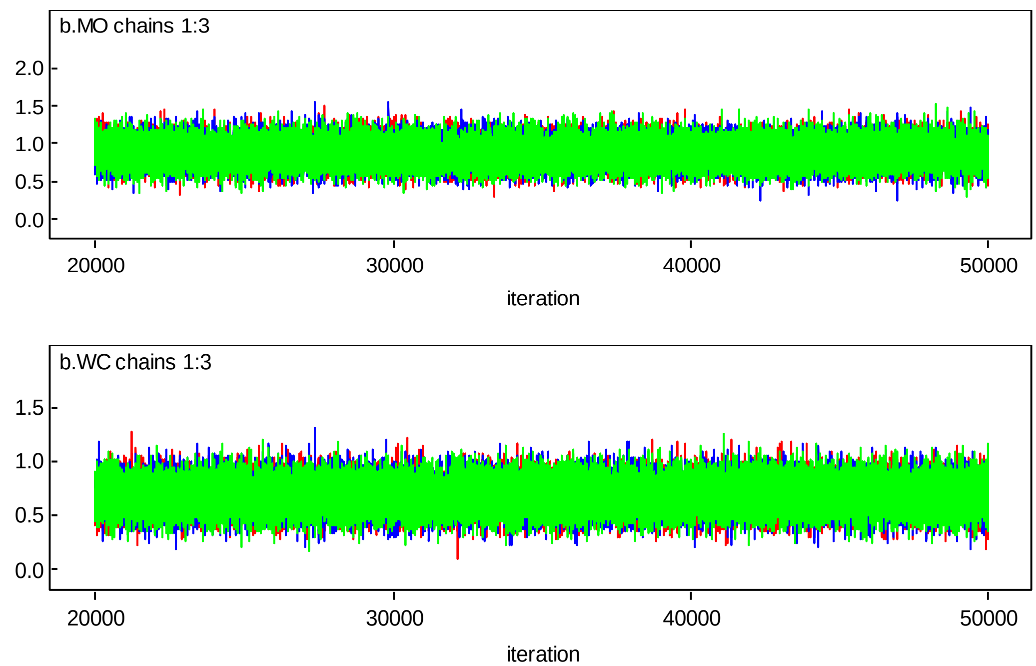

Figure 2: Convergence of Time Series Plots for the Coefficients of Motivation and Working Condition with Prior Variance 10.

Autocorrelation Plot

An autocorrelation plot is another recommended test for convergence in a

Bayesian analysis. For all simulated parameters, the plots of the first 20 to 40 lags of three independently generated chains demonstrated good “chain mixture” which was an indication of convergence of chains in Figure 3 and 4 and in Appendix (Figures A1.2 and A2.2).

The plots in Figure 3 and 4 and in Appendix (Figures A1.2 and A2.2), showed that the three independent chains were mixed or overlapped and died out for higher lags. Hence, this was an evidence of convergence.

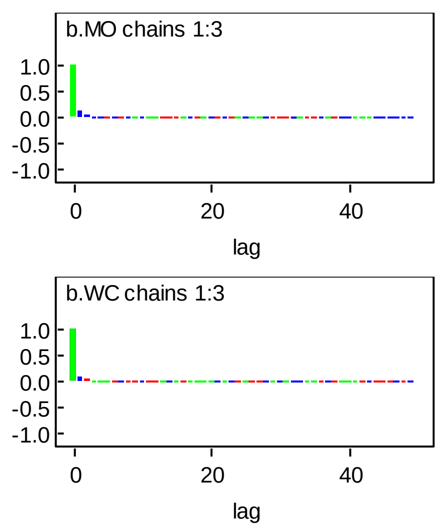

Figure 3: Convergence of Autocorrelation Plots for Coefficients of Motivation and Working Condition with Prior Variance 100.

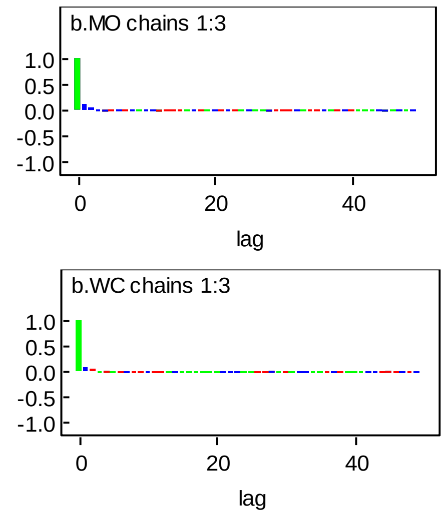

Figure 4: Convergence of Autocorrelation Plots for Coefficients of Motivation and Working Condition with prior variance 10.

Gelman-Rubin Statistic (GR)

In GR test also indicated that there was convergence of the chains in Figure 5 and 6 and in Appendix (Figure A1.3 and A2.3). The green line represents the between-variability, the blue line represents the within-variability, and the red line represents the ratio. In the plots, the red lines were very close to one indicating a clear convergence of chains.

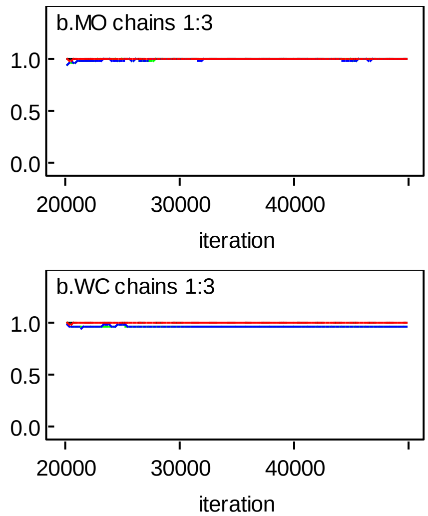

Figure 5: Convergence Using Gelman-Rubin Statistic for Motivation and Working Condition with Prior Variance 100.

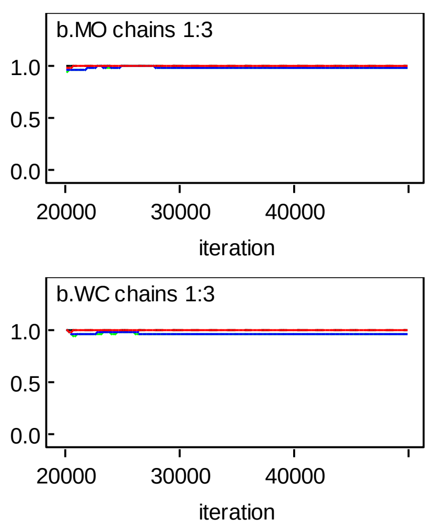

Figure 6: Convergence Using Gelman-Rubin Statistic for Motivation and Working Condition with Prior Variance 10.

Density Plots

This was another recommended technique for identifying convergence. The plots for all covariates indicated that none of the coefficients had bimodal density plot in Figure 7 and 8 and in Appendix (Figures A1.4 and A2.4), ensuring convergence of simulated parameters.

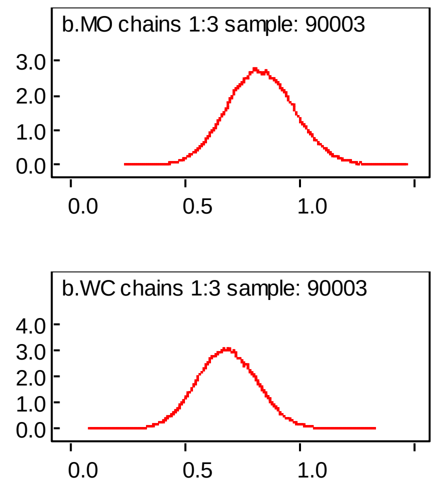

Figure 7: Convergence for Density Plot of the Parameter's for Motivation and Working Condition with Prior Variance 100.

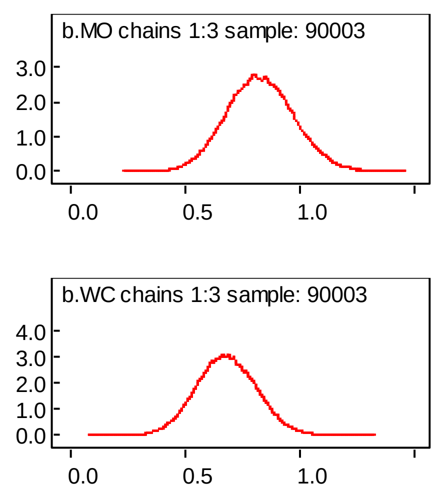

Figure 8: Convergence for Density Plot of the Parameter's for Motivation and Working Condition with Prior Variance 10.

4.4.2. Assessing Accuracy of the Bayesian Logistic Regression Model Fitting

The posterior summary estimates by the MCMC algorithm, especially by Gibbs sampler, like posterior mean, standard error, Monte Carlo error, and 95% confidence intervals were estimated using WinBUGS software. Tables 4.6 and 4.7 contain coefficients (β) and odds ratio (OR) of variables under column node, the estimated coefficients and odds ratio value under column mean, the standard error (SE), Monte Carlo errors and 95% confidence interval (CI).

Gibbs sampler algorithm with WinBUGS software was adopted. There were 50001 iterations in each chain having three different chains and the burn-in terms in each chain has been 20000 iterations and then 30001 samples taken for the analysis in each chain. The total samples taken were 90003 for the three chins.

The convergence of the chain can be initially checked visually using trace plots. Values within parallel zone without strong seasonality indicate convergence of the chain. If the MC error value is low in comparison to its posterior standard error, then the posterior density is estimated with accuracy. Especially, to have accurate posterior estimates, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of its posterior standard error, since small values of MC error indicate that one has calculated the quantity of interest with precision and hence, evidence for accuracy of posterior estimates in Bayesian logistic regression is accomplished. After convergence and accuracy of posterior estimates are attained, summarizing the posterior statistic is possible. In Table 4.6 and 4.7, MC error for each significant predictor is less than 5% of its posterior standard error. This implies convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics. So one can say that the predictor variables given in the Tables 4.6 and 4.7: motivation (MO),working condition (WC), payment condition (PC), school administration (SA), co-workers (CW), curriculum condition (CC), personal life (PL) and student behavior (SB), were statistically significant predictor variables. This is because the 95%

confidence intervals of parameters do not contain zero.

Table 4.2: Parameter Estimation for the Posterior Distribution Obtained from the Sampled Values that Reflect Smooth Kernel Densities with Prior Information of Variance 100 (Shashemene, 2011).

Coefficients &Odds of Variables (Node)	Mean	SE	MC error	95% CI	
				Lower	Upper
alpha	0.9069	0.1909	0.001175*	0.5384	1.288
b.Motivation	0.8316	0.1466	6.046E-4*	0.5523	1.126
OR.Motivation	2.322	0.3451	0.001417	1.737	3.083
b.Working Condition	0.6838	0.1311	4.917E-4*	0.4302	0.9448
OR. Working Condition	1.999	0.2642	9.871E-4	1.538	2.572
b.Payment Condition	0.5998	0.1511	6.082E-4*	0.3101	0.9026
OR. Payment Condition	1.843	0.2823	0.001136	1.364	2.466
b.School Adminstration	0.5932	0.1372	5.25E-4*	0.3276	0.8667
OR.School Adminstration	1.827	0.253	9.767E-4	1.388	2.379
b.Co-Workers	0.5366	0.1335	5.182E-4*	0.2795	0.8042
OR.Co-Workers	1.726	0.2324	9.092E-4	1.322	2.235
b.Curriculum Condition	0.4301	0.1277	4.908E-4*	0.1823	0.6823
OR. Curriculum Condition	1.55	0.1991	7.671E-4	1.2	1.978
b.Personal Life	0.3797	0.1255	4.35E-4*	0.134	0.6282
OR.Personal Life	1.473	0.1861	6.415E-4	1.143	1.874
b.Student Behaviour	0.3445	0.1342	5.271E-4*	0.08388	0.6081
OR. Student Behaviour	1.424	0.1924	7.582E-4	1.088	1.837
b.Type of School	0.2751	0.2697	0.001638	-0.254	0.8048
OR.Type of School	1.366	0.3756	0.002252	0.7757	2.236

Note: *Indicates Covariates which are Statistically Significant in the Bayesian Analysis.

Table 4.3: Parameter Estimation for the Posterior Distribution Obtained from the Sampled Values that Reflect Smooth Kernel Densities with Prior Information of Variance 10 (Shashemene, 2011).

Coefficients &Odds of Variables (Node)	Mean	SE	MC error	95% CI	
				Lower	Upper
alpha	0.9036	0.1901	0.001156*	0.5367	1.282
b.Motivation	0.8286	0.1461	6.01E-4*	0.5501	1.122
OR.Motivation	2.315	0.3429	0.001405	1.734	3.071
b.Working Condition	0.6815	0.1308	4.908E-4*	0.4285	0.9418
OR. Working Condition	1.994	0.2629	9.827E-4	1.535	2.565
b.Payment Condition	0.5974	0.1507	6.026E-4*	0.3083	0.8995
OR. Payment Condition	1.838	0.2807	0.001124	1.361	2.458
b.School Adminstration	0.5908	0.1369	5.236E-4*	0.326	0.8636
OR.School Adminstration	1.823	0.2517	9.72E-4	1.385	2.372
b.Co-Workers	0.5346	0.1332	5.186E-4*	0.2778	0.8017
OR.Co-Workers	1.722	0.2315	9.084E-4	1.32	2.229
b.Curriculum Condition	0.4286	0.1275	4.895E-4*	0.1813	0.6802
OR. Curriculum Condition	1.548	0.1985	7.635E-4	1.199	1.974
b.Personal Life	0.3784	0.1253	4.352E-4*	0.1332	0.6265
OR.Personal Life	1.471	0.1854	6.407E-4	1.143	1.871
b.Student Behaviour	0.343	0.1339	5.249E-4*	0.08291	0.6062
OR. Student Behaviour	1.422	0.1916	7.54E-4	1.086	1.833
b.Type of School	0.2764	0.2685	0.001607	-0.2502	0.8038
OR.Type of School	1.367	0.3741	0.002211	0.7787	2.234

Note: *Indicates Covariates which are Statistically Significant in the Bayesian Analysis.

However, comparing the results of the two Tables 4.6 and 4.7 in (Table 4.8) there were a slight differences in coefficients (b), standard deviation (SD), odds ratio (OR) and credible interval (95%).

Table 4.4: The Model Comparison Based on Standard Error with Different Prior Variances (Shashemene, 2011).

Variables	Prior variance of 100		Prior variance of 10		SE Comparison
	Mean($\hat{\beta}$)	SE_1($\hat{\beta}$)	($\hat{\beta}$)	SE_2($\hat{\beta}$)	
Motivation	0.8316	0.1466	0.8286	0.1461	$SE_2 < SE_1$
Working condition	0.6838	0.1311	0.6815	0.1308	$SE_2 < SE_1$
Payment condition	0.5998	0.1511	0.5974	0.1507	$SE_2 < SE_1$
School administration	0.5932	0.1372	0.5908	0.1369	$SE_2 < SE_1$
Co-workers(colleagues)	0.5366	0.1335	0.5346	0.1332	$SE_2 < SE_1$
Curriculum	0.4301	0.1277	0.4286	0.1275	$SE_2 < SE_1$
Personal life	0.3797	0.1255	0.3784	0.1253	$SE_2 < SE_1$
Student behavior	0.3445	0.1342	0.343	0.1339	$SE_2 < SE_1$

4.5. Interpretation and discussions of the Results

From the Bivariate analysis and Bayesian logistic regression in the present study, the socio-demographic variables have not been found to be significant for the level of job satisfaction. Similar results have also been reported by Sebsibe (2002) and Spector (1996).

In the Bivariate analysis, type of school is a significant variable but in the Bayesian multiple logistic regression, it becomes insignificant which may be due to the influence of other factor variables.

The result of Table 4.6 shows that, the first factor which affects the level of teachers job satisfaction is motivation, the odds ratio indicates that teachers with high motivation are more likely to have high overall satisfaction level (satisfied) keeping other covariates constant and its odds ratio 2.322 shows that for a unit increase in motivation measurement of its factor score, the probability of being satisfied increases 2.322 times keeping other covariates constant. The second factor is working conditions of a teacher which affects the overall level of satisfaction in job and its odds ratio 1.999 shows that for a unit increase in working conditions measurement of its factor score, the probability of being satisfied increases 1.999 times keeping other covariates constant. Motivation and working condition is positively correlated with overall level of job satisfaction since their odds ratios are greater than one. The third significant predictor is payment condition of a teacher which affects the overall level of satisfaction in job and its odds ratio 1.843shows that for a unit increase in payment condition of its factor score, the probability of being satisfied increases 1.843times keeping other covariates constant. The other factors are also interpreted in the same fashion.

In general, all the predictor variables are positively correlated with the outcome variable, since all the parameters are greater than zero and its odds ratio are greater than one.

From Table 4.8 of model comparison, The SE with prior variance 10 is smaller than that with prior variance 100 as expected due to the fact that smaller variance is assumed with information in the priors.

Convergence assessment indicates that the two priors are almost showing the same result with a slight difference. In both cases, the predictor variables which are significant are identified at 5% level of significance. The variables, which are not significant, were also identified in both prior cases. While comparing the two models, where their variances are 10, they are better than the other, because the standard errors (SE) of the parameters, in small variance priors, are less than the other.

5.1 Conclusions

This study has been an attempt to identify and measure overall job satisfaction among school teachers at Shashemene town. Based on the descriptive analysis, it was found that 68.4% of the respondent teachers had high level of satisfaction (satisfied) while 31.6% of them had low level of satisfaction (dissatisfied) in their jobs.

Motivation has the highest value of parameter coefficient indicating that motivation to teach, had the greatest influence on teachers overall job satisfaction than other factors. The other variables like socio economic and demographic variables had no significant influence on teachers' job satisfaction in the current study.

From this research, one can conclude that, the overall level of teachers' job satisfaction is greatly determined by the psychological attitude of a teacher rather than his/her age, sex, educational level, marital status, work load, level of school and type of school.

Finally, it can be concluded that with appropriate choice of prior distributions the application of Bayesian logistic regression model as the posterior distribution of the unknown parameters given data can be conducted using Gibbs sampler algorithm. A more reliable inference can be obtained with the Bayesian approach.

5.2 Recommendations

In order to know the overall level of teachers' job satisfaction, it is better to see the psychological attitudes of the population rather than observing only the socio demographic problem.

Teachers are to be appreciated to have motivation in order to increase their positive feeling towards the profession. School administrations and Ministry of education need to implement interventions to raise teachers' motivation towards high job satisfaction, which in turn raises education quality.

6. REFERENCES

Alan, A. (2004). Bayesian Inference for Categorical Data Analysis, Department of Statistics, University of Florida, Gainesville, Florida, USA 32611-8545.

Albert, J. (2008). Bayesian Computation with R, Second Edition, Springer.

Amina, O., Ilse, V.L., Gabrielle, B., Gogot, S., Aminatou, D. and Sandrine, R. (2010). Teacher Attrition in Sub-Saharan Africa: The Neglected Dimension of the Teacher Supply Challenge.

Anderson, D. (2007). Markov Chain Monte Carlo Methods in Bayesian Computation, New York: Springer-Verlag.

Casella, G. and Berger, R. L. (2002). Statistical Inference. Duxbury Press.

CSA. (2007). Summary and Statistical Report of 2007 Populations and Housing Census, Addis Ababa.

Clark, T. and Griffiths, R. C. (2002). Bayesian Logistic Regression Using a Perfect Phylogeny, Department of Epidemiology and Public Health Imperial College, LondonUK.

Cochran, W.G. (1977). Sampling Techniques. 3rd Edition, John Wiley & Sons Inc., New York.

Education Bureau of Shashemene Town, 2010

Gamerman, D. (2007). Efficient Sampling from the Posterior Distribution in Generalized Linear Models. *Statist Comput. 7:57-68.*

Gill, J. (2002). Bayesian Methods: A Social and Behavioral Science Approach, a CRC Press Company.

Herzberg, F., Mausner, B., Peterson, R.O. and Campbell, D.E. (1957). Job Attitudes: Review of Research and Opinion, Psychological Services of Pittisberg, Pittisberg.

Kim, I. and Loadman, W. (1994). Predicting Teacher Job Satisfaction [ED 383-707].

Merkle, E. Sheu, C. and Trisha, G. (2005). Simulation-Based Bayesian Inference Using WinBUGS. www.mrcbsu.cam.ac.uk/bugs/winbugs/contents.shtml Tutorial.

Mila, A. L., Yang, X. B. and Carriquiry, A. L. (2003). Bayesian Logistic Regression of Soybean Sclerotinia Stem Rot Prevalence in the U.S. North-Central Region.

Porter, L.W. and Steers, R.M. (1973). Organizational Work and Personal Factors.

Sebsibe, Y. (2002). Some Major Determinants of Job Satisfaction among Employees of Oromia Head and Zonal Cooperative Promotion Offices; A.A.U.

Spector, P. E. (1996). Industrial and Organizational Psychology: Research and Practice. New York: John Wiley & Sons, Inc.

UNESCO. (2010). New Projections on the Global Demand for Teachers. UIS.

Walsh, B. (2004). Introduction to Bayesian Analysis. Lecture Notes for EEB 596z.