Statistical analysis of saving habits of employees: a case study at Debre Birhan Town in North Shoa, Ethiopia

Timerga, Genanew and Gotu, Butte and Alem, Yegnanew

Debre Birhan University, Alpha University College/Addis Ababa University, Hawassa University

15 November 2011

Statistical Analysis of Saving Habits of Employees: A Case Study at Debre Birhan Town in North Shoa, Ethiopia

Genanew Timerga [1] (M.Sc), Butte Gotu (PhD) [2], Yegnanew Alem [3] (M.Sc)

[1]Debre Birhan University, Po. Box 445, Tel: +251913054542, tichdar@gmail.com

[2]Alpha University College/Addis Ababa University, Tel: +251911405942, buttegotu@gmail.com

[3]Hawassa University, Po. Box 05, Tel: +251911380766 yegna2009@gmail.com

ABSTRACT

Saving represents one of the most predictable determinants of successful personal and economic development. People desire to save although they tend to postpone saving until they have higher-paying jobs or some stability in their lives. However, in developing countries, where opportunities for structured and institutionalized saving are rare, people could perhaps begin saving earlier than expected. The purpose of this study has been to assess saving habits and identify factors that influence the saving habits of employees at Deber Birhan town. A sample of 480 was collected from employees on saving habit at Debre Birhan town during February, 2010 to October, 2011. Saving habit was measured according to multi method tool that incorporates self report, visual analog scale and economic identification test. Descriptive, Binary logistic regression and Bayesian statistical methods were used. The result indicates that 47.29% employees had no saving experience and 52.71% of the respondents have been involved in saving part of their income. It was also found that government employees have lower saving habits than the private employees. The results obtained from the analysis of binary logistic regression indicate that age, education, number of dependent family members, transport service, job satisfaction in the sector, cost of expenditures and inflation significantly affect the saving habits of employees. Being a member of saving association, cost of recreation and housing are also significantly related with saving habits of employees. Results from binary logistic regression indicate that after controlling other variables in the model, the odds of saving decreases for instance by 43.4% for one unit increases in number of dependent family members. Employees who had job satisfaction in the sector were less likely to be in no saving habit with odds 2.491. The result from Bayesian analysis indicates that monthly salary, distance from home to work place and supporting others with money were significant predictors of saving habits. Furthermore, education, number of dependent family members, cost of expenditures and monthly salary are important factors affecting saving habits of employees.

***Key words**: Bayesian Logistic Analysis, Binary Logistic Regression, Saving Habit,*

CHAPTER ONE: INTRODUCTION

1.1. Background of the Study

Personal saving has two primary functions. First, savings provide the economic security of a safety net. By transferring resources from the present to the future via savings, individuals are prepared to face unexpected and irregular financial circumstances. Second, saving leads to accumulation of wealth that enables individuals to improve their living standard and to respond to new opportunities (Gokhale, 2000). Everyone agrees that starting to save early has merit in it and "Money grows on the tree of patience" and there are benefits of “power of compounding”, but few actually practice it.

When it comes to saving, people in general and the poor in particular might not be completely rational and completely knowledgeable (Karim, 2010).The goal of promoting financial saving habit is to make people more aware of financial opportunities, choices, and possible consequences. There is a growing recognition of the importance of financial education as it relates to saving (Greenwald et al., 2001; Gill, 2004). Financial education is one way of increasing savings and asset accumulation. Sherraden et al., (2007) say that the extent to which an individual understands the process and benefits of asset accumulation is likely to affect their willingness to save.

As few studies are available on saving in Ethiopia, this study intends to fill this gap with a focus on the ANRS. The regional capital of ANRS, Bahir dar, is the only really big urban center in the region, and since the last dozen years, it has grown substantially in size and economic activity. Assessing employees’ saving habits has been crucial, which is as good as null at present, and hence Debre Birhan Town has been selected for the present study. It is, therefore, important to conduct researches on employees’ saving habits.

Furthermore, the statistical models which the researcher mainly concentrated in this study are the logistic regression model and Bayesian logistic regression model to identify significant factors associated to no-saving related to employees.

1.2 Statement of the Problem

Most of the developing countries have low rate of saving habits, so that improving saving habit is a primary goal for people living in this part of the world (Michael et al., 2001). Improving saving habit of individuals is given attention to look at a variety of savings services used by people/customer in the community. Improving this depends not only on attending of government provider but also on appropriate execution of recommended components of saving in household of the community and institutions or sectors.

However, most of the studies focus on descriptive statistics produce hard-to-generalize on a wider perspective. So that the following important issues is identified.

- Saving habit and institutions of saving area have become increasingly important for safety nets of the employees.
- Investigation of factors associated with saving habits assumes critical importance.
- Provision of saving services varies widely across individuals in Ethiopia.
- Understanding these variations at the individual and in the community level in order to successfully implement any saving related policies and programs.

Thus, the purpose of this study is to assess the current status of saving habits services, and factors that influence utilization of these services, and to find out the possible reasons for underutilization of saving habit services using primary data collected from Debre Birhan, one of the cities in ANRS, Ethiopia.

1.3 Objectives of the Study

1.3.1 General Objective:

To assess saving habits and identify factors that influence the saving habits of employees at Deber Birhan town.

1.3.2 Specific Objectives:

- To assess the current status of saving practices of the employees at Deber Birhan town;
- To identify factors that influence the saving habits of employees;
- To evaluate the capacity of saving institutions that employees they used;
- To provide or document information for decision makers, planners and researchers.

1.4 Significance of the Study

Assessing saving habits is very important to understand and evaluate the achievement gained in the implementation of saving programmers as a source of information for the employees about the benefit of saving in improving employee life. It is hoped that results of this study will improve policy makers', planners' and researchers' understanding of the determinants of saving habits of employees in the study area and may serve as an important tool for any possible information towards improving saving habits.

This research is essential as it contributes to the efforts of the country in improving saving habits of employees.

CHAPTER TWO: LITERATURE REVIEW

2.1 Literature Review in Ethiopia

The study of saving has a contribution to change personal behavior and economical growth in the country. Dedebit Credit and Saving Institution (DECSI) and Amhara Credit and Saving Institutions (ACSI) take more than 65% share in serving clients in the market (Befekadu, 2007). Higher Gross Domestic Savings (GDS) has improved the saving income function of employees. The major determinants of GDS are demographic aspect, interest rate, tax, broad money, stage and duration of disease, association of infectious disease, export, etc. The result of his study shows export has high value to saving habit; tax has a significant positive effect when it regresses on private saving.

The impact assessment of micro finance records positive results of saving habits in the country by different researchers. Mengistu (1997) shows that increase in the number of program beneficiaries is one indicator of the contribution of assistance program to employment creation and income generation.

Kassa (1998) shows in South Ethiopia in the area of education, consumption expenditure, medical expenditure, family assistance, employment creation, income generation are very important to saving habits. In addition, Kassa (1998) reported a 30.82%, 10.5% and 19.7% annual growth in consumption expenditure of the first, second and third credit cycle beneficiaries, respectively. About 7%, 8.03% and 2.07% growth in first, second and third saving cycle beneficiaries' case, respectively. Using Wilcoxon grouped pair's nonparametric test, he did not accept the null hypothesis of average income before and after saving are the same at 5% level of significance implying that the average income after saving is greater than before saving in the first, second and third credit cycle.

2.2 Saving Institutes

At this time, there are different types of saving system in the world. This is categorized in to three ways, and these are formal saving sector (saving and credit co-operatives (SACCOS), bank and insurance companies), semi-formal saving sector (microfinance institutions (MFI)) and informal saving sector (save at home, save at clubs, deposit collector, reciprocal lending, rotating savings and credit association (ROSCA), accumulation savings and credit association (ASCAS) etc) (Ziorklui and Barbee, 2003). Informal savings have different names in different countries.

In Ethiopia, the effective financial structure has 95% of the productive asset which is composed of 70-80% loan and 10-20% liquid investment and the remaining 5% is unproductive assets composed of land, buildings and equipments. On the other hand, 70-80% of credit union liability should be composed of members' savings to achieve financial independence. In order to finance non-performing assets, improve earnings and absorb losses, members share capital and institutional capital should be greater or equal to 20% and to 10% of total asset respectively. Rate of return and costs operating expense to total assets ratio is set to be less than 10% and other return and costs to be greater or equal to market rate. However, administrative cost should not be greater than 5% of the average total assets (Gebeyaw, 2008).

In our country, there are different traditional financial systems which have long history and paramount contribution to economic betterment and social wellbeing of the society. Traditional institutions organized with a sense of cooperation and risk sharing has enabled Ethiopians to experience saving and financial management within its cultural context. "Eqqub", "Eddir", "Mahiber" etc are some of the informal financial institutions that shaped the social bond and interaction (Gebeyaw, 2008).

Characteristics of Saving Associations

The most important purpose of saving institutions is to make mortgage loan on residential property. These organizations are known as saving associations, building loan associations and cooperative banks. As home-financing institutions, they give primary attention to single-family residences and are equipped to make saving in this area. Some of the most important characteristics of a savings association are:

1. It is generally a locally owned and privately managed home financing institution.
2. It receives individuals' savings and uses these funds to make long-term amortized loans to home purchasers.
3. It makes save for the construction, purchase, repair, or refinancing of houses.
4. It is state or federally chartered.

Ugandan, survey 87% of respondents said that a service delivery point in their sub-district would satisfy their needs as savings clients (Pelrine et al., 2005). As a result, informal and semi-formal savings mechanisms can out-compete formal sources on proximity, helping explain why relatively risky informal mechanisms are still the most widely used by low-income Ugandans. And also distance of financial sector affects saving rate of households. Return on Equity (ROE) of NBE WSACA was 37%, 32% and 30% in 2005, 2006 and 2007 respectively, which was closer to ROE of commercial banks (32.7%, 40.7%and 29% in the respective years). ROE of MFI industry, on the other hand was 55%, 45% and 50% during the respective years, which was significantly higher than the later year (NBE WSACA, 2008). Similarly, in rural Ethiopia, most of the people use informal saving mechanisms because they are far away from the financial institutes or sectors.

CHAPTER THREE: METHODOLOGY

3.1 Description of the Study Area

This study was conducted at Debre Birhan town, the capital of North Shoa or "Semen Shewa" zone from February to October, 2011. According to CSA (2008) report, the estimated population size of Debre Birhan town was 65,244 out of these 31,658 were male and 33,556 were female.

3.2 Study Population

The target population consists of all employees at Debre Birhan town. A total of 5113 employees within the age interval of 18 to 75 years were considered.

3.3 Sampling Design and Techniques

The data which is cross-sectional were collected from the target population through a structured questionnaire.

3.3.1. Sampling

Stratified random sampling is used where the researcher draws simple random samples from certain aggregation units of interest (from certain census block groups) when the strata are heterogonous geographic units and also from successively smaller aggregations until the individual subject level is reached.

In this study using stratified random sampling method, government and private employees were selected with probability proportional to size without replacement (ppswor) as primary sample units (strata), individual employee (secondary unit samples) from each selected stratum with systematic sampling. Strata 1 : All employee that are work in the government organization. Strata 2: All employee that are work in the private organization. The sampling frame was prepared from government and private employees by giving sequential order number. Finally an employee selected by systematic random sampling procedure was contacted.

3.3.2. Sample Size Determination

The sample size for this study was determined based on stratified sampling for proportions of 95 percent confidence level. Moreover, 3 percent of the sample size is added to compensate for non-response rate. The sample size formula is given by (Cochran, 1977)

$$n = \frac{\sum_{h=1}^{2}\left({W_h}^2{S_h}^2 / w_h\right)}{V + \frac{\sum_{h=1}^{2} W_h {S_h}^2}{N}} \quad \text{………………………………………… 3.1}$$

where $V = \left(\frac{d}{Z\alpha_{/2}}\right)^2$, is the specified variance of the estimate and in case of estimating proportion, the stratum variance is given by ${S_h}^2 = \frac{N_h}{N_h - 1} P_h Q_h = {S_h}^2 = \frac{N_h}{N_h - 1} PQ$ for common p. Since we use proportional allocation sample, $w_h = n_h/n = N_h/N$, w_h stands for stratum weight, N=N_1 + N_2

$$n = \frac{\sum_{h=1}^{2}\left({W_h}^2 N_h PQ / W_h(N_h - 1)\right)}{V + \frac{\sum_{h=1}^{2}\left(\frac{W_h N_h PQ}{N_h - 1}\right)}{N}} \quad \text{……………………… (3.2)}$$

where Z be the upper $\alpha/_2$ point of standard normal distribution, where $\alpha =$ 0.05 significance level, which is $Z_{\alpha/_2} = Z_{0.025} = 1.96$. Suppose the relative error d is usually set by the investigator desired which is used from a similar study undertaken by the "Saving habits, needs and priorities in urban Uganda" in 2005, with a sample size of 852 is taken as a reference for the purpose of fixing the standard deviation. One of the explanatory variables used in that study is customer saving with a standard deviation of 0.45. This helps us to determine the sample size to represent the population by calculating the acceptable absolute error, d, where $d = Z\alpha_{/2} S/\sqrt{n}$. If we adopt a significance level α = 0.05, then the calculated margin of error from the above information is 0.03. Previous studies indicate that the proportion of saving habits of employees varies from study to study. In our study, it is indicated that the proportion of saving habits of employees p was determined small proportion from the results of previous work of similar population of saving habits of employees in Chinese and Americans 14.02% (p = 0.1402) of (Feifei, 2010). Thus, the sample size calculation is

$$n = \frac{\sum_{h=1}^{2}\left({W_h}^2 N_h PQ / W_h(N_h - 1)\right)}{V + \frac{\sum_{h=1}^{2}\left(\frac{W_h N_h PQ}{N_h - 1}\right)}{N}} = \frac{0.12059}{0.00023 + 0.000024} = \frac{0.12059}{0.000254} = 474.75 \approx 475$$

Finally, 3 percent of the sample size, which is 15, was added to the determined sample size 475 to compensate for none response rate. Thus, the required sample size for this study is 490 employees from 5113 who live at Debre Birhan town. Next, we carried out sample size allocation to each stratum with proportional allocation.

$$k_h = \left[\frac{nN_h}{N}\right] = nW_h \ ,h = 1,2 \qquad \dots\dots\dots\dots\dots\dots\dots\dots (3.3)$$

Where, h = types of employee = 2

$$n = \sum_{i=1}^{2} k_h = 474.76 \approx 475 + 15 = 490$$

$W_h = {}^{N_h}/_{N}$ = Probability of stratum weight selection of h^{th} employee, N_h = Total number of employees within the h^{th} strata, N = Total number of employee in DB town

Table 3.1 Number of Employee Taken from the Selected PSU at Debre Birhan

Employee	N_h	W_h	Sample
Government (N_1)	3503	0.685	336
Private (N_2)	1610	0.315	154
Total	5113		490

3.4 The Study Variables

3.4.1. Dependent Variable

The response variable in this study is the status of saving habits at Debre Birhan town. The habits of employees are identified either save out of income or no save out of income. The response variable is a dichotomous category and my interest of the study is no saving out of income. Thus, coded as the value 0 for 'save out of income' and 1 for 'no save out of income'.

3.4.2. Independent Variables

The independent variables in this study are classified as occupational variables, economic variables and personal relationship and contextual variables.

3.5 Logistic Regression Model

Logistic regression analysis extends the techniques of multiple regression analysis to research situations in which the outcome variable is categorical.

Generally, the response variable is binary, such as (save or no save, presence or absence, success or failure etc) in logistic regression.

3.5.1. Assumptions of Logistic Regression

Assumptions were should consider for the efficient use of logistic regression as given below (Hosmer and Lemeshow, 1989).

- Logistic regression assumes meaningful coding of the variables. Logistic coefficients were difficult to interpret if not coded meaningfully. The convention for binomial logistic regression is to code the dependent class of interest as 1 and the other class as 0.
- Logistic regression does not assume a linear relationship between the dependent and independent variables.
- The dependent variable must be categorical.
- The independent variables need not be interval, no normally distributed, no linearly related and no equal variance within each group.
- The groups must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups.
- Larger samples are needed than for linear regression because maximum likelihood coefficients are large sample estimates.
- The logit regression equation should have a linear relationship with the logit form of the dependent variable.
- Absence of multicollinearity.

3.5.2. Model Description

Since the response variable in logistic regression is usually dichotomous, we were define such a response variable as Y, and denote the even y=1, when the subject has the characteristic of interest and y=0, when the subject does not have that characteristic of interest. So an alternative form of the logistic regression equation is the logit transformation of P_i given as

$$\text{logit}\left[P_i\right] = \log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} \ldots\ldots\ldots\ldots (3.5)$$

The coefficient can be interpreted as the change in the log-odds associated with a one unit change in the corresponding independent variable or the odd increases multiplicatively by e^{β_i} for every one unit change increase in X_i $i = 1,2, \ldots, k$.

3.5.3. Parameter Estimation for Logistic Regression

The maximum likelihood and non-iterative weighted least squares are the two most computing estimation methods used in fitting logistic regression model (Hosmer and Lemeshow, 1989).

Consider the logistic model $P(X_i) = \frac{e^{X\beta}}{1+e^{X\beta}}$, since observed values of Y say, Y_i's (i=1, 2 …, n) are independently distributed as binomial and, the maximum likelihood function of Y is given by:

$$L(\beta|y) = \prod_{i=1}^{n} p(y_i|X_{i1}, X_{i2}, \ldots, X_{ik}) = \prod_{i=1}^{n} \left[\frac{e^{X\beta}}{1+e^{X\beta}}\right]^{y_i} \left[\frac{1}{1+e^{-X\beta}}\right]^{(1-y_i)} \ldots\ldots\ldots\ldots (3.6)$$

3.6.1. Model Selection

In model selection, there are two competing goals: on one hand the model should be complex enough to fit the data well. On the other hand, it should be simple to interpret, smoothing rather than over fitting the data (Agresti, 2002).

3.6.2. Goodness of Fit of the Model

Assessing goodness of fit involves investigating how close values predicted by the model with that of observed values (Bewick and Jonathan, 2005). The comparison of observed to predicted values using the likelihood function is based on the statistic called deviance.

$$D = -2\sum_{i=1}^{n} \left[y_i \ln\left(\frac{\hat{p}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{p}_i}{1-y_i}\right)\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.7)$$

For purposes of assessing the significance of an independent variable, the value of D are compared with and without the independent variable in the equation as given below:

χ^2= D (model without the variable) − D (model with the variable)

The goodness-of-fit χ^2 process evaluates predictors that are eliminated from the full model, or predictors (and their interactions) that are added to a smaller model.

3.6.3. Likelihood-Ratio Test

The likelihood ratio test statistic (G^2) is the test statistic commonly used for assessing the overall fit of the logistic regression model. The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the full model (L_1) over the maximized value of the likelihood function for the simpler model (L_0). The likelihood-ratio test statistic is given:

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1) = -2(L_0 - L_1)] \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.8)$$

It is compared with a χ^2 distribution with 1 degree of freedom. This log transformation of the likelihood functions yields a chi-squared statistic.

3.6.4. The Hosmer and Lemeshow Test Statistic

The final measure of model fit is the Hosmer and Lemeshow goodness-of-fit statistic, which measures the correspondence between the actual and predicted values of the dependent variable.

$$\hat{C} = \sum_{k=1}^{g}\left({(O_k \quad O'_k\bar{P}_k)^2} \Big/ {n'_k\bar{P}_k(1-\bar{P}_k)} \right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.9)$$

Where g is the number of groups, n'_k is the number of covariate patterns in the kth group, $O_k = \sum_{j=1}^{n'_k} y_j$ is the number of responses among the n'_k covariate patterns, and $\bar{P}_k = m_j\bar{P}_j/n'_k$ is the average estimated probability.

3.6.5. The Wald Statistic

The Wald statistic is an alternative test, which is commonly used to test the significance of individual logistic regression coefficients for each independent variable (that is to test the null hypothesis in logistic regression model that a particular logit coefficient is zero).

If the Wald test is not significant, then these explanatory variables can be omitted from the model. Wald $\chi 2$ statistic was used to test the significance of individual coefficients in the model and is calculated as:

$$Z = \frac{\hat{\beta}_j}{SE(\beta_j)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.10)$$

Each Wald statistic is compared with a χ^2distribution with 1 degree of freedom. Wald statistic is easy to calculate but their reliability is questionable, particularly for small samples.

3.6.6. R^2 Statistic

A number of measures have been proposed in logistic regression as an analog to R^2 in multiple linear regressions. The Cox and Snell measure is based log-likelihoods and considers sample size. The maximum value that the Cox and Snell R^2 attain is less than 1. The Nagelkerke R^2 is an adjusted version of the Cox and Snell R^2 and covers the full range from 0 to 1, it is often preferred.

Therefore, in this study R^2 statistic to indicate how useful the explanatory variables are in predicting the response variables were used (Bewick and Jonathan, 2005).

$$R_{cs}^{2} = 1 - \exp\left[-\frac{2}{n}\left[\mathrm{D(model\ without\ the\ variable)} - \mathrm{D(model\ with\ the\ variable)}\right]\right]$$

The Nagelkerke measure is as follows: $R_N{}^2 = \frac{R_{cs}{}^2}{R_{MAX}{}^2}$,

Where $R_{MAX}{}^2 - 1 - \exp[2(n)^{-1}\mathrm{D(model\ without\ the\ variable)}]$

3.7 Bayesian Analysis

3.7.1. Bayesian Logistic Regression

The foundation of Bayesian statistics is the Bayes' theorem which states that if A and B, are events and $P(B)$, the probability of event B, is greater than zero, then:

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

3.7.2. The Likelihood and Prior Function

Likelihood Function

The likelihood function of the Bayesian formulation for the joint distribution of n independent Bernoulli trials was still the product of each Bernoulli distribution, the sum of independent and identically distributed Bernoulli trials in which the sum has a Binomial distribution. Specifically, let $y_1, y_2 \ldots\ldots y_n$ be independent Bernoulli trials with success probabilities $P_1, P_2, P_3, \ldots\ldots, P_n$, that is $y_i = 1$ with probability P_i or $y_i=0$ with probability $1 - P_i$, for $i = 1,2,\ldots,n$. The trials are independent, the joint distribution of $y_1, y_2, \ldots y_n$ is the product of n Bernoulli probabilities. As usual, the likelihood, function used by Bayesians matches that from frequents inference. binomial:

$$L(y \mid \beta) = \prod_{i=1}^{n} \left[P_i^{y_i} (1 - P_i)^{(1-y_i)} \right]$$

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 $y_i=0$ of the event for that subject.

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

where: P_i = the probability of i^{th} employees being save, since individual subjects are assumed independent from each other likelihoods function over a data set of subjects is:

$$L(y \mid \beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta_o + \beta_1 x_{i1} + \dots \beta_k x_{ik}}}{1 + e^{\beta_o + \beta_1 x_{i1} + \dots \beta_k x_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_o + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_o + \beta_1 x_{i1} + \dots \beta_k x_{ik}}} \right)^{(1-y_i)}$$

Prior Function

The choice of an informative prior distribution typically involves a certain amount of subjectivity; historically, this has been a reason for disagreement between Bayesian and classical statisticians. The assumed prior normal distribution for parameter β_j is given by

$$f(\beta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \left(\frac{\beta_j - \mu_j}{\sigma_j} \right)^2 \right) \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.11)$$

3.7.3. The Posterior Distribution

Based on the prior distribution given above, the posterior distribution of the Bayesian logistic regression contains all the available knowledge about the parameters in the model like

$$f(\beta|y) = \prod_{i}^{p} \frac{p_i^{y_i} (1 - p_i)^{1-y_i} p(\beta)}{p(x_1, x_2, \dots, x_p)} \propto \prod_{i}^{p} p_i^{y_i} (1 - p_i)^{1-y_i} p(\beta)$$

$$f(\beta|y) = \prod_{i=1}^{p}\left[\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip}}}{1+e^{\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip}}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip}}}{1+e^{\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip}}}\right)^{1-y_i}\right]$$
$$\times\prod_{i=0}^{p}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{1}{2}\left(\frac{\beta_i-\mu_i}{\sigma_i}\right)^2\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.12)}$$

Where $f(\beta|y)$is the posterior distribution which is the product of the logistic regression likelihood and the normal prior distributions for the β parameters.

Note that the posterior distribution is a conditional distribution, conditional observing the sample. The posterior distribution is now used to make statements about β, which is still a random quantity. The most popular method of simulation technique is Markov Chain Monte Carlo (MCMC) methods.

3.7.4. Markov Chain Monte Carlo (MCMC) Methods

The use of Markov chain Monte Carlo (MCMC) methods to evaluate integral quantities has exploded over the last fifteen years. The primary distinction made here is between standard Monte Carlo simulation and the Markov chain type of Monte Carlo methods. The initial definition required is that of a more primitive concept that underlies for the second MC which is called Markov chains.

3.7.5. The Gibbs Sampler Algorithm

The Gibbs sampler (David, 2006) 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 $\pi(\beta)$ where β is an 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 β and defined by π (β) = $\pi(\beta\,|\,\beta_i)$ for i = 1, 2… k, where the notation β_i indicates a specific parametric form from β without the β_i coefficient. These requirement facilities the iterative nature of the Gibbs sampling algorithm described as:

1. Start with an initial value: $\beta^{[0]} = \beta^{[0]}{}_0, \beta^{[0]}{}_1, \beta^{[0]}{}_2, \ldots, \beta^{[0]}{}_k$
2. Sample for each i = 0, 1, 2... , n-1:

Generate $\beta_0^{(i+1)}$ from $f\left(\beta_0 \mid \beta_1^{(i)}, \beta_2^{(i)}, \beta_3^{(i)}, \ldots, \beta_k^{(i)}\right)$

Generate $\beta_1^{(i+1)}$ from $f\left(\beta_1 \mid \beta_0^{(i+1)}, \beta_1^{(i)}, \beta_2^{(i)}, \ldots, \beta_k^{(i)}\right)$

Generate $\beta_2^{(i+1)}$ from $f\left(\beta_2 \mid \beta_0^{(i+1)}, \beta_1^{(i+1)}, \beta_2^{(i)}, \ldots, \beta^{(i)}{}_k\right)$

$\vdots$

Generate $\beta_k^{(i+1)}$ from $f\left(\beta_k \mid \beta_0^{(i+1)}, \beta_1^{(i+1)}, \beta_2^{(i+1)}, \ldots, \beta_{k-1}^{(i+1)}\right)$......... .(3.16)

3. Return $\beta^{(1)}, \ldots\ldots\ldots\ldots, \beta^{(n)}$

Once convergence is reached, all simulation values are from the target posterior distribution and a sufficient number should then be drawn so that all areas of the posterior are explored.

3.7.6. Assessment of Convergence

There are basically three approaches to determining convergence for Markov chains: assessing the theoretical and mathematical properties of particular Markov chain, diagnosing summary statistics from in-progress models, and avoiding the issue altogether with perfect sampling, which uses the idea of "coupling from the past" to produce a sample from the exact stationary distribution (Congdon, 2005). The second convergence assessment methodology involves monitoring the performance of the chain as part of the estimation process and making an often subjective determination about when to stop the chain. Among several ways the most popular and straight forward convergence assessment methods are:

1. **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. These can be accessed in WinBUGS using the "autcor" button on the Sample Monitor Tool.
2. **Time series plots**: Iteration numbers on x-axis and parameter value on y-axis are commonly used to assess convergence. If the plot looks like a horizontal band, with no long upward or downward trends, then we have evidence that the chain has converged. These can be accessed in WinBUGS by using the "history" button**.**

3. **Gelman-Rubin statistic**: for a given parameter, this statistic assesses the variability within parallel chains as compared to variability between parallel chains. 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 between variability, the blue line represents 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.
4. **Density plot:** a classic sign of non convergence is multimodality of the density estimate.

3.7.7. The Burn-in Period

Burn-in is a colloquial term that describes the practice of throwing away some iteration at the beginning of an MCMC run. A sufficient period, the chain approaches its stationary distribution, from which we throw away all the data. This is the burn-in period. After the burn-in we run normally, using each iterate in our MCMC calculations or posterior summary statistics. To showing evidence for poor mixing, time series trace will be seen for a minimum burn-in period for some starting value (Merkle et al., 2005).

3.7.8. Assessing Accuracy of the Bayesian Logistic Regression

Once we are happy that convergence has been achieved, we were needed to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference. 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 we are using as our 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% of the sample standard deviation. The Monte Carlo error (MC error) and sample standard deviation (SD) are reported in the summary statistics table of WinBUGS statistical package.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Descriptive Statistics

The data comprised a sample of 480 employees, who were working in government and private sectors in Debre Birhan town. The data collection period was April to June, 2011. The response variable considered in this study was the saving habits of the employees (either save or not save).

As can be seen in Table 4.1, out of 480 employees considered in the analysis, 47.29% employees have no saving habits and 52.71% have saving habits at the time of data collection. Of the total sample, 31.5% of the employees were in private or non government organization whereas 68.5% were in government organization. With regard to the sex composition, 43.7% were female and 56.3% were male employees. The age distribution indicates that 26% of the respondents were in the age category below 25 years, 37.4% in 25-35 years, 20.8% in 36-45 and 15.8% were above 45 years old. In case of marital status, 44%, 52.1%, 3.8% and 0.1% of the respondents were single, married, separated and widowed respectively. It was observed that among the employees the percentage of Orthodox, Muslim, Protestant and others religion followers were 83.1%, 4.2%, 9.4% and 3.3% respectively. Considering income, about 17.1 % of the respondents reported their income per month during the survey period was below 1000 birr, and 18.1% of the respondents were above 3000. With regard to education, 11.5%, 37.3%, 41.8% and 9.4% of the respondents were certificate & below, diploma, first degree master and above degree holders respectively.

From cross-tabulation, it can be seen that 61.9% among male and 41% among female employees exercised saving. This indicates that proportion of male employees who practice saving is higher than that of female employees.

Table 4.1: Results of Demographic Characteristics (Debre Birhan Town, 2011)

Explanatory Variables	Category	Current Status of Saving Habits at DB in 2011				Total	
		No Saving habit		Saveing habit			
		Count	%	Count	%	Count	%
Sex of the respondents	Female	124	59.0	86	41	210	43.7
	Male	103	38.1	167	61.9	270	56.3
Age of the respondents	Below 25	74	59.2	51	40.8	125	26.0
	25—35	87	48.6	92	51.4	179	37.4
	36—45	41	41	59	59	100	20.8
	Above 45	25	32.9	51	67.1	76	15.8
Education Level	Certificate & below	37	67.3	18	32.7	55	11.5
	Diploma	105	58.7	74	41.3	179	37.3
	1st degree	78	38.8	123	61.2	201	41.8
	Masters and above	7	15.6	38	84.4	45	9.4
Religion	Orthodox	189	47.4	210	52.6	399	83.1
	Muslim	8	40	12	60	20	4.2
	Protestant	23	51.1	22	8.7	48.9	9.4
	Other	7	43.8	9	56.2	16	3.3
Ethnicity	Oromo	17	36.2	30	63.8	47	9.8
	Amhara	185	50.4	182	49.6	367	76.4
	Tigre	11	36.7	19	63.3	30	6.3
	Gurage	6	46.2	7	53.8	13	2.7
	Other	8	34.8	15	65.2	23	4.8
Monthly salary	Below 1000	53	64.6	29	35.4	82	17.1
	1000—2000	103	52	95	48	198	41.3
	2001—3000	48	42.5	65	57.5	113	23.5
	Above 3000	23	26.4	64	73.6	87	18.1
Marital status	Single	106	50.2	105	49.8	211	44.0
	Married	109	43.6	141	56.4	250	52.1
	Separated	12	66.7	6	33.3	18	3.8
	Widowed	0	0.0	1	100	1	0.1
Employee	Government	167	50.8	162	49.2	329	68.5
	Private	60	39.7	91	60.3	151	31.5

Table 4.2 shows 84.1% of the respondents had no saving experience due to low income in the job whereas 15.9% of the respondents gave the following reasons for not saving: distance of financial institution, low interest rate earned and no benefit of saving.

Table 4.2: Employees with no Saving Habit by Various Reasons

Reasons for not saving	Category	Number of employees	
		Count	%
Low income in the job	No	36	15.9
	Yes	191	84.1
	Total	227	100.0
Far location of financial institution	No	202	89.0
	Yes	25	11.0
	Total	227	100.0
Low interest rate earned	No	121	53.3
	Yes	106	46.7
	Total	227	100.0
Thinking no benefit of saving	No	175	77.1
	Yes	52	22.9
	Total	227	100.0

From those individuals who have saving habits, 86.2% of the respondents have used formal method of saving institutes whereas 75.1% of the respondents have used informal method of saving institutes. This show that formal method of saving institutes is slightly higher than informal saving methods of saving institutes (Table 4.3).

Table 4.3: Results of Saving institutes at Debre Birhan Town in 2011

Methods of saving institutes	Category	Number of employees	
		Count	%
Formal methods of saving institutes	No	35	13.8
	Yes	218	86.2
	Total	253	100.0
Semi-formal methods of saving institutes	No	151	59.7
	Yes	102	40.3
	Total	253	100.0
Informal methods of saving institutes	No	63	24.9
	Yes	190	75.1

Total	253	100.0

When we compare government and private employees saving habits, 73.6% of government employees had no saving experience whereas 26.4% of private employees have no saving habits. Similarly saving habits in government employees was 64% while 36% in private employees. Therefore, the proportionality of private employees to save is more than government employees (Table 4.4).

Table 4.4: Results of Employees' Saving habits of money at Debre Birhan in 2011

Employer	No saving habit		Saving habit		Total	
	Count	%	Count	%	Count	%
Government	167	73.6	162	64.0	329	68.5
Private	60	26.4	91	36.0	151	31.5
Total	227	100.0	253	100.0	480	100.0

4.2 Bivariate Analysis

This section reports the association between the outcome variable, saving habits of employee not to save status and each predictor of the variables. This was done by cross-tabulating each predictor variables against the outcome variable and tested using chi-square and likelihood ratio tests. Also, frequency distributions of each category of predictor variables were included. No association is found between saving habits status and the explanatory variable: religion, frequency of visiting spiritual places, credit access and interest rate of employee. Because the asymptotical significance value of both tests result exceeds level of significance (α=0.25).

Based on the Wald test results, 23 covariates with p-values less than 0.25 are selected for inclusion in the multiple logistic regressions and Bayesian logistic regression analyses. First, the results of multiple logistic regression models are presented and then extended it to the Bayesian approach.

Bivariate Association of Saving Habits and Explanatory Variables

Pearson's chi-square was used to investigate the association between saving habits and explanatory variables of demographic, socio-economic and service provision explanatory variables and Spearman's correlation coefficients were used to measure of association between saving habits and other nominal variables. Table 4.5 indicates that gender, employer, parent asset on saving, membership of saving association, getting any additional allowance, transport service, job satisfaction and inflation were associated with saving habits of respondents at 0.05 levels of significance.

Table 4.5: Nominal Measure Characteristics of Bivariate Association of Saving habits

	Cramer's V		Contingency Coefficient		Pearson Chi-Square		
	Value	Sig.	Value	Sig.	Value	df	Sig.
Gender	.208	0.000	.203	0.000	20.699	1	0.000
Employer	.103	0.025	.102	0.025	5.047	1	0.025
Parent asset on saving	.095	0.038	.094	0.038	4.310	1	0.038
Credit access in the sector	.029	0.519	.029	0.519	0.416	1	0.519
Member of saving association	.216	0.000	.211	0.000	22.410	1	0.000
Transport Service	.102	0.026	.101	0.026	4.984	1	0.026
Getting any allowance	.172	0.000	.170	0.000	14.242	1	0.000
Job Satisfaction	.139	0.002	.137	0.002	9.246	1	0.002
Income Satisfaction	.089	0.050	.089	0.050	3.837	1	0.050
Inflation	.103	0.025	.102	0.025	5.046	1	0.025
Interest rate	.027	0.561	.027	0.561	0.338	1	0.561
Community development	.022	0.634	.022	0.634	0.227	1	0.634

4.3 Multiple Logistic Regression Analysis

A forward multiple logistic regression analysis is carried out to select the most important covariates among the 23 covariates provided from the bivariate analyses. As a result, 10 of the variables are found to be significant using the forward selection likelihood ratio test of the multiple logistic regressions procedure at significance level of 0.05 (Table 4.11).

4.3.1. Assessing Model Fit

After the logistic model is formed using the selected predictor variables in the forward likelihood ratio selection procedure, the first step is to assess the overall fit of the model to the data. As described in the methodological part, the recommended test for overall fit of a logistic regression is the Hosmer and Lemeshow test also called the chi-square test.

The Hosmer and Lemeshow goodness of fit test divides cases into deciles based on predicted probabilities (Table 4.6) and then computes chi-square value from observed and expected frequencies presented in Table 4.7. The SPSS output in Table 4.6 shows the non-significance of the chi-square value. Hence, we do not reject the null hypothesis that there is no difference between the observed and expected frequencies which indicates that the model adequately fits the data. We conclude that the model adequately fits the data.

Table 4.6: Results of Hosmer and Lemeshow Test

	Chi-square	Df	Sig.
Final step	6.160	8	0.629

Table 4.7: Results of Contingency Table for Hosmer and Lemeshow Test

		Saving habits of employee = Yes		Saving habits of employee = No		
		Observed	Expected	Observed	Expected	Total
Final step	1	44	45.503	4	2.497	48
	2	41	41.072	7	6.928	48
	3	37	36.121	11	11.879	48
	4	35	30.091	13	17.909	48
	5	20	24.499	28	23.501	48
	6	21	19.496	27	28.504	48
	7	16	14.522	32	33.478	48
	8	7	9.044	41	38.956	48
	9	5	4.944	43	43.056	48
	10	1	1.707	47	46.293	48

The omnibus tests are measures of how well the model performs, the chi-square tests measures the difference between the initial model and the regression model in terms of number of correctly classified subjects or it is the change in the -2log-likelihood from the previous step. Since the omnibus test is significant, the model in final step is considered be appropriate (Table 4.8).

Table 4.8: Results of Omnibus Tests of Model Coefficients

		Chi – square	df	Sig.
Final step	Step	7.694	1	0.006
	Block	207.548	19	0.000
	Model	207.548	19	0.000

Cox & Snell R Square and Nagelkerke R Square - these are pseudo R - squares. The model summary with -2log-likelihood statistic shows the overall fit of the model (Table 4.9). Cox and Snell R square is 0.351, that is 35.1% of the variation in the dependent variable is explained by the predictor variables and the nagelkerke R-square shows that approximately 47% of in the dependent variable is explained by the predictor variables.

Table 4.9: Likelihood and pseudo R square

	-2log likelihood	Cox and Snell R square	Nagelkerke R square
Final step	456.464	0.351	0.469

Another way of assessing the goodness of the fitted model is to see how well the model classifies the observed data. In the classification table, cases with probabilities ≥ 0.50 are predicted as having the event; other cases are predicted as not having the event. The results (Table 4.10) indicate that 76.2% of the employees who haven't saved were correctly predicted and 81% of employees who saved were correctly predicted. Our model shows an overall high percentage (78.8%) of correct classification of employees in their habits either save or not save. That is, the fitted model has an overall predictive accuracy of 78.8% and this may ordinarily considered adequate.

Table 4.10: Results of Classification Table

	Observed		Predicted		
			Saving habits of employees		Percentage Correct
			Yes	No	
Final Step	Saving habits of employees	Yes	205	48	81.0
		No	54	173	76.2
	Overall Percentage				78.8

The cut off value is .500

From the results of the classification table and Hosmer and Lemeshow test we can conclude that the fitted model with 10 covariates is satisfactory (Table 4.11). Since, most of the covariates are categorical to compute odds ratio we need to have a reference category. The multiple logistic regression coefficients can be estimated using the maximum likelihood estimation method implemented in the SPSS package (Table 4.11).

Table 4.11: Results of the Final Multiple Logistic Regression Model

Variables		β	S.E.	Wald	df	Sig.	Exp(β)	95% CI for EXP (β) Lower	Upper
Age	Below 25 (Ref.)			33.076	3	.000			
	25—35	.603	.294	4.202	1	.040	1.828	1.027	3.253
	36—45	1.409	.383	13.533	1	.000	4.094	1.932	8.674
	Above 45	2.484	.444	31.247	1	.000	11.991	5.019	28.649
Education				13.988	3	.003			
	Certificate & below(Ref.)	--	--	--	--	--	--	--	--
	Diploma	.347	.396	.767	1	.381	1.414	.651	3.071
	1st degree	.763	.403	3.581	1	.058	2.144	.973	4.724
	Masters & above	2.265	.652	12.065	1	.001	9.628	2.683	34.552
Housing	Owner (Ref.)			11.799	2	.003			
	Rent private	.908	.271	11.241	1	.001	2.478	1.458	4.213
	Rent Government	.820	.485	2.858	1	.091	2.271	.878	5.875
Number of dependent family		-.569	.097	34.518	1	.000	.566	.468	.685
Member of saving association	No (Ref.)	--	--	--	--	--	--	--	--
	Yes	1.273	.268	22.557	1	.000	3.571	2.112	6.039
Transport service	No (Ref.)	--	--	--	--	--	--	--	--
	Yes	.793	.291	7.444	1	.006	2.211	1.250	3.909
Job satisfaction in the sector									
	Unsatisfied (Ref.)	--	--	--	--	--	--	--	--
	Satisfied	.913	.286	10.167	1	.001	2.491	1.421	4.365
Inflation affect saving	No (Ref.)	--	--	--	--	--	--	--	--
	Yes	1.639	.543	9.115	1	.003	5.152	1.777	14.935
Cost of expenditures				14.634	3	.002			
	Below 1000 (Ref.)	--	--	--	--	--	--	--	--
	1000—1500	.100	.306	.106	1	.745	1.105	.606	2.014
	1501—2500	.497	.318	2.433	1	.119	1.643	.880	3.067
	Above 2500	-1.712	.640	7.157	1	.007	.181	.051	.633
Cost of recreation				19.850	3	.000			
	Below 100 (Ref.)	--	--	--	--	--	--	--	--
	100—250	.922	.311	8.813	1	.003	2.514	1.368	4.621
	251—350	1.202	.411	8.548	1	.003	3.327	1.486	7.450
	Above 350	1.397	.461	9.197	1	.002	4.042	1.639	9.968
	Constant	-4.265	.779	29.989	1	.000	.014		

Ref. = Reference group

The estimated coefficients and standard errors of the estimates that are used in computing the Wald statistic and the odds ratio (Exp(β)) are presented in Table 4.11.The significance of the Wald statistic indicates the importance of the predictor variable in the model.

Discussion

The study has provided an insight into the factors that determine the saving habits of employees at Debre Birhan town. According to the results, about 47.3% of the respondents have no saving habits. Out of the employees who have no saving habits 68.5% and 56.3% were government employees and males respectively.

The most important covariates identified in the multiple logistic regressions are age, education, housing, number of dependent family member, member of saving association, transport service, job satisfaction, inflation, expenditures and cost of recreation.

The first factor which affects no saving is the number of dependent family member of the employees. The odds ratio indicates that for every one number increase of dependent family member, the probability of not to save of the same employee increases 0.566 times that of an employee who has no dependent family member (coeff -0.569, OR 0.566, P=0.000, CI 0.468, 0.685). This indicates that employees with large number of dependent family members are more likely not to save money as compared to those with small number of dependent family members. This binary logistic regression indicate that after controlling other variables in the model, the odds of saving is decreased by 43.4% for each unit increases in number of dependent family members.

Job satisfaction of the employees is the other covariate which shows significant impact on the saving habits. The chance of "no saving" depends on the job satisfaction of employees. Those who were satisfied in the sector they are working in could have 2.491 times saving habit as compared to those who are not satisfied in the sector (coeff 0.913, OR 2.491, P=0.001, CI 1.421, 4.365).

The availability of transport service has also an influence on the saving habit of employees. The odds ratios of employees who had no saving habits and no transport service is 2.211 times compared to those who have transport service use (coeff 0.793, OR 2.211, P=0.006, CI 1.250, 3.909). This implies, absence of transport service has positive relation with "not to save".

Other covariate that has significant influence on the saving habit is the rate of inflation. The probability of not saving is high for those employees who are highly affected by inflation. The odds ratio related to high rate of inflation is 5.152 times compared to low rate of inflation (coeff 1.639, OR 5.152, P=0.003, CI 1.777, 14.935).

Besides, recreation cost has impact on saving habits. The odds ratio indicates that employees whose monthly average recreation cost were above 350 have about 4.042 times high chance not to save when compared to the average recreation cost below 100 (coeff 1.202, OR 3.327, P=0.003, CI 1.486, 7.450).

In addition, housing was found to be a significant factor on the saving habit of employees. The odds ratio of "not to save" for employees who rent private house was 2.478 times higher when compared to house owner employees (coeff 0.908, OR 2.478, P=0.001, CI 1.458, 4.213). This result indicates that employees who rent private house have low saving habits than house owners. This result is similar with the study conducted in New Zealand by Andrew Coleman (2008), which reported the saving habits of house owners is more than those who rent house (Jong et al., 2006).

Furthermore, expenditure has also impact on the saving habit of employees. The results of binary logistic analysis shows that the probability of "not to save" for employees with monthly expenditure above 2500 birr is 0.181 times to that of employees whose expenditure is below 1000 birr in the town (coeff -1.712, OR 0.181, P=0.007, CI 0.051, 0.633). This result shows that employees that have expenditure above 2500 birr per month have a low rate of saving habits than those whose expenditure below 1000 birr.

The level of education of employees was also found to affect the saving habit of employees. The probability of "not to save" for masters and above is 9.628 times as compared to those whose level of education was certificate and below (coeff 2.265, OR 9.628, P=0.001, CI 2.683, 34.552). Similar Studies indicate that savings increase with level of education, (Beverly and Clancy, 2001). Solomon (1975) found that highly educated individuals tend to have higher average saving habits.

Finally, being a member of saving association has also impact on the saving habit of employees. Those who were membership of saving association have 3.571 times more chance to save when compared to those who are not membership of saving association (coeff 1.273, OR 3.571, P=0.000, CI 2.112, 6.039). Gardiol (2002) and Masson et al., (1998) indicate that being a membership of saving association improves saving habits of employees.

4.3.2. Multicolinearity Diagnostics

The assumptions of logistic regression is no multicolinearity problem between the explanatory variables. Since the calculated standard error values are small (less than 1), there is no problem of multicolinearity in the explanatory variables included in the final model of the multiple logistic regressions (Table 4.11).

4.3.3. Diagnosis of Outliers and Influential Observation

We can use Cook's Distance and DFBETAS to identify influential observations. Cook's Distance, combines information on the residual and leverage. The lowest value that Cook's Distance can assume is zero, and the higher the Cook's Distance is, the more influential in the point. In this data there is no observation whose Cook's Distance is greater than one. The minimum and maximum values of the test results for Cook's influence statistics were 0 and .53029, respectively. DFBETAs assess how each coefficient is changed by including a given observation. DFBETAs less than unity indicate no specific impact of an observation on the coefficient of a particular predictor variable at 5% level of significance (Table 4.12).

Table 4.12: Results of Diagnostic Tests for Influential Values

Tests for influential cases of DFBETA	N	Minimum	Maximum
Analog of Cook's influence statistics	480	0.00000	0.53029
DFBETA for constant	480	-0.15330	0.20123
DFBETA for Age(1)	480	-0.04821	0.04123
DFBETA for Age(2)	480	-0.08250	0.05929
DFBETA for Age(3)	480	-0.11566	0.05918
DFBETA for Education(1)	480	-0.11216	0.09072
DFBETA for Education(2)	480	-0.11353	0.09673
DFBETA for Education(3)	480	-0.28816	0.11225
DFBETA for Housing(1)	480	-0.06861	0.04716
DFBETA for Housing(2)	480	-0.18395	0.11775
DFBETA for N0. of dependent family	480	-0.01251	0.02293
DFBETA for MEM_SAV association(1)	480	-0.06347	0.04394
DFBETA for T_ Service(1)	480	-0.07111	0.04787
DFBETA for Job _Satisfaction(1)	480	-0.07721	0.05016
DFBETA for Inflation(1)	480	-0.20266	0.17532
DFBETA for Expenditure(1)	480	-0.04144	0.05787
DFBETA for Expenditure(2)	480	-0.05372	0.05689
DFBETA for Expenditure(3)	480	-0.20523	0.37187
DFBETA for REC _Cost(1)	480	-0.07436	0.07177
DFBETA for REC _Cost(2)	480	-0.14738	0.09417

4.4 Bayesian Logistic Analysis

For the Bayesian logistic regression analysis, Gibbs sampling method in the WinBUGS is used. The Bayesian model used is normal-normal, in which the dependent variable saving habits is assumed to follow a normal distribution with the prior of the coefficients normal distributed uninformative priors, we assume that the regression parameters of interest all follow a normal distribution with mean = 0 and precision = 1.0e-3 and the inverse Gamma distribution as a prior for σ^2 with shape parameters 0.01. Among the variables considered the number of family member, number of dependent family member and distance of working place are continuous variables. All the other predictor variables are categorical and so dummy variables are used in setting up the model in WinBUGS. Three chains of parameters were simulated for 50000 iterations each. A total of 30000 posterior samples and the first 20000 iterations are discarded as the burn-in stage by checking the time Series (history plots) of all the parameters.

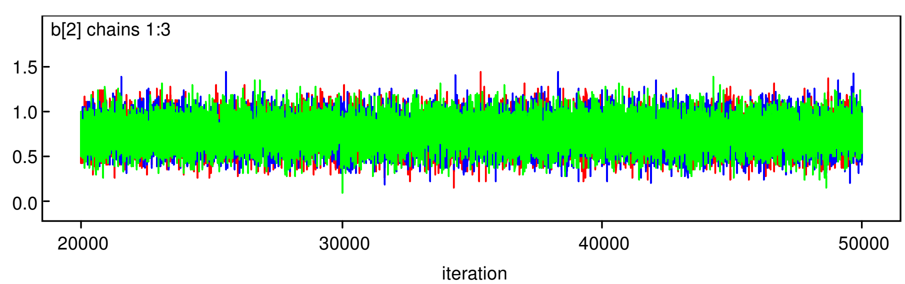
4.4.1. Assessment of Model Convergence

Four methods of checking the convergences of MCMC output were used in this study. Plots are shown in Figures 4.1- 4.4.

Time Series

Time series plots (iteration number on x-axis and parameter value on y-axis) are commonly used to assess convergence. If the plot looks like a horizontal band, with no long upward or downward trends, then we have evidence that the chain has converged.

The four independently generated chains demonstrated good "chain mixture" an indication of convergence (Figure 4.1). The Time series plots show that the chains with three different colors overlap one over the other.



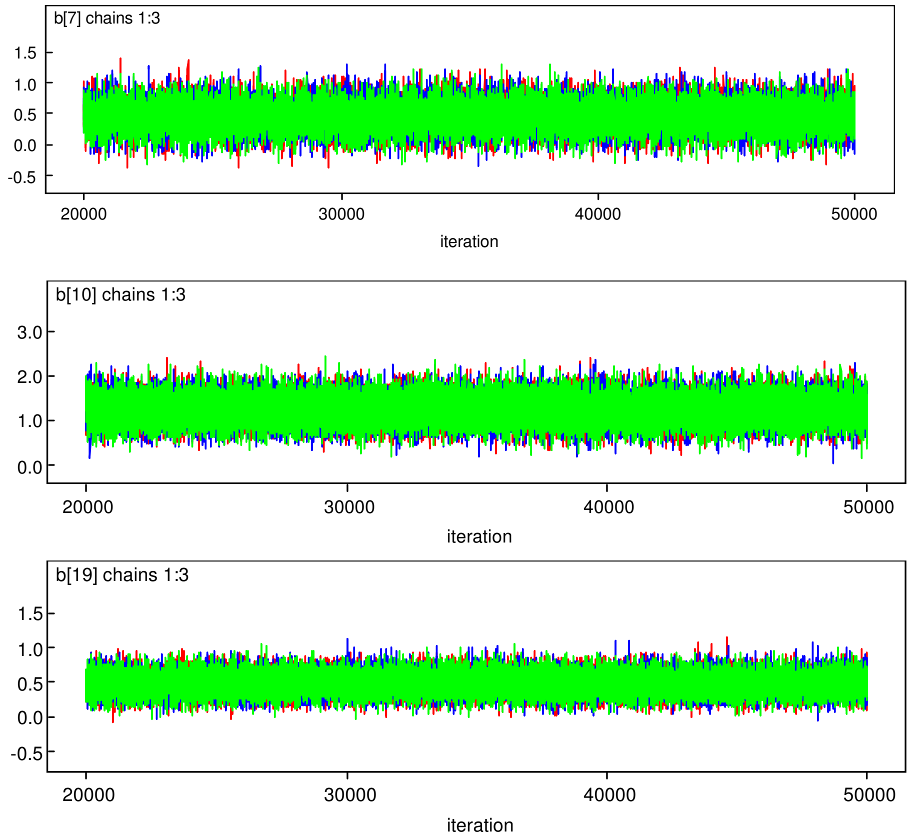
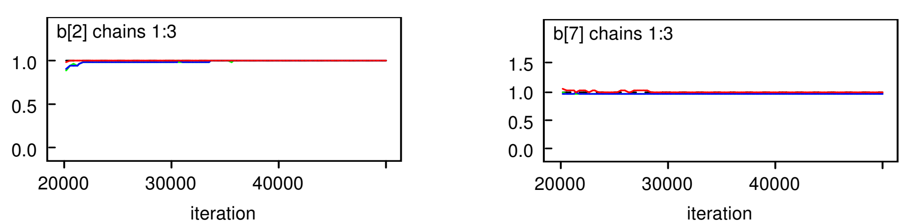


Figure 4.1: Convergence of Time Series Plots for the Coefficients of Age of the respondent, Housing, Membership of Saving Association, and Recreation cost

Gelman-Rubin

The Gelman-Rubin (GR) statistic is used for assessing convergence. For a given parameter, this statistic indicates the variability within parallel chains as compared to variability between parallel chains. The model is judged to have converged if the ratio of between to within variability is close to 1. 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 across the width of the plot. Since in our plot the red line is close to one, we can consider this is evidence for convergence.



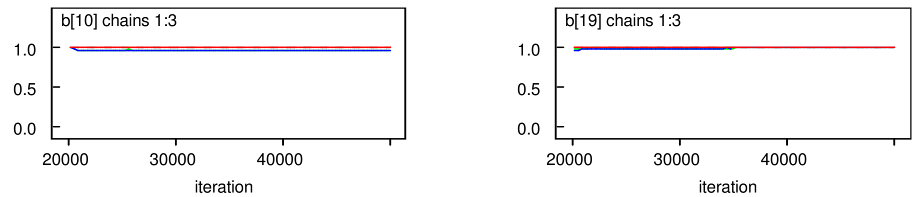
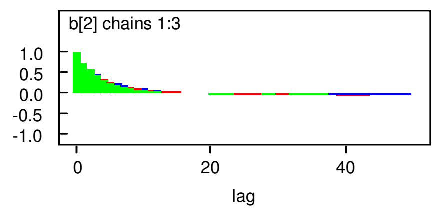


Figure 4.2: Convergence Using Gelman-Rubin Statistic for the Coefficients of Age of the respondent, Housing, Membership of Saving Association, and Recreation cost

Autocorrelation Function

This option produces lag-autocorrelations for the monitored parameters within each chain. High autocorrelations indicate slow mixing within a chain and, usually slow convergence to the posterior distribution. The plots show that the four independent chains were mixed or overlapped to each other and hence this is an evidence of



convergence.

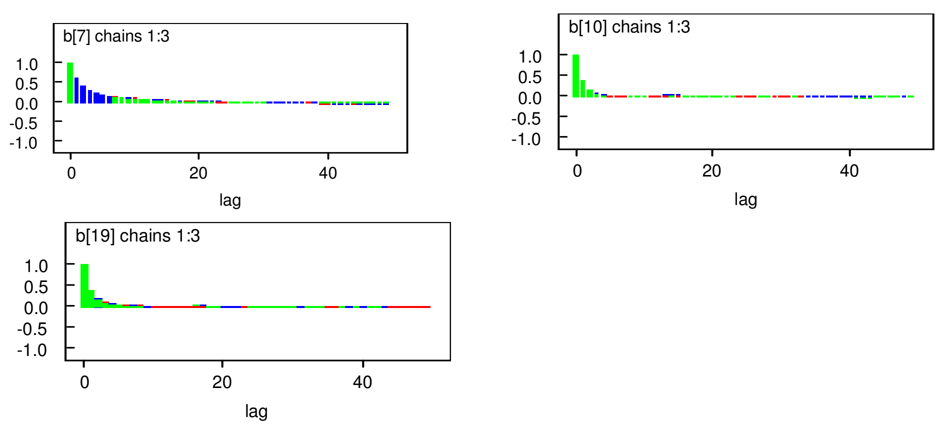


Figure 4.3: Convergence of Autocorrelation Plots for the Coefficients of Age of the respondent, Housing, Membership of Saving Association, and Recreation cost

Kernel Density

Density is also used as an alternative method for identifying model convergence. The plots for all statistically significant covariates indicated none of the coefficients have bimodal density, and hence the simulated parameter values were converged.

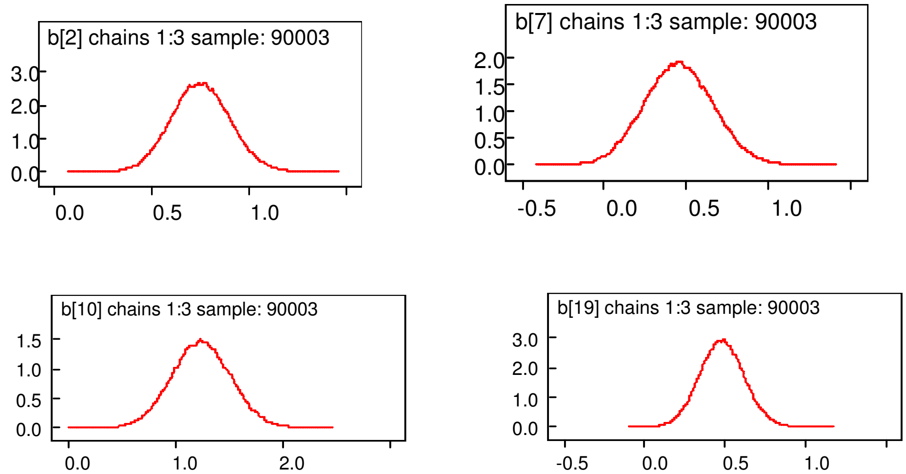


Figure 4.4: Convergence for Density Plot of the Parameter's of Age of the respondent, Housing, Membership of Saving Association, and Recreation cost

4.4.2. Assessing Accuracy of Bayesian Model

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 the sample standard deviation. In Table 4.13, MC errors for all predictor variables are 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 statistic.

Table 4.13: Results of Comparison of MC error with 5% of sd

Parameters	node	5% of sd	MC error
Sex of the respondents	b_{1}	0.01327	0.002018
Age of the respondents	b_{2}	0.009525	0.002214
Level of education	b_{3}	0.00751	0.001222
Employer	b_{4}	0.014065	0.001894
Family size	b_{5}	0.006195	0.002209
Monthly Salary	b_{6}	0.008835	0.001919
Housing	b_{7}	0.01052	0.001745
Parent asset on saving	b_{8}	0.01336	0.002074
Number of dependent family member	b_{9}	0.007855	0.002478
Member of saving association	b_{10}	0.013695	0.001531
Years live at Debre Birhan town	b_{11}	0.055	7.944E-4
Saving money for quality of life	b_{12}	0.01701	0.004269
Distance home to work place in km	b_{13}	0.004782	7.163E-4
Transport service for working sector	b_{14}	0.015725	0.001675
Getting any allowance service in the sector	b_{15}	0.013455	0.001219
Job Satisfaction in the sector	b_{16}	0.016255	0.003534

Income satisfaction on the job	b_{17}	0.0134	0.001945
Inflation	b_{18}	0.028555	0.01229
Recreation cost	b_{19}	0.01239	0.001603
Supporting others with money	b_{20}	0.006875	8.57E-4
Expenditure cost	b_{21}	0.00782	0.001218
Ethnicity	b_{22}	0.00614	6.158E-4
Marital status	b_{23}	0.01243	0.001842

Once model convergence and model accuracy is achieved we can talk about the variables which have significant contribution for the prediction of the response variable. Using Bayesian logistic regression we get 13 variables which are significant at 5% level of confidence: age of the respondents, level of education, monthly salary, housing, number of dependent family member, membership of saving association, distance from home to work place, transport service for working sector, job satisfaction in the sector, inflation, supporting others with money, recreation cost and expenditures cost. To see assessing independence of the samples after burn in, we look at the lag 4 autocorrelations. It indicates that there is high autocorrelation between the samples after burn-in. For each Bayesian model posterior mean (β_i), standard deviation (σ) and Monte Carlo (MC) error were calculated to assess the accuracy of the simulation.

The results obtained were similar to the findings by Congdon, 2005; Gerlach et al, 2007; Marjerison, 2006. The variables, monthly salary, distance from home to work place and supporting others with money showed significant impact on saving habits, and were used in the MCMC method of Bayesian model.

As seen in Table 4.14, the results of this analysis showed that the 95% credible interval for intercept of monthly salary, the posterior standard deviation σ excludes 0. The mean value related to "not to saved" with monthly income less than 1000, whose posteriori expected to be equal to β_6=0.1479 with a standard deviation of 0.1767. Comparing the posterior mean of the parameter, supporting others with money β_{20}= 0.334 with sd=0.2478 than employees who are not supporting others.

The 95% credible set displayed in Table 4.14 covers negative values for distance from home to work place β_{13}= -0.1275 with CI = (-0.3158, -0.0593), monthly salary β_6=0.1479 with CI = (-0.2019, -0.4935) and supporting others with money β_{20}=0.334 with CI = (-0.1488, -0.8245) relations to saving habits of employees respectively.

Model of the Bayesian Logistic Regression

Based on the binary logistic regression analysis and the Bayesian logistic regression analysis the following model is written as follows:

$$\ln\left(\frac{p}{1-p}\right) = -4.683 + 0.7537 \times AG + 0.4721 \times ED + 0.1479 \times MS + 0.455 \times HO$$
$$-0.6898 \times NDF + 1.243 \times MSA - 0.1275 \times DIS + 0.6651 \times TRS + 0.8164 \times JS$$
$$+1.761 \times IN + 0.4793 \times REC + 0.334 \times SUP - 0.1041 \times EXP$$

Table 4.17: Results of Bayesian Model

Model parameters	Node	Posterior Mean	Standard deviation	MC error	95% Credible set	
Alpha		-4.683	0.8701	0.02222	-6.399	-3.012
Sex of the respondents	b_1	0.3478	0.2654	0.002018	-0.1736	0.8702
Age of the respondents *	b_2	0.7537	0.1502	0.001222	0.4648	1.053
Level of education *	b_3	0.4721	0.1905	0.002214	0.1053	0.852
Employer	b_4	0.4533	0.2813	0.001894	-0.09576	1.008
Number family member	b_5	0.0519	0.1239	0.002209	-0.1858	0.2971
Monthly Salary **	b_6	0.1479	0.1767	0.001919	-0.2019	-0.4935
Housing *	b_7	0.455	0.2104	0.001745	0.0502	0.8711
Parent asset on saving	b_8	-0.204	0.2672	0.002074	-0.7304	0.3183
Number of dependent family *	b_9	-0.6898	0.1571	0.002478	-1.005	-0.3865
Member of saving association *	b_{10}	1.243	0.2739	0.001531	0.7127	1.785
Years live at Debre Birhan town	b_{11}	-0.0102	0.11	7.944E-4	-0.2236	0.2053
Saving money for quality of life	b_{12}	0.4249	0.3402	0.004269	-0.2405	1.088
Distance home to work place **	b_{13}	-0.1275	0.09564	7.163E-4	-0.3158	-0.0593
Transport service in working area *	b_{14}	0.6651	0.3145	0.001675	0.05394	1.284
Getting any allowance service	b_{15}	0.2497	0.2691	0.001219	-0.2778	0.7797
Job Satisfaction in the sector *	b_{16}	0.8164	0.3251	0.003534	0.1773	1.452
Income satisfaction on the job	b_{17}	0.00118	0.268	0.001945	-0.5215	0.5282
Inflation *	b_{18}	1.761	0.5711	0.01229	0.6938	2.935
Recreation cost *	b_{19}	0.4793	0.1375	8.57E-4	0.2144	0.7542
Supporting others with money **	b_{20}	0.334	0.2478	0.001603	-0.1488	-0.8245
Expenditure cost *	b_{21}	-0.1041	0.1564	0.001218	0.4122	0.2001
Ethnicity	b_{22}	0.1763	0.1228	6.158E-4	-0.06252	0.419

Marital status	b_{23}	0.1208	0.2486	0.001842	-0.3666	0.6087

* indicates significance in Binary logistic regression and Bayesian analysis

** indicates significance in Bayesian analysis

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The main objective of this study was to identify factors that affect the saving habits of employees. The descriptive analysis of saving habits shows that of the employees considered, 47.29% were found to have no saving habits and 52.71 % of them were found to be saving habits at the time of the study period. The binary logistic regression and Bayesian logistic analysis showed that age, education, housing, number of dependent family member, member of saving association, transport service, job satisfaction in the sector, inflation, cost of expenditures and cost of recreation were the major factors that affect the saving habits of employees in the town. Family size and lower income are also indicators of low saving habits of employees.

The capacity of institutions employees utilized, formal method of saving institutes is higher than informal saving methods of saving institutes.

The binary logistic regression indicated that number of dependent family members and job satisfaction in the sector were significantly related with saving habits of employees. Married employees have higher saving habits than those who are not married employees. This study also indicated that sex, religion, ethnicity, credit access in the sector, living at DB, income satisfaction, interest rate were not significant predictors of saving habits of employees.

Monthly salary, distance from home to work place and supporting others with money are significant predictor in addition that were identified using Bayesian analysis.

Based on Bayesian analysis, number of dependent family, distance home to work place and cost of expenditure are correlated negatively with the parameter of saving habits of employees.

5.2 Recommendations

To minimize factors of no saving habits of employees in the town, the level of education should be considered when an employee has a complicated living condition. When this is the case, appropriate monthly salary and house should be given special attention.

Putting the above consequences of no saving habits, the following recommendation should be implemented by the concerned bodies:

- Since the level of education is one of the problems identified in this study, attention should be given education and training of employees.
- Above half percent (50%) of employees are private house rents. This indicates that reduce their saving habits of employees, thus, government should give attention for those people to solve this problem.
- Extended working time has been observed, and as a result employees' satisfaction to their work is low. Hence, either additional allowance should be given or their monthly salary should be increased for the employees.
- Transport service from home to work place and back to home should be encouraged in the sector.
- Work more on awareness creation and advocacy.
- Support and promote the activities of saving associations.

REFERENCES

Agresti, A. (2002). An Introduction to Categorical Data Analysis. John Wiley and So, Inc.

Andrew, Coleman. (2008). Inflation and The Measurement of Saving and Housing Affordability.

Befekadu Kereta. (2007). Outreach and Financial Performance Analysis of Microfinance Institutions in Ethiopia.

Beverly, S., & Clancy M. (2001). Financial Education in A Children and Youth Savings Account Policy Demonstration: Issues and Options (Research Background Paper 01-5). St. Louis, MO: Washington University, Center for Social Development.

Bewick, L. and Jonathan, B. (2005). Statistics Review 14: Logistic Regression.

Central Statistical Agency (CSA), (2008). Summary and Statistical Report of Population and Housing Census Results. 2007. Addis Ababa, Ethiopia.

Cochran, W.G., (1977): Sampling Techniques. Third Edition, John Wiley and Sons (ASIA) Pte Ltd., Singapore.428p

Congdon P. (2005). Bayesian Models for Categorical Data, Queen Mary, University, UK

David P.M. Scollnik (2006). Bayesian Reserving Models Inspired by Chain Ladder Methods and Implemented Using Winbugs, Department of Mathematics and Statistics, University of Calgary, Canada.

Feifei Wang, (2010). A Comparison of Employee Saving Motives Between Chinese and Americans

Gardiol A. L., Yang X. B., and Carriquiry A. L. (2002). Bayesian Logistic Regression of Soybean Sclerotinia Stem Rot Prevalence in the U.S. North-Central Region: Accounting for Uncertainty in Parameter Estimation, Kearney Agricultural.

Gebeyaw Aychile, (2008). Financial Performance of National Bank of Ethiopia Workers' Savings and Credit Association with Special Emphasis to Adjustments

Gerlach, R. B. and Hall A. (2007). A Bayesian Approach to Variable Selection in Logistic Regression with Application to Predicting Earnings Direction from Accounting Information School of Finance & Economics, University of Technology, Australia

Gill.J (2004). Bayesian Methods for The Social and Behavioral Science Approach, Academic Price, Inc

Gokhale, (2000). 'Personal Saving in Developing Nations, Further Evidence'.

Greenwald, Grinstein-Weiss, M. Zhan, and M. Sherraden, (2001). Saving Performance in

Individual Development Accounts: Does Marital Status Matter? Journal of Marriage and Family, 68(1), 192-204.

Hosmer and Lemeshow (1998). Applied Logistic Regression, 2nd Ed., University of Massachusetts and the Ohio State University, Massachusetts and Columbus http: //www.Bu.Cam.Uk/Winbugs/Cont.Shml (Accessed On 12 August 2011).

Jong-Youn Rha, Catherine P. Montalto, and Sherman D. Hanna, (2006). The Effect of Self-Control Mechanisms on Household Saving Behavior

Karim Moussaly, (2010). Participation in Private Retirement Savings Plans, 2008

Kassa Woldesenbet (1998). Impact Analysis of Southern Ethiopia Micro Enterprise Project on Beneficiaries, Social & Economic. A Paper Presented to the National Micro Enterprise Work Shop (Amharic Version), Addis Ababa.

Marjerison W. M., Jr. (2006). Bayesian Logistic Regression with Spatial Correlation: An Application to Tennessee River Pollution, Masters of Science in Applied Statistics, Worcester Polytechnic Institute

Masson, P., Bayoumi, T. And Samiei, H. (1998). International Evidence on the Determinants of Private Saving, the World Bank Economic Review, 14 (3):393

Mengistu Bediye (1997). Determinants of Micro Enterprises Loan Repayment and Efficiency of Screening Mechanism in Urban Ethiopia: The Case of Bahirdar & Awassa Towns, Department of Economics, Addis Ababa University.

Merkle, E., Sheu, C. And Trisha, G. (2005). Simulation-Based Bayesian Inference Using Winbugs. Winbugs Tutorial Outline:

Michael Sherraden, Mark Schreiner, Margaret Clancy, Lissa Johnson, Jami Curley, Michal Grinstein-Weiss, Min Zhan, Sondra Beverly, 2001. Savings and Asset Accumulation in Individual Development Account

Pelrine, Richard and Olive Kabatalya, (2005). “Savings Habits, Needs and Priorities in Urban Uganda” Kampala: USAID/Rural SPEED.

Sherraden, M. S., Johnson, L., Elliott, W., Porterfield, S., & Rainford W. (2007). School Based Children’s Saving Accounts for College: The I can Save Program. Children and Youth Services Review, 29(3), 294-312.

Solomon, Lewis C. (1975). The Relation Between Schooling and Savings Behavior: In Education, Income and Human Behavior, Edited by, F. Thomas Juster (253-293), New York: Mcgraw Hill Book

Ziorklui S.Q and Barbee W (2003). Financial Sector Reforms Strategies and Financial