

**EMPIRICAL STATISTICAL MODELLING FOR CROP YIELDS PREDICTIONS:  
BAYESIAN AND UNCERTAINTY APPROACHES**



**UNIVERSITY OF CAPE TOWN**

Authored by

Rasheed A. Adeyemi

STD.NO: ADYRAS001

MSc Thesis Submitted on 19<sup>th</sup> September, 2014

MSc Thesis Presented for the Degree of Master of Science in Mathematical Statistics in the  
Department of Statistical Sciences, University of Cape Town, South Africa.

Supervisor: Professor Renkuan Guo

Co-supervisor: Professor Tim Dunne

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

## Declaration

I, Rasheed A. Adeyemi, hereby:

1. Grant the University of Cape Town free licence to reproduce the above thesis in whole or part for the purpose of research.
2. This thesis is my own work, both in concept and execution.
3. Neither the substance nor any part of the above thesis has been submitted in the past, is being, or is to be submitted for a degree at this University or at any other University.
4. Each significant contribution and quotation in this thesis from the work of other people has been duly acknowledged and listed in the Bibliography.

---

Signature

---

Date

## **Acknowledgements**

First and foremost, praises be to God, the Lord of Seven Heavens and Earth, the giver of all knowledge for His Mercy and Sustenance on me in the course of this programme.

I am indebted to my supervisor, Professor Renkuan Guo for his invaluable advice, directions, and his continuous support and help in my thesis.

My sincere thanks go to Professor Tim Dunne for his support and encouragement. His guidance throughout my research, professional expertise and proofreading the manuscripts severally and useful comments has led to completion of this master thesis. Your helpfulness and support is highly appreciated.

I am grateful to the current HOD, Department of Statistical Sciences, Associate Professor Francesca Little and immediate past HOD, Associate Prof Christien Thiar for their timely interventions to assist me in meeting my obligations to my funders. I appreciate all the supporting staff in the department for creating a friendly environment for learning. Thanks goes to Hassan Saqid for his assistance in R code.

Finally, I am grateful to my employer, Federal University of Technology, Minna and Tertiary Education Trust Fund (TEFT) Abuja for providing me with the financial support for this programme.

I appreciate the support of my family for their love, prayers and belief in me.

## Abstract

This thesis explores uncertainty statistics to model agricultural crop yields, in a situation where there are neither sampling observations nor historical record. The Bayesian approach to a linear regression model is useful for prediction of crop yield when there are quantity data issues and the model structure uncertainty and the regression model involves a large number of explanatory variables. Data quantity issues might occur when a farmer is cultivating a new crop variety, moving to a new farming location or when introducing a new farming technology, where the situation may warrant a change in the current farming practice.

The first part of this thesis involved the collection of data from experts' domain and the elicitation of the probability distributions. Uncertainty statistics, the foundation of uncertainty theory and the data gathering procedures were discussed in detail. We proposed an estimation procedure for the estimation of uncertainty distributions. The procedure was then implemented on agricultural data to fit some uncertainty distributions to five cereal crop yields. A Delphi method was introduced and used to fit uncertainty distributions for multiple experts' data of sesame seed yield. The thesis defined an uncertainty distance and derived a distance for a difference between two uncertainty distributions. We lastly estimated the distance between a hypothesized distribution and an uncertainty normal distribution. Although, the applicability of uncertainty statistics is limited to one sample model, the approach provides a fast approach to establish a standard for process parameters. Where no sampling observation exists or it is very expensive to acquire, the approach provides an opportunity to engage experts and come up with a model for guiding decision making.

In the second part, we fitted a full dataset obtained from an agricultural survey of small-scale farmers to a linear regression model using direct Markov Chain Monte Carlo (MCMC), Bayesian estimation (with uniform prior) and maximum likelihood estimation (MLE) method. The results obtained from the three procedures yielded similar mean estimates, but the credible intervals were found to be narrower in Bayesian estimates than confidence intervals in MLE method. The predictive outcome of the estimated model was then assessed using simulated data for a set of covariates.

Furthermore, the dataset was then randomly split into two datasets. The informative prior was later estimated from one-half called the "old data" using Ordinary Least Squares (OLS) method. Three models were then fitted onto the second half called the "new data": General Linear Model (GLM\*) (M1), Bayesian model with a non-informative prior (M2) and Bayesian model with informative prior (M3). A leave-one-out cross validation (LOOCV) method was used to compare the predictive performance of these models. It was found that the Bayesian models showed better predictive performance than M1. M3 (with a prior) had moderate average Cross Validation (CV) error and Cross Validation (CV) standard error. GLM performed worst with least average CV error and highest (CV) standard error among the models. In Model M3 (expert prior), the predictor variables were found to be significant at 95% credible intervals. In contrast, most variables were not significant under models M1 and M2. Also, The model with informative prior had narrower credible intervals compared to the non information prior and GLM model. The results indicated that variability and uncertainty in the data was reasonably reduced due to the incorporation of expert prior / information prior. We lastly investigated the residual plots of these models to assess their prediction performance.

---

\* Note GLM here has a different meaning from Generalized Linear Model in McCullagh and Nelder(1989)

Bayesian Model Average (BMA) was later introduced to address the issue of model structure uncertainty of a single model. BMA allows the computation of weighted average over possible model combinations of predictors. An approximate AIC weight was then proposed for model selection instead of frequentist alternative hypothesis testing (or models comparison in a set of competing candidate models). The method is flexible and easy to interpret instead of raw AIC or Bayesian information criterion (BIC), which approximates the Bayes factor. Zellner's g-prior was considered appropriate as it has widely been used in linear models. It preserves the correlation structure among predictors in its prior covariance. The method also yields closed-form marginal likelihoods which lead to huge computational savings by avoiding sampling in the parameter space as in BMA. We lastly determined a single optimal model from all possible combination of models and also computed the log-likelihood of each model.

**KEY WORDS:** Crop yields, general linear model, Bayesian Model Average, Model Uncertainty, Uncertainty theory, Expert opinion, Expert Prior.

# Acronyms and Abbreviations

Normal/Gaussian	Normal distribution in Probability Theory
Uncertainty Normal	Uncertainty Normal distribution in Uncertainty Theory
CBN	Central Bank of Nigeria
FAO	Food and Agricultural Organization
GDP	Gross Domestic Product
Ha	hectare (one hectare is equivalent to 10,000 square meters)
Kg	Kilograms
Km	Kilometres
NERICA	New Rice for Africa
NAFCON	National Fertilizer Company of Nigeria, Ltd
NPK	N+ P <sub>2</sub> O <sub>5</sub> + K <sub>2</sub> O (Nitrogen + Phosphate + Potash)
K	Potash/Potassium
P	Phosphorus
RTEP	Root and Tuber Expansion Programme
SADC	Southern African Development Community
SSA	Sub-Saharan Africa
NEWUs	National Early Warning Units

## Acronyms

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BMA	Bayesian Model Average
CDF	Cumulative Distribution function
CI	Confidence interval/Credible interval
CV	Cross Validation
GLM	General Linear Model
LOOCV	Leave-one-out-Cross validation
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
OLS	Ordinary Least Squares
PIP	Posterior Inclusion Probability
VIF	Variance Inflation factor

## Contents

1.	Introduction .....	16
1.1	Background of the Study.....	16
1.2	Motivation for Uncertainty modelling .....	17
1.3	Aims and Objectives of the Study.....	19
1.4	Overview of the Study .....	20
2.	Literature Review.....	23
2.1	Farmer Participation Techniques in Crop Yield Modelling.....	23
2.2	Density Distributions for Crop Yields .....	24
2.3	Common Parametric Distributions of Crop Yields.....	25
2.3.1	Gaussian Probability Distribution.....	25
2.3.2	Lognormal Distribution.....	26
2.3.3	Gamma Distribution.....	27
2.3.4	Weibull Distribution .....	28
2.4	Elicitation of Expert Probability Distribution in Agriculture.....	29
2.5	Data Sources and Processing .....	30
2.5.1	Agricultural Survey of smallholder farmers .....	30
2.5.2	Experts Survey Data.....	32
3.	Linear Regression Model and Inference .....	35
3.1	Linear Regression Model.....	35
3.2	Goodness of fit for Linear Models.....	37
3.3	Model Selection .....	38
4.	Bayesian Statistics and Inference.....	40
4.1	Bayesian Inference.....	40
4.1.1	Bayesian Inference of Linear Regression Model .....	41
4.1.2	Prior Distribution for model Parameter.....	42
4.1.3	Prior Specification in Linear Regression Model.....	44
4.1.4	The Posterior Distribution.....	45
4.1.5	Bayesian Predictive Model .....	46
4.1.6	Empirical Bayes Method.....	47
4.1.8	Markov Chain Monte Carlo Sampling Method .....	48
4.1.9	Bayesian Credible Region Estimation .....	48
4.2	Analytical Procedure for Split Data .....	50
4.3	Leave-One-Out-Cross-Validation (LOOCV).....	51

4.4	Bayesian Model Selection and Model Average .....	51
4.4.1	Bayesian Model Averaging (BMA).....	52
4.4.2	Bayes Factor.....	53
4.4.3	AIC Weights for Model Comparison.....	55
4.4.4	Bayesian Model Selection using Zellner`s g-prior.....	57
5.	Foundation of Uncertainty Theory and Estimation Procedures .....	60
5.1	Review of Uncertainty Statistics.....	60
5.2	Uncertainty Measures and Relations.....	61
5.3	Axioms and Definitions .....	63
5.4	Uncertainty Measures .....	64
5.4.1	Some Uncertainty Distributions.....	65
5.4.2	Inverses of Uncertain Distribution.....	68
5.4.3	Uncertain Normal Distribution .....	70
5.4.4	Uncertain Expected Value .....	72
5.4.5	Expectation of a function of an uncertain variable .....	74
5.4.6	Uncertain Variance .....	77
5.5	Uncertain Distance.....	78
5.5.1	Distance for Uncertainty Normal Distributions .....	80
5.5.2	Distance between Two Uncertainty Linear Distributions.....	81
5.5.3	Distance for Uncertainty Linear and Uncertainty Normal Distributions .....	83
5.7	Uncertainty Inference and Estimation.....	84
5.7.1	Expert experimental data .....	85
5.7.2	Nonparametric Estimation Method.....	86
5.7.3	Method of Least Squares.....	87
5.7.4	Method of Moments.....	88
5.7.5	Methods of Delphi .....	89
5.8	Summary .....	91
6.	Results - Analysis of Crop Yields Data Using Uncertainty Statistics .....	92
6.1	Estimation of Uncertainty Distributions .....	92
6.1.1	Empirical Uncertainty Distributions .....	92
6.1.2	Results of Analysis of Uncertainty Empirical Estimation .....	94
6.2	Results of Estimation of Uncertainty Distributions .....	95
6.2.1	Results of Uncertainty Linear Distribution analysis .....	96
6.2.2	Results of Uncertainty Normal Distribution analysis .....	99

6.3	Fitting data to a Deterministic Production function.....	101
6.4	Estimating Uncertainty Distribution from Multiple Expert Data.....	102
6.4.1	Aggregate Method.....	102
6.4.2	Analysis of Multiple Experts' Data Using Delphi Method.....	103
6.5	Fitted Uncertainty Distributions for Sesame Seed Yield .....	106
6.7	Distance Estimates between Expert Distributions and Uncertainty Normal Distribution ..	107
6.8	Summary and Discussion.....	108
7.	Results of Linear Regression Analysis on Cereal Crop Yields .....	109
7.1	Descriptive Statistics of Agricultural survey data.....	109
7.2	Model Specification.....	112
7.3	Results Regression Analysis on full Data.....	113
7.3.1	Regression Analysis for Full Data on Sorghum Yield.....	114
7.3.2	Regression Analysis on Millet Yield for Full Data.....	119
7.4	Bayesian Prediction Analyses.....	122
7.5	Results of Regression Analysis on Split Data.....	123
7.5.1	Linear Regression Model on old Data for Sorghum .....	124
7.5.2	Linear Regression Model on New Data .....	128
7.5.3	Bayesian Linear Regression Model on new Data .....	132
7.5.4	Leave One Out Cross Validation for Model Comparison for Sorghum .....	138
7.6	Results on Linear Regression Analysis on Split Data for Millet .....	139
7.6.1	Bayesian Regression Model on New Data.....	140
7.6.2	Leave One Out Cross Validation for Model Comparison.....	143
7.7	Assessing the Model Accuracy and Residual Plots .....	144
7.7.1	Residual Plots on Sorghum.....	144
7.7.1	Residual Plots on Millet Yields .....	146
7.8	Summary and Discussion.....	148
8.	Results - Analysis of Bayesian Model Selection and Model Average.....	149
8.1	Bayesian Model Average (BMA) analysis .....	149
8.1.1	Bayesian Model Averaging (BMA) analysis for Sorghum.....	149
8.1.2	Model Selection Using AIC Weight of Evidence on Sorghum Yield.....	154
8.1.3	Bayesian Model Average analysis for Millet.....	155
8.2	Models selection by Zellner's g-prior.....	160
8.2.1	Bayesian Model Selection for predicting Sorghum yield .....	161
8.2.2	Bayesian Model Selection for predicting Millet yield .....	162

8.3	Summary and Discussion.....	163
9.	Conclusions and Suggestions .....	165
9.1	Conclusions.....	165
9.2	Suggestion for Further Research.....	165
	Bibliography .....	168
	Appendix A: Mathematical Derivations .....	175
A.1	Jeffreys' Prior.....	175
A.2	Derivation of Posterior Distribution for a normal prior .....	175
	Appendix B: Expert Survey Questionnaire .....	177
B.1:	Instructions.....	177
B.2:	Expert Eliciting Information on Uncertain Variable.....	178
B.3:	R Code for Estimation of Uncertainty moments .....	180
	Appendix C: R Code for Regression Model .....	191
C.1:	R2WinBUGS Code for Bayesian model with non- informative priors.....	191
C.2:	R code OLS and Direct MCMC Estimation.....	196
C.3:	Bayesian Predictive model for set of covariates .....	198
	Appendix D: Regression analysis on SPLIT data .....	203
D.1:	R Code for Linear Regression analysis on Split data.....	203
D.2:	R Code for Bayesian Models on SPLIT data .....	205
D.3:	R Code for LOOCV cross –validation on SPLIT data.....	206
D.4:	Cross Validation ouput.....	208
	Appendix E: Raw Data .....	211
E.1:	Raw data for Sorghum (15 Observations).....	211
E.2:	Raw data for Millet (15 observations).....	211
	Appendix F: R code Bayesian Model Average.....	212
F.1:	Sorghum (BMA analysis).....	212
F.2:	Millet (BMA analysis).....	213
	Appendix G: R code Bayesian Model Selection Using Zellner's g-prior .....	215
G .1:	Bayesian Model Selection for Sorghum .....	215
G .2:	Bayesian Model Selection for Millet .....	216

## List of Tables

2.1:	Description of model variables.....	31
2.2:	Annual Estimates Crop yield (ton/ha) provided by Agronomy Experts.....	33
2.3:	Annual Sesame seed yield (kg/ha) from Experts domain knowledge.....	34
4.1:	List of Likelihood, Prior Distributions and their corresponding Posteriors .....	43
4.2:	Interpretation of Evidence for Bayes factor.....	54
5.1:	Relationship between Probability theory versus Uncertainty theory.....	62
5.2:	Relationship between Credibility Theory versus Uncertainty Theory.....	62
5.3:	Example of Uncertainty measures.....	66
5.4:	Computation of Uncertainty Distribution and Density Measures.....	66
5.5:	Comparison between Uncertainty normal variable and Gaussian random variable ....	73
6.1:	Summary of Observed and Expected values of Crop yields by Experts .....	96
6.2:	Summary estimates of Uncertainty Linear Distribution of crop yields (kg/ha).....	98
6.3:	Summary estimates of Uncertainty Normal Distribution of crop yields (kg/ha).....	100
6.4:	Composite data from multiple agricultural experts.....	102
6.5:	First round Data Computation for multiple agricultural experts.....	103
6.6:	Decision Table.....	103
6.7:	Adjusted data after the first round Computation and feedback.....	104
6.8:	Second round iteration computation for multiple agricultural experts.....	105
6.9:	Decision Table .....	105
6.10:	Estimates of Uncertainty Linear and Uncertainty Normal Distribution on Sesame....	106
6.11:	Distance Estimates between Standard Normal Distribution and estimated Expert Distributions for Crop yield.....	107
7.1:	Descriptive Statistics of variables on Sorghum yield .....	110
7.2:	Descriptive Statistics of the predictors of model variables on Millet .....	111
7.3:	Summary estimates of MLE and Posterior of Bayesian Estimation on Sorghum yield	115
7.4:	Summary estimates of MLE and Posterior of Bayesian Estimation on millet yield....	119
7.5:	Predicted and Expected Estimates of Simulated yield for set of Covariates.....	122
7.6:	Variable types and Description of Variables in the data.....	124
7.7:	Linear regression model fitted on old data for Sorghum.....	124
7.8:	Correlation Matrix of the numerical independent on old data set.....	125
7.9:	Variance inflation factors (VIF) of numerical independent variables on old data	125
7.10:	Difference in the Parameters when possible leverage and influential observations are removed from old data set.....	126 127
7.11:	Linear regression model fitted on the new data for Sorghum. ....	132

7.12:	Correlation matrix of the numerical independent variables on the new Data .....	129
7.13:	Variance inflation factors (VIF) of numerical independent variables on the new data	129
7.14:	Difference in the Parameters when possible leverage and influential observations are removed from old data set.....	131
7.15:	Linear regression model fitted on the new data .....	128
7.16:	Prior Parameters for an informative Bayesian linear regression model .....	133
7.17:	Summary of Bayesian Regression Estimates with Informative prior on the new data.	134
7.18:	Geweke diagnostic statistics for each variable of the Bayesian linear regression model with informative prior.....	136
7.19:	Bayesian Regression Estimates with Non-informative Prior on new Data .....,	137
7.20:	Geweke diagnostic statistics for each variable of the Bayesian linear regression model with non-informative prior.....	137
7.21:	Models comparison using the Leave one out cross validation.....	138
7.22:	Linear regression model fitted on the new Data for Millet.....	139
7.23:	Summary of Bayesian Regression Estimates with informative prior on new data .....	140
7.24:	Geweke diagnostic statistics for each variable of the Bayesian linear regression Model with informative prior .....	141
7.25:	Bayesian Regression Estimates with Non-informative Prior on new data .....	142
7.26:	Geweke diagnostic statistics for each variable of the Bayesian linear regression model with non-informative prior .....	142
7.27:	Models comparison using the Leave One Out Cross Validation .....	143
8.1:	Summary estimates and inclusion Probabilities using BMA and p-value on sorghum	150
8.2:	AIC Weight Estimates for Model Comparison on Sorghum .....	154
8.3:	Summary estimates and inclusion Probabilities using BMA and p-value on Millet	156
8.4:	AIC Weight Estimates for Model Comparison on Millet.....	159
8.5:	Summary Estimates of -2 log likelihood and marginal posterior probability for Model Selection using Zellner's g-prior on Sorghum.....	161
8.6:	Summary Estimates of -2 log likelihood and marginal posterior probability for for Model Selection using Zellner's g-prior on Millet.....	162

## List of Figures

2.1:	Normal Distributions .....	27
2.2:	Lognormal Distributions .....	27
2.3:	Gamma Distributions .....	28
2.4:	Weibull Distribution .....	29
4.1:	Plot of five different prior distributions .....	44
4.2:	Effect of constant $c$ on posterior estimates using Zellner's $g$ - prior .....	58
5.1:	Uncertainty Linear Distribution .....	68
5.2:	Uncertainty Normal Distribution .....	68
5.3:	Uncertainty Zigzag Distribution .....	68
5.4:	Uncertainty Lognormal Distribution .....	68
5.5:	Inverse Uncertainty Linear Distribution .....	69
5.6:	Inverse Uncertainty Zigzag Distribution.....	69
5.7:	Inverse Uncertainty Normal Distribution .....	70
5.8:	Inverse Uncertainty Lognormal Distribution .....	70
5.9:	Standard Uncertainty Normal and Gaussian Distribution .....	72
5.10:	Lebesque Finite Integral .....	73
6.1:	Expert Distribution for Maize Yield .....	95
6.2:	Expert Distribution for Rice Yield .....	95
6.3:	Expert Distribution for Sorghum Yield .....	95
6.4:	Expert Distribution for Soybean Yield .....	95
6.5:	Expert Distribution for Cowpea Yield .....	95
6.6:	Uncertainty Linear Distribution for Maize Yield .....	99
6.7:	Uncertainty Linear Distribution for Rice Yield .....	99
6.8:	Uncertainty Linear Distribution for Sorghum Yield .....	99
6.9:	Uncertainty Linear Distribution for Cowpea Yield .....	99
6.10:	Uncertainty Linear Distribution for Cowpea yield .....	99
6.11:	Uncertainty Normal Distribution for Maize .....	99
6.12:	Uncertainty Normal Distribution for Rice Yield.....	100
6.13:	Uncertainty Normal for Sorghum Yield .....	100
6.14:	Uncertainty Normal Distributions for Soybean Yield .....	100

6.15:	Uncertainty Normal Distribution for Cowpea Yield .....	100
6.16:	Fitted Deterministic function for Maize yield response to NPK fertilizer .....	101
6.17:	Expert Distribution function on Maize Yield .....	101
6.18:	Fitted Uncertainty Linear Distribution for Sesame seed yield .....	106
6.19:	Fitted Uncertainty Normal Distribution for Sesame seed yield.....	106
7.1:	Density Plots of Bayesian posterior means indicated by green bar and MLE coefficient estimates by red bar for Sorghum yields .....	119
7.2:	Density Plots of Bayesian Posterior mean indicated by green bar and MLE coefficient estimates by red bar for Millet yield.....	121
7.3:	Histogram of simulated draws of the predictive Distribution for sets of Covariates on sorghum .....	123
7.4:	Histogram of simulated draws of the predictive Distribution for sets of Covariates on millet .....	123
7.5:	Half-Normal plot of residuals for the old data. ....	126
7.6:	Half-Normal plot of leverages for the old data.....	126
7.7:	Half-Normal plot of the Cook's distance statistics for the old data.....	126
7.8:	Half-Normal plot of residuals for the new data .....	127
7.9:	Half-Normal plot of leverages for the new data .....	130
7.10:	Half-Normal plot of the Cook's distance statistics on the new data .....	130
7.11:	Trace and density plots of the posteriors for the first four variables using an informative prior .....	135
7.12:	Trace and density plots of the posteriors for the first four variables using a non- informative prior .....	136
7.13:	Average Cross Validation error and Cross Validation standard error on Sorghum.	139
7.14:	Average Cross Validation error and Cross Validation standard error on Millet ...	143
7.15:	Plot of Predicted versus observed(GLM). ....	145
7.16:	Half of Predicted and Observed versus Residual Plot (GLM).....	145
7.17:	Plot of Predicted versus observed (Non informative).....	145
7.18:	Half of Predicted and Observed versus Residual Plot (Non informative).....	145
7.19:	Plot of Predicted versus observed(Informative). ....	145

7.20:	Half of Predicted and Observed versus Residual Plot (Informative).....	145
7.21:	Combined Model: Half of Predicted and Observed versus residual plot(Sorghum)	146
7.22:	Plot of Predicted versus observed(GLM). ....	147
7.23:	Half of Predicted and Observed versus Residual Plot (GLM).....	147
7.24:	Plot of Predicted versus observed (Non informative).....	147
7.25:	Half of Predicted and Observed versus Residual Plot (Non informative).....	147
7.26:	Plot of Predicted versus observed(Informative). ....	147
7.27:	Half of Predicted and Observed versus Residual Plot (Informative).....	147
7.28:	Combined Model: Half of Predicted and Observed versus residual plot(Millet)	147
8.1:	BMA image plot of selected models and Predictor Coefficients on Sorghum	152
8.2:	Density plots of the predictor variables on Sorghum using BMA .....	153
8.3:	BMA image plot of selected models and Predictor Coefficients on Millet	158
8.4:	Density plot of the Predictor variables on Millet using BMA .....	158

# 1. Introduction

## 1.1 Background of the Study

The agricultural sector is particularly important for the Nigerian economy. The agricultural sector employed over 60 percent of the total labour force and 40 percent of the GDP in Nigeria in 2001 (Adeoti & Olubamiwa, 2009). The sector also grew by 7.4% in 2007 (CBN, 2007). The agricultural sector is central to Nigerian households and the national economy. This has made it a critical component of interest for programmes that seek to reduce poverty and attain food security in Nigeria. In the light of this significant contribution, the government has invested significantly in the agricultural sector by introducing new seed varieties, subsidizing and reinforcing the use of agrochemicals and fertilizers, encouraging soil management and afforestation, promoting ecological control, ensuring farmer' registration, introducing e-wallet facilities and constructing farmer depots, amongst other interventions.

An analysis of the sectoral real GDP conducted by the CBN (2006) indicated that the agricultural sector contributed about 42% to the national GDP compared to 41.2% contributed in 2005. The contribution of the agricultural sector to the national GDP grew steadily in 1990 from 4.2% in 2002 to 7.2% in 2006. This upsurge in the agricultural productivity was driven largely by putting into practice different agricultural policies and programmes. These include the application of zero tariffs on imported agro-allied products and the persistent implementation of the temporary embargo on the importation of some agricultural products, notably livestock products and rice. The presidential initiatives on Root and Tuber Expansion Programme (RTEP) and New Rice for Africa (NERICA), developed a hybrid strain of rice through biotechnology research in the West Africa subregion. As a result these initiatives, rice and cassava production have grown by 7.7% and 7.4% respectively. However, there was a decrease in other crops like sorghum, millet, beans, yam and cocoa (CBN, 2007). The overall objectives of these agencies and programmes aimed at improving resource use, farmers' income, productivity, food security and accelerating rural development.

Despite of all these programmes and initiatives, Nigeria still experiences food price fluctuations. Crop yields continue to decline due to inappropriate use of resources, climate change and soil degradation. Reduction in soil nutrients is also identified as a basic cause of

decreasing food security in sub-Saharan Africa (SSA) countries. Sanchez and Jama (2002) stated that the depletion in soil fertility has been estimated at an average of 600 kilogram of Nitrogen per hectare, 75 of kilogram Phosphorus per hectare and 400 kilogram of Potassium per hectare from about 200 million hectares of cultivated lands in sub-Saharan Africa (SSA). The depletion in soil fertility is attributed to land degradation and consequent decrease in farmland productivity as a primary cause of low per capita food production in SSA (Bremen et al. 2001; Sanchez, 2002).

Crop production and the prediction of crop yield have direct impact on year-to-year national and international economies and play an important role in food management (Hayes and Decker, 1996). Efforts have been made to develop various indices for different crops in various regions throughout the globe. For instance, the Southern African Development Community (SADC) countries use a combination of subjective procedures (i.e. extension officers and/or growers' assessment) and objective ones involving direct measurement. In 2009, the SADC Secretariat's Food, Agriculture and Natural Resources Directorate prepared a handbook *National Early Warning Units (NEWUs) for Food Security*. This compiles selected working papers presented as blue print for forecasting Crop yield in SADC countries.

## **1.2 Motivation for Uncertainty modelling**

Statistical modelling of crop yield can be used for many purposes such as to: determine the crop yield risk factors of the individual farmer or groups, design and rate crop yield for insurance contracts, aid decision-making in agricultural production, improve risk management under uncertain climatic variations and enhance government policy on agriculture.

Research on model building for agricultural yields has often ignored the issue of model uncertainty. Example of such studies includes, Time series models (Box et al. 2008, Tongkhaw & Kantanantha, 2012), and exponential smoothing (Montgomery, et al. 2008) and model based on crop growth indices (Jain, et al. 1992, Ramasubramainian, et al. 1999). The general principle is to select a particular model from a group of models and then progress as if the sample data was generated by chosen model.

In a standard statistical analysis such as regression analysis, one typically chooses a specific model, estimates the model parameters and substitutes the point estimates into the proposed model to make prediction. Such a method pays no attention to issues of uncertainty in the

selected model, thus it results to over-reliance on the deductions and therefore leads to biased conclusions. Hence, ignoring the uncertainty may adversely affect the predictive performance of the final model.

The functional form of a model can also generate another source of uncertainty, where model forecast is often dependent on the model used and the exact model specification for the data may not be known. There are many different models that researchers can use and the implications of the choice between them are not known with certainty. Thus, the uncertainty over which model to use is an important aspect of prediction or indeed any inference from data. The distributional form of these models could give different answers to the scientific question relating to the data at hand. This is a source of uncertainty in drawing the inference. Typically, any approach that conditions on a single model deemed to be the best, but which ignores this source of uncertainty may underestimate the uncertainty.

Few studies employ standard procedures on a related methodology for dealing with large number of regressors in forecasting is based on principal components or factor analysis as in e.g Stock and Watson (2002), Kitchen and Monaco (2003). Stock and Watson (2004), Aiolfi and Timmermann (2004) use static factor models for forecast combining; they found that the factor model forecasts improved equal-weighted averages in few situations, such as price forecasts and country growth output.

In Bayesian paradigm, the approach known as Bayesian Model Averaging (BMA) is the traditional method to tackle uncertainty in modelling, where it is natural to reflect uncertainty through probability. The BMA approach obeys a straight implementation of Bayes' theorem. BMA overcomes the difficulty of model uncertainty by conditioning, not on a single best model, but on the entire ensemble of predictor combinations. BMA used weights in combining forecasts to estimate posterior model probabilities within a Bayesian framework. The weights adopted in combining different forecasts could simply be uniform or they could be based on historical performance. For more detail, see Hendry and Clements, 2002; Diebold and Lopez, 1996; Raftery et al. (1997), Hoeting, et.al. (1999).

In related studies, model parameters are mixed up to have all possible combinations of predictors using BMA which could guarantee the reduction in the expected predictive squared error loss provided it is within the set of candidate models (Min and Zellner, 1993). This is supported by Raftery et al. (1997) who argued that if the predictive ability of the model is measured by a logarithmic scoring rule, the BMA could yield the optimal model.

Therefore, within the background of the uncertainty model, the use of BMA goes along with practical functionality considerations. In addition, strong argument for the ability of BMA to yield a better prediction can be found in, (e.g., Raftery et al. (1997), Fernández et al. (2001) and Clyde & George, 2004).

The thesis structure is organized as follows.

In the first part of this thesis, we use data obtained from subjective expert judgement. The conceptual axioms and foundation of uncertainty theory is formalised in detail. An estimation procedure for determining the model parameters in uncertainty statistics was introduced. This work leads to the development of uncertainty model that utilizes not only small sample size but is applicable in a situation where sampling distributions are not feasible. The proposed methods were implemented on experts' crop yield data.

A Bayesian Model Average approach is then introduced to investigate model structure uncertainty and determine the best model based on posterior model probability. BMA facilitates the identification of potential predictors to be included in the optimal model. BMA analysis computes the posterior model probabilities and the Akaike information criterion (AIC). An approximated AIC weight was proposed for the purpose of model comparison.

For a model involving several predictors and that ensures a stable covariate structure in the model a procedure is introduced that minimizes the influence of priors on the posterior coefficient estimates. A Zellner's  $g$ -prior method is later introduced to ensure a consistent estimation in model selection for prediction, which enhances predictive performance over the maximum likelihood estimation under BMA.

### **1.3 Aims and Objectives of the Study**

The broad aim of the present study was to explore uncertainty statistics in modelling agricultural crop yields data in a situation where the data sample is small

The first purpose utilized uncertainty theory for the elicitation of probability from domain expert data on agricultural productivity on some cereal crops. The specific objectives are to:

- Establish the foundation concept of uncertainty theory and uncertain distributions;
- Investigate the functional relationship between uncertainty theory and probability theory through their distributional properties,

- Describe the concept of domain expert survey and the associated data collection procedure; and introduce the Delphi method for multiple expert data
- Propose a method for estimation of uncertainty distributions and fit some uncertainty distributions to agricultural crop yields
- Construct empirical distributions for the expert data on crop yields
- Define an uncertainty distance measure and extend the concept to derive a distributional form for the difference between two uncertainty distributions
- Estimate the distance between expert distributions and Normal distribution as a measure of model accuracy.

The second aim explored Bayesian framework to address the issue of model uncertainty and implement findings to an agricultural survey data. The specific objectives here are to:

- Identify determinant factors that have significant effects on cereal crop yields from Nigerian farmers' agricultural production practice;
- Propose a general linear model for agricultural crop yields and estimate the regression parameters
- Compute the predicted and expected crop yields for a set of covariates
- Evaluate the effect of the expert prior or expert belief on the parameter estimation in linear regression model;
- Compute the posterior probability for the models sampled in the parameter space using BMA and evaluate the posterior inclusion probability of the predictor variables
- Make model comparison and decide the best model for prediction of crop yield using an approximate AIC weights
- Design an optimal model for a set of covariates using Zellner  $g$ -prior.

#### **1.4 Overview of the Study**

The thesis is structured into nine chapters.

Chapter 2 reviews related statistical modelling procedures for crop yields and some relevant probability distributions. The chapter also presents a review of previous expert elicitation of probability distributions in the agricultural sector. Two datasets for analysis are presented in this chapter, the data collected from an agricultural sampling survey and a domain expert survey of cereal crop yields.

In Chapter 3, we present a classical approach to a linear regression model and the related inference.

Chapter 4 gives a detailed description of the Bayesian framework and the related inferences to a linear regression model; and it describes procedures for prior estimation. The chapter also illustrates the procedure of Bayesian Model Average (BMA) and method of model selection. The chapter further proposes an approximate AIC weight for model comparison. The chapter later describes a model selection procedure under Zellner  $g$ -prior.

Chapter 5 describes uncertainty theory, the theoretical foundation of uncertainty statistics, uncertainty variable and uncertainty distributions. The chapter defines formulae for computation of uncertainty expected value, uncertainty variance and a nonparametric empirical estimation method. It establishes the functional relationship between uncertainty theory and probabilistic (randomness) uncertainty. The chapter further describes a procedure to generate data from expert domain and a method of elicitation of probability distribution for such expert domain data. A relevant estimation method was prescribed and used to fit the uncertainty distributions. We proposed an uncertainty distance as a measure for assessing the closeness between a generic model and the reality in the expert data. Initially, we have envisaged developing an algorithm for uncertainty multiple regression estimation. This could however not be achieved in uncertainty theory as there cannot be a centered covariate vector and no valid correlated estimation between uncertain variables. Therefore, the proposed model approach is been limited to one sample model.

In chapter 6, we present the results from uncertainty statistics analysis. The chapter estimates the empirical distribution for the experts and fits some uncertainty distribution functions to the cereal crop yields for expert survey data. The chapter lastly evaluate the model accuracy using distance estimate between the expert distribution and uncertainty normal distribution.

Chapter 7 presents the results of the regression analysis of general linear model to the full data set of the agricultural survey. A simulated result was presented to evaluate model predictive performance. The chapter later analysed the data using Bayesian method on a split data. The three models were generated and compared their predictive performance using a leave-one-out-cross validation method. Separate residual plots were constructed for each model as diagnostic tools to evaluate the model predictions.

Chapter 8 presents the results of the Bayesian Model selection and model Average analysis. The image plots of the regression coefficients along side model weights were visually

presented. The chapter also carried out the density plots of the predictor variables that show discernible association with the response variable of interest. The relative AIC weight was analyzed for model comparison for a set of selected candidate models. Zellner's  $g$ -prior method was also carried out to determine the optimal model. The approach also computes the log likelihood and marginal posterior probabilities of the selected models.

Chapter 9 assesses the stated objectives of this thesis, draws conclusions and makes suggestions for further studies.

## 2. Literature Review

This chapter presents a review of related literature on probability distribution of crop yields and relevant density estimation. A retrospective review of expert elicitation of probability distributions in agriculture was also discussed. Two datasets for analysis in this thesis were described: a survey of agricultural production in Niger state of Nigeria and expert survey.

### 2.1 Farmer Participation Techniques in Crop Yield Modelling

Historically, the appraisal of crop yields has been made by farmers themselves from the preceding production year, ahead of the planting season. Farmers may make these forecasts in order to plan for next cropping year and strategize their agronomic practices. For example, the planting windows, the choice of a cultivar, and the amount of fertilizer to apply all depend on the climate and soil condition. Forecasting crop yield may involve evaluation of the important production parameters. For example, it requires quantifying the land area to be planted for a specific crop at the start of growing season and the expected yield. Chandrasahas & Rai (2001) developed a Bayesian approach based on farmers' appraisal data. They incorporated actual harvest yield and farmers' appraisal data on yield from previous year(s) in order to obtain posterior probabilities which were then used for obtaining a Bayesian forecast of crop yield in the current year.

Another traditional method of yield forecasting is the evaluation of crop status by agricultural experts. The experts monitor crop growth development by observing and evaluating crop performance throughout the crop growing season, which may involve counting tiller number (i.e. number of shoots after the initial parent shoot), spikelet numbers and scoring the soil fertility percentage, percentage of crop damage from pests and fungi, percentage of weeds infestation, and so on. In this way, the yield can be forecast using regression methods, or by the knowledge from local expertise. In other cases, crop yield forecasts are based on reports by crop correspondents at regular intervals during the growing season, using crop appearance as an indicator. A means for meeting the challenges of a precise estimator at small area level is through carefully planned design known as Crop- Cutting Experiments (CCE). This is a cost effective technique that has been developed and implemented for estimation of yield crop at small area levels (Rao, 2003, 2004).

Nowadays, researchers have recognized that the weather affects crops differently at different stages of crop growth. Bellow (2007) recently studied weather condition on crops growing

and yields. He applied Bayesian mixed-effects for county yield estimation on spatial component by incorporating information from neighbouring counties. A similar crop yield forecasting method may be based on the information from remote sensing and crop simulation models. The Department of Agriculture (USDA) in United States is saddled with responsibility of density estimation of crop yield across counties. The National Agricultural Statistics Service (NASS) used to apply multi-stage stratification methods to estimate crop yields. Stasny et al. (1995) employed Bayesian estimation method incorporating weather data, cropping practices and other factors from neighbouring counties to simulate the crop yields for close counties that share similar geographic variations.

Ogallo et al., (2008); Jones & Thornton., (2003) conducted identical research works on the influence of the climate reliance on agricultural crop growth. They reported that crop growth development and yield components are affected by year-to-year climatic variability and significant effects of extreme events. In congruence to the report, Persson, et al., (2011) concluded that the vulnerability of crop production and uncertainties in changes in historical patterns of regional climate would have a severe effect on the economic and food security risks of many countries.

## **2.2 Density Distributions for Crop Yields**

Crop yield distributions are useful in risk modelling and as tools for crop rating, designing and marketing crop insurance. For decades, distributional forms underlying the crop yields have been widely disputed (Just and Weninger, 1999; Ker and Goodwin, 2000; Sherrick, et al.2004). Particularly, the shape of the density function has been discussed extensively. Just and Weninger (1999) argued that the crop yields density resembles a Gaussian distribution.

In modelling average soybean yields in U.S., Gallagher (1987) detected a negative skewness based on the fact that the yield cannot exceed the biological potential of the plant, but it is assumed that the yield can equal zero under extreme condition such as heat and disease infection. He further stressed that the skewness and changing variance of the soybean yields can be modeled using a gamma distribution function. However, Norwood et al. (2004) pointed out that it would be difficult to identify the maximum yield needed to implement the Gamma model when conducting forecasts.

In related studies, Nelson and Preckel (1989), under aprior assumption, derived a conditional beta distribution to model the distribution of corn yields at farm level in five Iowa counties

and found a negative skewness. They estimated the maximum attainable yield by maximum Likelihood and modeled deviations of yield from its maximum value as a conditional beta distribution. They found that beta distribution provides a flexible means of fitting the skewness in either direction of the bell-shape as suggested by Day, while ignoring correlation of yields between farms within the same county and pooling farm-level data to estimate corn yield response to fertilizer applications.

Other studies suggested different distribution forms to better reflect crop yield, Sherrick et al., (2004) proposed a parametric distribution, while Ozaki et al., (2007) adopted a semiparametric functions. Ker and Coble, (2003) first used Normal and beta distributions to estimate the corn yield densities and nonparametric kernel estimator to correct the crop density. They found that the semiparametric estimator with a normal distribution is more efficient than the competing parametric models (Normal and Beta) than the standard nonparametric kernel estimator.

Agricultural yield data is also been recognized to have irregular behaviour in its density structure. One of these possible variations is attributed to spatial dependence across farms. The spatial dependence may arise where the agricultural yield diminishes drastically as the distance increase. For more detail, see (Goodwin, 2001; Goodwin and Ker, 2002; Ozaki, 2005).

However, due to the skewed nature of most crop yields, other researchers have considered Weibull distribution as a flexible alternative for a wide range of skewness. Sherrick et al. (2004) compared Normal, Log-normal, Logistic, Beta, and Weibull distributions on farm level corn and soybean yields to determine best distribution for each crop. They found that the Weibull and beta distribution ranked highest based on the goodness-of-fit measures.

## **2.3 Common Parametric Distributions of Crop Yields**

In the literature, a wide range of probability distributions has been discussed in modelling crop yields. Several commonly used distributions are discussed in this section.

### **2.3.1 Gaussian Probability Distribution**

The density function of a normal distribution with mean  $\mu$  and standard deviation  $\sigma$

is

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (2.1)$$

where  $-\infty < \mu < \infty, \sigma > 0$

The entire distribution depends on two parameters,  $\mu$  and  $\sigma$ . The normal distribution is symmetric, bell shaped, and unbounded  $[-\infty, +\infty]$ . However, in most studies of crop distribution, the likelihood of a predicted  $f()$  being less than zero is negligible although the probability of 0 yields is finite. Typically, the normal distribution will be truncated at  $Y$  equal zero.

### 2.3.2 Lognormal Distribution

The density distribution of a random variable  $X$  is said to be lognormal distribution if its logarithm is a normally distribution. Suppose a random variable  $Y$  has a normal density distribution, and then  $X = \exp(Y)$  has a lognormal distribution. The log-normal distribution is very useful in modeling crop yields, income, and insurance claims. It is right-skewed, has a thick tail and fits many situations.

The lognormal probability density function is defined for  $y > 0$  in terms of parameters  $\mu$  and  $\sigma > 0$ . as follows:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2} \quad (2.2)$$

For modelling crop yields, Day (1965) recommended that the Gaussian distribution can be parameterized to yield the following versions

$$\log(x - \alpha_1) \sim N(\mu, \alpha), \quad x > \alpha_1 \quad (2.3)$$

$$\log(\alpha_2 - x) \sim N(\mu, \alpha), \quad x < \alpha_2 \quad (2.4)$$

$$\log\left(\frac{x - \alpha_1}{\alpha_2 - x}\right) \sim N(\mu, \alpha), \quad \alpha_1 < x < \alpha_2 \quad (2.5)$$

To determine the nature of unknown parameters  $\alpha_1$  and  $\alpha_2$ , Day suggested that a Pearson and Geary significant test should be performed to check the appropriate distributions among equations (2.3) – (2.5). The graphs of normal and lognormal distributions for different location and scale parameters are presented in Figure (2.1) and Figure (2.2)

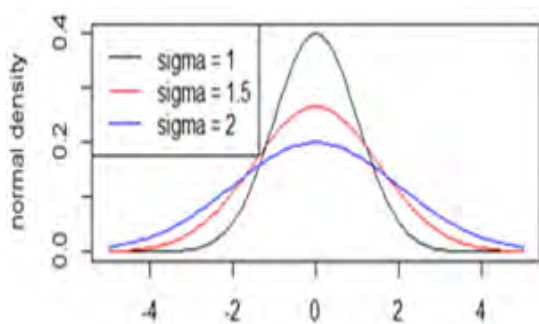


Figure 2.1: Normal Distributions

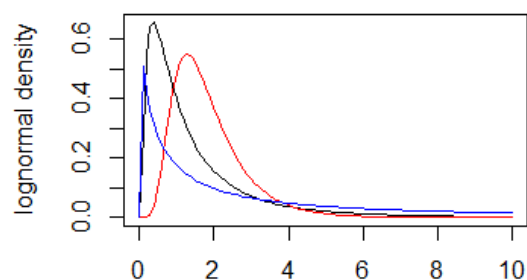


Figure 2.2: Lognormal Distributions

Figure 2.1 shows Normal probability density functions (pdfs) with parameters with location  $\mu = 0$  and different scale parameters  $\sigma = 1$  (black solid line),  $\sigma = 1.5$  (red dotted line), and  $\sigma = 2$  (blue dashed line).

Figure 2.2 shows log-normal probability density functions (pdfs) with parameters with location  $\mu = 0$  and  $\sigma = 1$  (black solid line),  $\mu = 0.5$  and  $\sigma = 0.5$  (red dotted line), and  $\mu = 1$  and  $\sigma = 2$  (blue dashed line). For a small  $\sigma = 0.5$  it resembles a normal distribution (see the right panel).

### 2.3.3 Gamma Distribution

In the literature, the density function of the Gamma distribution can be parameterized in various ways. In the context of regression analysis, the density is usually parameterized in terms of the mean and the scale parameters.

The density of a Gamma distribution for a random variable  $X$  is given by

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} \exp(-\frac{\alpha}{\beta} x)}{\beta^{\alpha} \Gamma(\alpha)} \quad x \geq 0 \quad (2.6)$$

The Gamma function  $\Gamma(\alpha)$  is given by  $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$

The mean and the variance of pdf (2.6) are obtained as  $E(x) = \frac{\alpha}{\beta}$  and  $Var(x) = \frac{\alpha^2}{\beta}$  and the distribution is simply written as  $x \sim \text{Gamma}(\alpha, \beta)$ .

As in the case of lognormal distribution, the gamma density can be reparameterized such that the shape  $\alpha$  and scale  $\beta$  are positive is given as

$$f(x) = \frac{\beta^{\alpha} x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)} \quad (2.7)$$

Equation (2.7) respectively yields the mean and variance as

$$E(x) = \alpha\beta, \text{ and } Var(x) = \alpha\beta^2 \quad (2.8)$$

where  $x \sim G(\alpha, \beta)$ . The gamma distribution is widely applied because it can assume a wide range of shapes. For example, the gamma distribution becomes an exponential distribution for  $\alpha=1$ , while the sum of  $n$  independent identically distributed (iid) exponential rv, with parameter,  $\beta$  has a gamma distribution, with parameters  $n$  and  $\beta$ . This flexibility results from gamma distribution with two parameters.

In most crop yield applications,  $\alpha$  and  $\beta$  are referred to as the shape and rate parameters respectively. For more applications of Gamma distribution, see Nelson and Preckel (2001). The plot for different shapes of Gamma distribution is presented in Figure (2.3)

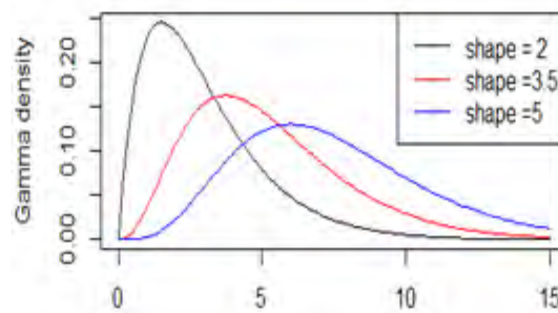


Figure 2.3: Gamma Distributions

#### 2.3.4 Weibull Distribution

Weibull distribution is also a generalized version of the exponential distribution. The full Weibull probability distribution for a random variable  $X$  having three parameters is given by

$$f(x; \gamma, \alpha, \beta) = \frac{\beta(x-\gamma)^{\beta-1}}{\alpha^\beta} \exp\left(-\frac{x-\gamma}{\alpha}\right)^\beta \quad (2.9)$$

where  $x > \gamma$ ,  $\alpha, \beta > 0$ .

Equation (2.9) becomes Gamma when  $\gamma = 0$ .

The probability distribution of a two parameter Weibull is given by

$$f(x; \alpha, \beta) = \beta\alpha^{-\beta}x^{\beta-1}\exp\left(-\frac{x}{\alpha}\right)^\beta \quad (2.10)$$

In the case of crop yield data, one can consider as a two-parameter Weibull distribution where the location  $\gamma = 0$  and thus,  $x \geq 0$  because yield data are strictly positive. Weibull distribution possesses many nice properties, it grants the estimation of higher moments and its density is bounded by zero.

The graph of the two parameter Weibull distribution is provided in Figure 2.4

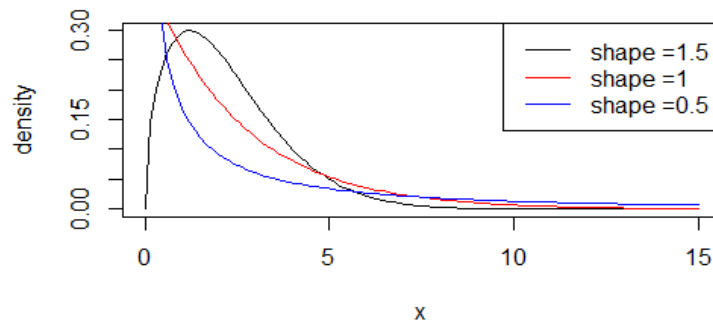


Figure 2.4: Weibull Distribution

## 2.4 Elicitation of Expert Probability Distribution in Agriculture

Expert opinion plays a major role in assessing problems for which data are lacking. Experts may have valuable knowledge about problems and possible solutions in their field. The elicitation and aggregation of the expert opinions may provide important knowledge to decision makers in many fields: Politics, Economics, Science and Technology.

As in many other applied sciences, Agriculture is an area where expert opinion has been successfully implemented to elicit probability distribution. Smith and Mandac (1995) reported the use of histogram method to elicit the distributions of rice yield from the Philippines rice farmers. They investigated some farmers, who are grouped into two, one cultivated a plain terrain and the second cultivated on the plateau. They had intended to use subjective elicitation method to determine the crop yield harvest for the next season and to evaluate the performance from three nitrogen fertilizer levels. They first determined the highest and lowest possible yield from each farmer and then partitioned the intervals into five equally spaced bins. The farmers were given equal number of chips and instructed to distribute them into bins in such a way as to represent the expected estimate of the crop yield into the intervals. The elicited farmer's distributions were compared with the estimates of econometric functions. The farmers' estimates from the plain were found to be closer to the the empirical model than those on the plateau. The overall estimated variances from the farmers' distributions were also found to be smaller than the estimate from the econometric model.

A similar study had been conducted by Grisley and Kellogg (1983) in Thailand. The rice farmers were given a monetary incentive (25 one baht coins) in place of the chip. They were asked to place the coins in the interval in which the correct value, they expected to get at harvest. Here, the farmers were not required to provide their opinions after a week instead of

immediately. Their estimates were then compared with the actual observed values from the farmer's field. It was concluded that they gave better accurate estimates for rice than for tobacco, soybeans and peanuts. The variability and uncertainty in the estimates was also found to be larger for the higher values of yield than for smaller values.

Recently, Sherrick (2002) elicited the farmers' probabilities of corn yield in response to weather variables from Illinois. He deduced the effect of past weather data on the corn yields and concluded that the farmers had poorly calibrated the crop yields. He concluded that the farmers had overestimated the probability of the extreme event, which might have adverse effect on the crop insurance deals.

Roosen and Hennessy (2001) used the histogram technique, where farmers' opinion were sought to elicit the percentage change in regional apple yield. McDaniels (1995) also used subjective elicitation as part of resource management in Fishery. He adopted the variable interval method to elicit the distribution of long-term equilibrium productivity. Despite all the usefulness of elicitation, it is worthwhile to note the criticism by Anand (1985) on the use of subjective probability in the context of agriculture. He identified three problems as a cause for concern: axiom violation, biased estimation and poor calibration.

## **2.5 Data Sources and Processing**

This section provides a description of the survey data used in this work. The datasets included a crop assessment survey of smallholder's farmers and expert opinion survey on agricultural production in Niger state, Nigeria.

### **2.5.1 Agricultural Survey of smallholder farmers**

The first data used in this work was obtained from an Agricultural crop survey conducted in Niger State, North Western part of Nigeria. The detailed description of the original survey was reported in Nmadu et al. (2012).

The target population is comprised of farming households in the study area clustered in farming communities (villages). The data collection was direct administration of questionnaire to the house head (the farmer). A structured questionnaire was administered to small-scale farmers between February and April, 2012. Other informal inquiry of data collection was also involved through discussion and personal interviews with household heads and the family members. The information was sought about agricultural productivity

on cereal crops such as sorghum, cowpea, millet and groundnut and the demographic characteristics, farm inputs and the farming practices of the study area. Other variables measured or observed were labour man-hours, total land cultivated, land acquisition and fragmentation, farm inputs, problems relating to pests, diseases and drought as well as farmers' adaptation to climate change, cultural and risk management practices.

For the purpose of this thesis, information on crop yields of sorghum and millet and farm inputs and socioeconomic characteristics of the farmers was used, as presented in Table 2.1.

**Table 2.1: Description of model variables**

<b>Terms</b>	<b>Description</b>
Crop yield	Total grain weight (kg) of harvested crop , shred and dried at the end of 2011/ 2012 growing season.
Farmers	Respondents (smallholders farmers) in the sampling survey A respondent is a person who provides answers the questionnaire.
Land size	Total acreage of land available to the farmer measured 1 <i>ha</i> = 10,000 <i>m</i> <sup>2</sup> a parcel of land for the crop(s)
Crop yield per hectare	Total yield weight (in <i>kg</i> ) divided by the total hectares cultivated by smallholders farmers
NPK fertilizer	Quantity (in <i>kg</i> ) of inorganic NPK fertilizer use to boost the soil nutrient
Seed quantity	Quantity (in <i>kg</i> ) of seed sown on entire parcel of land
Respondents	Number of smallholders farmers interviewed for a particular crop e.g. sorghum(n)=135, cowpea(n)=159 and millet(n)=95
Farm size	Hectares of land cultivated in the cropping season for a particular crop
Fertilizer used	Quantity of NPK fertilizer applied per hectare of land cultivated in the cropping season ( <i>kg/ha</i> )
Seed used	Seed per hectares of land cultivated in the cropping season ( <i>kg/ha</i> )

Soil supplements	Soil conservation strategy adopted by the farmer (dummy variable)  No additional soil supplement added ( <i>coded=0</i> ),  Solely organic strategy adopted (animal manure/poultry droppings by the farmer ( <i>coded =1</i> ), Solely inorganic supplement applied by the farmer ( <i>coded =2</i> ), Integrated soil supplement (organic and inorganic) adopted by the farmer ( <i>coded =3</i> )
Education level	Highest level of education attained by the respondent at the time of interviewing (dummy variable) by ranking the years spent in schools.  No formal education or Quranic School= 0- 2 years in school,  Primary School(1) = 2-7 years, High school (2) = 9-13 years, Degree/Diploma(3)= 14-17years,  Masters degree and higher Qualifications (4) = over 18years
Extension services	Binary variable (0=No contact by extension officer, 1= visited by extension officer)
Credit facility	Binary variable (0=no credit, 1= access to credit)
Farmer's experience	Number of years the respondent (farmer) has been involved in farming activities

### 2.5.2 Experts Survey Data

Expert opinion can be reported in either quantitative or qualitative form. Expert opinion is commonly characterised in form of distribution or empirical data from experiments or physical. The main difference between expert opinion and empirical data according to Cooke (1991) is that expert opinion is a form of personal opinion (e.g. subjective probabilities). He points out a difference between the subjective and normal probabilistic in the view on probabilities; but that this doesn't mean that the probabilistic view is more objective than the subjective view.

This thesis utilized two datasets collected through a designed questionnaire administered on agricultural experts as presented in Tables 2.2 and 2.3. The first data set was an evaluation of annual yield production of some cereal crops as provided by agricultural experts, while the second was the multiple experts' judgements on a particular crop, which was sesame seed production. The procedure of data collection from experts is described in Chapter 5 (Uncertainty theory). A sample questionnaire used in elicitation of information from experts is provided in Appendix B.

***Data set 1: Cereal Crop yields***

In November 2012 five agronomists (experts) are asked to complete an independent survey questionnaire for five cereal crop production. The following estimates were elicited on annual production of five cereal crops in Niger State, Nigeria and the associated degree of belief, denoted by  $(x_i, \alpha_i)$ . The expert data  $(x_i, \alpha_i)$  represents an ordered pair satisfying certain condition described in chapter 5, where  $x_i$  is the value of the uncertain variable about the crop yield, and  $\alpha_i$  is the degree of belief or intuition level attached to the  $i^{th}$  data point by the expert. The coded information is presented in Table 2.2.

**Table 2.2: Annual Estimates Crop yield (ton/ha) provided by agronomy experts**

Agronomists	Sample point				
	1	2	3	4	5
Maize	(1.8, 0.1)	(2.0, 0.5)	(2.5, 0.75)	(2.7, 1.0)	
Rice	(0.8, 0.0)	(1.0, 0.4)	(1.2, 0.70)	(1.5, 0.9)	(2.0, 1.0)
Sorghum	(0.8, 0.2)	(1.0, 0.5)	(1.2, 0.90)	(1.5, 1.0)	
Soybean	(1.0, 0.2)	(1.5, 0.5)	(1.8, 0.75)	(2.0, 1.0)	
Cowpea	(0.8, 0.0)	(1.0, 0.4)	(1.2, 0.80)	(1.5, 0.9)	(1.8, 1.0)

Source: Survey 2012

***Data set 2: Sesame Seed Production***

Five agricultural scientists were asked to complete a questionnaire on the sesame seed yield production in the last growing season of November 2011, in Niger State, North Western Nigeria. Each agronomist provided the following estimates (coded from their linguistic evaluation) of sesame seed yield on the basis of degree of belief and experience. At the same time, it was assumed that every respondent had fair knowledge about the subject matter, with each respondent assigned equal weight of 1/5. The information provided by the experts and

presented in Table 2.3 coded as  $(x_i, \alpha_i)$ . The coding shares the same meaning as defined in data set in Table 2.2 above.

**Table 2.3: Annual Sesame seed yield (*kg/ha*) from Experts domain**

Agronomists	Data Points				
	1	2	3	4	5
$A_1$	(260, 0.40)	(270, 0.7)	(275, 0.8)	(280, 0.95)	(290, 1.0)
$A_2$	(270, 0.40)	(275, 0.5)	(280, 0.6)	(290, 0.80)	(300, 1.0)
$A_3$	(250, 0.10)	(260, 0.3)	(270, 0.5)	(280, 0.70)	(290, 0.9)
$A_4$	(260, 0.20)	(270, 0.4)	(275, 0.6)	(280, 0.80)	(290, 1.0)
$A_5$	(230, 0.15)	(240, 0.4)	(250, 0.8)	(260, 0.95)	(270, 1.0)

Source: Expert survey 2012

Where  $A_i$   $i=1,2,3,4,5$  represents the  $i^{th}$  agronomist. From the data above, five empirical uncertainties distribution  $f_i(x), i = 1, \dots, 5$  can be constructed corresponding to each of the five agriculturists. The Delphi method discussed in chapter 5 was applied to obtain an optimal or consensus data for the composite uncertainty distribution.

### 3. Linear Regression Model and Inference

This chapter describes the method of classical estimation to a multiple linear regression model and its inference. The chapter later states some criteria for model selection and model fit.

#### 3.1 Linear Regression Model

A multiple linear regression model is widely used statistical tool to express the functional link between a response variable and explanatory variables called covariates or regressors. The regressors are assumed to have discernible association with the variable of interest (the output). Therefore, in many applied sciences, the interest is to evaluate the output from the control variables. The underlying functional relation between the output and influential variables can be expressed as a linear combination of the explanatory variables with appropriate weights called regression coefficients.

Suppose we denote the response variable by  $Y$  and the explanatory variables by  $X_1, \dots, X_n$ , then a general model can be written as

$$E[Y | X_1 = x_1, \dots, X_k = x_k] = f(x_1, \dots, x_k) \quad (3.1)$$

In particular, linear regression models are an important class of regression models in which the response variable is a linear combination of explanatory variables, that is,

$$f(x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Suppose  $E(Y) = f(x_1, \dots, x_k) = \eta$ , i.e.,  $Y$  fluctuates about an unknown parameter  $\eta$ ,  $Y = \eta + \varepsilon$ , where  $\varepsilon$  is the error. The focus is to model the linear relationship between  $Y$  and  $X_j; j = 1, \dots, k$ . Therefore, for observations  $(Y_i; X_{i1}; \dots; X_{ik}); i = 1, \dots, n$ , the model of Equation (3.1) can be expressed in the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (3.2)$$

where  $i = 1, \dots, n$ , or, equivalently in matrix notation,

$$Y = \mathbf{X}\beta + \varepsilon \quad (3.3)$$

where

- $\mathbf{X}$  is a non-stochastic  $n \times k$  matrix with  $k < n$ . The matrix  $\mathbf{X}$  has rank  $k+1$  that is,  $\mathbf{X}$  is of full column rank;
- the elements of the  $n \times 1$  vector  $\mathbf{Y}$  are observable random vectors;

- the elements of  $n \times 1$  vector  $\boldsymbol{\varepsilon}$  are non-observable random error such that  $E(\boldsymbol{\varepsilon}) = 0$  and  $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$  with  $\sigma^2 > 0$ . We can write  $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$

To obtain the estimation of parameters in model (3.3), there are two basic methods: ordinary least squares technique and maximum likelihood estimation.

The principle of the ordinary least squares method is perhaps the best known and most applied method for estimating the regression parameters. If we assume that the Equation (3.3) is solvable with respect to  $\boldsymbol{\beta}$ , then a solution  $\boldsymbol{\beta}^*$  clearly satisfies  $\|Y - X\boldsymbol{\beta}^*\|^2 = 0$ . On the other hand, if equation (3.3) is not solvable, then we can determine a vector,  $\widehat{\boldsymbol{\beta}}$ , which satisfies

$$\|Y - X\widehat{\boldsymbol{\beta}}\|^2 \leq \|Y - X\boldsymbol{\beta}_*\|^2 \quad (3.4)$$

for every vector  $\boldsymbol{\beta}_* \in R_k$ . Such a vector  $\widehat{\boldsymbol{\beta}}$  is called the least squares estimator for the parameters  $\boldsymbol{\beta}$  in Equation (3.3) because the condition (3.4) can be equivalently expressed as

$$\sum_{i=1}^n (Y_i - X_i \widehat{\boldsymbol{\beta}})^2 \leq \sum_{i=1}^n (Y_i - X_i \boldsymbol{\beta}_*)^2 = \sum_{i=1}^n \varepsilon_{*i}^2 \quad (3.5)$$

where  $\varepsilon_{*i} = Y_i - X_i \boldsymbol{\beta}_*$  is the  $i^{\text{th}}$  residual of the estimate of  $\widehat{\boldsymbol{\beta}}$ . Then the sum of squared residuals is minimized for  $\boldsymbol{\beta}_* = \widehat{\boldsymbol{\beta}}$ , so that  $\widehat{\boldsymbol{\beta}}$  has the smallest sum of squared residuals. To obtain  $\widehat{\boldsymbol{\beta}}$ , we can differentiate the function  $\|Y - X\boldsymbol{\beta}_*\|^2$  with respect to  $\boldsymbol{\beta}_*$  i.e.,

$$\frac{\partial}{\partial \boldsymbol{\beta}_*} \|Y - X\boldsymbol{\beta}_*\|^2 = 2X'X\boldsymbol{\beta}_* - 2X'Y \quad (3.6)$$

Setting equation (3.6) equal to 0 and solving for  $\boldsymbol{\beta}_*$ , we can obtain the least squares estimator as  $\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y$

Under the linear regression model with assumptions (i) to (iv), the function

$$f(\boldsymbol{\beta}_*) = \|Y - X\boldsymbol{\beta}_*\|^2 = (Y - X\boldsymbol{\beta}_*)'(Y - X\boldsymbol{\beta}_*)$$

is minimized for  $\boldsymbol{\beta}_* = \widehat{\boldsymbol{\beta}}$ , where  $\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y$ . The vector  $\widehat{\boldsymbol{\beta}}$  is called the ordinary least squares estimator of  $\boldsymbol{\beta}$ .

The least squares estimator  $\widehat{\boldsymbol{\beta}}$  has many favourable properties, including that it is the best linear unbiased estimator of  $\boldsymbol{\beta}_*$ . More details about linear regression models and related estimation methods can be found in, e.g., Grob (2003).

Another commonly used estimation technique is the maximum likelihood method. The likelihood of a set of data is the probability of obtaining that particular set of data, given the chosen probability distribution model. This expression contains the unknown model

parameters. The values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimators or simply MLE's. Maximum likelihood estimation is a totally analytic maximization procedure. It has some desirable properties for large sample:

- They become unbiased minimum variance estimators as the sample size increases.
- They have approximate normal distributions and approximate sample variances that can be calculated and used to generate confidence bounds.
- Likelihood functions can be used to test hypotheses about models and parameters.

For the linear regression model (3.3), let us assume the error terms  $\varepsilon_i$  are normally distributed, i.e.,  $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$ . Then, given the observations  $(\mathbf{Y}_i, X_i) = (Y_i, X_{i1}, \dots, X_{ik})$ ;  $i = 1, \dots, n$ , the likelihood function  $l(\beta, \sigma^2; \mathbf{Y}, X)$  for the regression model is the probability density function of  $\mathbf{Y}$

$$l(\beta, \sigma^2; \mathbf{Y}, X) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta\|^2\right\}$$

The log likelihood,  $l(\beta, \sigma^2; \mathbf{Y}, X)$ , is therefore

$$l(\beta, \sigma^2; \mathbf{Y}, X) = -\frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta\|^2 \quad (3.7)$$

By taking the derivative on both sides of Equation (3.7) and is written as

$$\frac{\partial}{\partial \beta} \|\mathbf{Y} - \mathbf{X}\beta\|^2 = \frac{1}{2\sigma^2} \{2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'\mathbf{Y}\} :$$

Setting  $\frac{\partial l}{\partial \beta} = 0$ , we obtain the maximum likelihood estimator of  $\beta$ , which exactly equals

the least squares estimator in this case.

Other common linear models include both one way and two way Analysis of Variance (ANOVA) models can be carried out and significance tests of the effects factors on the response variable. More reading about maximum likelihood inference in normal regression models can be found in (Long, 1997; Draper and Smith, 1998; and Fahrmeir et al. 1996).

### 3.2 Goodness of fit for Linear Models

The goodness-of-fit test determines whether the data conform to the model assumptions. For linear regression and other linear models, the main assumption is that  $Y_i \sim i.i.d. N(\mu_i, \sigma^2)$  and normality can be assessed by generating a quantile - quantile (QQ) plot of the residuals or errors,  $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ . The QQ-plot compares the ordered residuals to quantiles of the

standard Normal distribution. Further, model residuals are generally plotted versus fitted values  $\hat{Y}_i$  and versus individual covariates to check for homogeneity of  $\sigma^2$ .

There are several other hypothesis tests for normality and constant variance. These include the Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests which can be applied to model residuals to check for normality and the Brusch-Pagan and White tests for heteroscedasticity. For more detail, see Seber (1977) and Stewart (1977).

### 3.3 Model Selection

The most commonly used methods are stepwise forward selection or backward elimination (Draper and Smith, 1981), Akaike information criterion (AIC) (Akaike, 1973), Schwarz Bayes Criterion (SBC) (Schwarz, 1978), residual sum of squares (RSS), and various functions of Residual Sum Squares (RSS) such as  $R^2$  and adjusted  $R^2$  (Hocking, 1972).

While checking goodness-of-fit test, a modeller must ensure that factors that have no significant effects on the response should not be retained in the model. This process is generally known as model selection and deals with how to pick the best  $X\beta$ .

An ideal model will be parsimonious, that is, retaining fewest parameters that have significantly explanatory power in the model. In linear regression, the coefficient of determination,  $R^2$  can be calculated by

$$\begin{aligned} R^2 &= \frac{SS_{REG}}{SYY} = 1 - \frac{RSS}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \\ &= 1 - \frac{MSE}{s_y^2} \end{aligned} \quad (3.8)$$

where  $SS_{REG} = SYY - RSS$  and  $RSS = (y - X\hat{\beta})'(y - X\hat{\beta})$

Equation (3.9) measures the proportion of variation in the data that can be explained by the model. For competing models,  $R^2$  values can be directly compared. Further, for nested models, a likelihood ratio test can be performed to see if incorporating additional variables would provide enough explanatory power to make them worth adding to the model.

One drawback of  $R^2$  is that it will generally increase as the number of parameters increases. Thus, a related measure, adjusted  $R^2$  (Hocking, 1976) can be calculated such that

$$R^2_{adj} = 1 - (1 - R^2) \frac{n-1}{n-p+1} \quad (3.11)$$

This  $R^2_{adj}$  will only increase if a new term enhances the model more than what is expected just by chance. Other criteria such as Mallows's  $C_p$  and the PRESS statistic can also be used for choosing between models (Yan and Su, 2009).

Several quantities that have been proposed for estimating Mean Square Errors, goodness of fit and model selection are not entirely unrelated concepts. For example, if we consider two models, the smaller of which is supposed to be fully contained in the bigger, and we find the relation

$$\overline{R}^2_{p+q} < R^2_p \quad (3.12)$$

then the smaller model obviously shows a better fit to the data.

An abstraction somewhat related to Mallows's  $C_p$  is Akaike's (1974) Information Criterion (or AIC) which can be used contiguously to compare different models. AIC is defined as

$$AIC = -2 \log(\text{likelihood}) + 2p \quad (3.13)$$

AIC is essentially twice the negative log likelihood with a penalty for the number of parameters in the model. It maximizes model log likelihood value for fixed  $p$  and  $p$  is the number of explanatory variables in the model (including the constant). A small AIC value is better. AIC is very useful in comparing and selecting non-nested model specifications, but may be biased towards a model that overfits with extra parameters, for more detail, see Carlin and Louis (2000).

Similar to AIC, a Bayesian Information Criterion (BIC) has been developed for comparison between models, where BIC is

$$BIC = -2 \log(\text{likelihood}) + 2p \log(n) \quad (3.14)$$

By comparison to AIC, BIC takes the sample size  $n$  into account, and leans less quickly towards complex models as  $n$  increases. For more readings, see (Kuha, 2000; Spiegelhalter et al. 2002)

In other words, AIC and BIC are commonly used model selection criteria, which check overfitting the model. However, BIC penalizes free parameters more strongly than does the AIC. The optimal model is the one that tends to have a smallest discrepancy between its fitted values and true outcome probabilities, that is minimizes criterion.

## 4. Bayesian Statistics and Inference

This chapter first gives a detailed description of the Bayesian paradigm and related inference on the linear regression model under the classical assumptions. This chapter describes a procedure on “split” data analysis and introduced a leave-one-out cross validation method for model comparison. The chapter later introduces Bayesian Model Average and AIC weights for model comparison. The chapter lastly describes the Zellner  $g$ -prior approach for Bayesian variable selection.

### 4.1 Bayesian Inference

There are two well-studied approaches to perform probabilistic inference in the analysis of data. The classical frequentist approach views probabilities as relative frequencies of occurrence of random variables. This approach is often associated with the work of Neyman and Pearson who described the logic of statistical hypothesis testing. A Bayesian approach, in the other hand, regards any distinction between a random variable and model parameters as artificial, and treats all quantities as having an associated probability distribution, representing a degree of uncertainty.

In the Bayesian approach, the likelihood of observable  $y$  is specified by a joint conditional distribution on the parameters  $\beta$  and  $\sigma^2$  and is given as,

$$y | \beta, \sigma^2 \sim N(X\beta, \sigma^2 I) \quad (4.1)$$

The indeterminateness about the parameters can be represented by a utility function known as prior distribution for the parameters  $p(\beta, \sigma^2)$ .

Given a set of sample observations  $(y_i, x_i)$ , applying Bayes' theorem and prior distribution is updated by the empirical data as

$$p(\beta, \sigma^2 | y) = \frac{p(y | \beta, \sigma^2) p(\beta, \sigma^2)}{\int p(y | \beta, \sigma^2) p(\beta, \sigma^2) d(\beta, \sigma^2)} \quad (4.2)$$

Equation (4.2) is called the joint posterior distribution of the parameters  $(\beta, \sigma^2)$ . The denominator is a normalizing constant, which guarantees that the posterior density is a valid density function. Point estimates and confidence interval of the parameters can then be obtained from the joint posterior density and its associated marginal posterior for each of components of the parameters. A marginal posterior is derived by integrating other

components out of the joint posterior of equation (4.2). The posterior distribution takes the data into account; therefore, it produces less uncertainty than the prior distribution.

#### 4.1.1 Bayesian Inference of Linear Regression Model

Bayesian computation essentially involves derivations of probability distribution by solving integrals. In regression analysis, Bayesian method requires the prior distinct specification of the model parameters. The prior is chosen for mathematical convenience, such that it is a conjugate prior resulting in a posterior distribution from the same family, but prior needs not be conjugate in some cases. Most often, the integrals involve may be intractable, thus we appeal to analytical techniques such as numerical approximation or Laplace transforms. For more information Bayesian computation, see Carlin and Louis (2000).

The joint conjugate prior for  $(\beta, \sigma^2)$  has the structure

$$p(\beta, \sigma^2) = p(\beta | \sigma^2)p(\sigma^2) \quad (4.3)$$

where the conditional prior for the parameter vector  $\beta$  is the multivariate Gaussian distribution with mean  $b_0$  and covariance matrix  $\sigma^2 B_0$ :

$$p(\beta | \sigma^2) = N(b_0, \sigma^2 B_0) \quad (4.4)$$

and the prior for  $\sigma^2$  is the inverse gamma distribution with hyperparameters  $\nu_0$  and  $V_0$ :

$$p(\sigma^2) = IG(\nu_0, V_0) \quad (4.5)$$

Hence, the joint posterior distribution is given by

$$\begin{aligned} p(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) &\propto p(\mathbf{y} | \mathbf{X}, \beta, \sigma^2) p(\beta, | \sigma^2) p(\sigma^2) \\ &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)\right) \\ &\times (\sigma^2)^{-(k+1)/2} \exp\left(-\frac{1}{2\sigma^2} (\beta - b_0)' B_0 (\beta - b_0)\right) \\ &\times (\sigma^2)^{-(\nu_0+1)/2} \exp\left(-\frac{V_0}{\sigma^2}\right) \end{aligned}$$

Therefore, the joint posterior distribution of  $\beta$  and  $\sigma^2$  can be parameterized as follows:

$$p(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto p(\beta | \mathbf{X}, \mathbf{y}, \sigma^2) p(\sigma^2 | \mathbf{X}, \mathbf{y},) \quad (4.6)$$

where the two factors proportional to the product of a multivariate normal densities of  $\mathbf{N}(\mathbf{b}_n, \mathbf{B}_n)$  and Inv – Gamma( $v_n, V_n$ ) distributions, with the parameters of these given by

$$\mathbf{B}_n = (\mathbf{X}^T \mathbf{X} + \mathbf{B}_0^{-1})^{-1} \quad (4.7)$$

$$\mathbf{b}_n = (\mathbf{B}_n) \mathbf{X}^T \mathbf{y} + \mathbf{B}_0^{-1} \mathbf{b}_0 \quad (4.8)$$

$$v_n = v_0 + \frac{1}{2} n \quad (4.9)$$

$$V_n = V_0 + \frac{1}{2} (\mathbf{y}' \mathbf{y} + \mathbf{b}_0' \mathbf{B}_0^{-1} \mathbf{b}_0 - \mathbf{b}_n' \mathbf{B}_n^{-1} \mathbf{b}_n) \quad (4.10)$$

The error variance  $\sigma^2$  can be evaluated by integrating it out from the joint prior distribution of  $\boldsymbol{\beta}$  and  $\sigma^2$ , the resulting unconditional prior for  $\boldsymbol{\beta}$  is proportional to a multivariate Student distribution with  $2v_0$  degrees of freedom, location  $\mathbf{b}_0$  and dispersion matrix  $V_0 / v_0 \mathbf{B}_0$ . The detailed derivation can be found in Fahrmeir et al. (2009).

The density function of  $\boldsymbol{\beta}$  is given as

$$p(\boldsymbol{\beta}) \propto t(2v_0, \mathbf{b}_0, V_0 / v_0 \mathbf{B}_0)$$

Bayesian estimates the linear model (3.3) by specifying the prior  $(\boldsymbol{\beta}, \sigma^2)$  as a non-informative prior, which yields a constant prior for parameter  $\sigma^2$

$$p(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2} \quad (4.11)$$

#### 4.1.2 Prior Distribution for model Parameter

A prior is a probability distribution that epitomizes the uncertainty about the model parameter for the present data. In Bayesian analysis, prior specification is one of the most crucial issues and represents the information about an uncertain parameter. Bayesian approach combines the prior information and the likelihood of new data to yield the posterior distribution of the parameter, which in turn is used for prediction of the future values. Opponents of Bayesian approach criticize the arbitrariness in the choice of prior, while proponents praise it as a manageable way of introducing flexibility in Bayesian analysis. For more information on the choice of prior see Kass and Wasserman (1996).

Berger (2006) noted that whenever historical or subjective information can be incorporated in estimation of the unknown parameter, an informative prior is said to have been used. On the other hand, extracting prior from historical or subjective information might be difficult for a real problem, thus automatic or default prior distributions are needed. Non-informative priors

are sometimes used as default priors, vague priors, or priors of ignorance. Bayesian analysis with non-informative priors preserves the appearance of objectivity, and is being increasingly recognized by classical statisticians. The procedures of generating priors are discussed below.

### **Non-informative Priors**

The most commonly used approaches to develop a non-informative prior are discussed in Kass and Wasserman (1995).

#### **The Uniform Prior**

A natural idea for choosing a non-informative prior is a uniform prior. A uniform prior assumes that population mean is equally likely in a region for all items.

#### **The Jeffrey's Prior**

Jeffrey (1961) proposed in Jeffrey's prior as a solution to the problem of the uniform prior that does not yield an analysis invariant to parameter tractability.

#### **Conjugate Prior**

The conjugate prior approach was introduced in Raiffa and Schlaifer (1961). One must ensure posterior distribution is in the same family as the prior probability distribution, they are known to be conjugate distributions. For example, the Gaussian family is conjugate to itself (or self-conjugate) with respect to a Gaussian likelihood function. For a Gaussian likelihood function, one needs to choose a Gaussian prior to yield a Gaussian posterior distribution. Gelman et al. (2004) presents a catalog of these priors. Spiegelhalter, et al. (2003) studied the gamma distribution as a conjugate prior distribution for the precision of a normal distribution ( $\tau \sim \text{IG}(0.001, 0.001)$ ).

A brief list of priors and the corresponding posteriors is given in Table 4.1.

**Table 4.1:** List of Likelihood, prior distributions and their corresponding posteriors

<b>Likelihood</b>	<b>Prior</b>	<b>Posterior</b>
Normal (mean unknown)	Normal	Normal
Normal (variance unknown)	Inverse Gamma	Inverse Gamma
Poisson	Gamma	Gamma
Multinomial	Dirichlet	Dirichlet
Binomial	Beta	Beta
Negative Binomial	Beta	Beta
Geometric	Beta	Beta

## Expert Elicitation Prior

Beta distribution provides a flexible representation of expert opinion and it is also useful to model discrete phenomena such as defects, nonconformities in items, fraction or prevalence and uncertainty event. For example, given a number of trials  $n$  with recorded successes  $s$ . In these situations,  $\alpha$  is set to the value  $(s + x)$  and  $\beta$  is set to  $(n - s + y)$ , where  $\text{Beta}(x, y)$  is the prior.

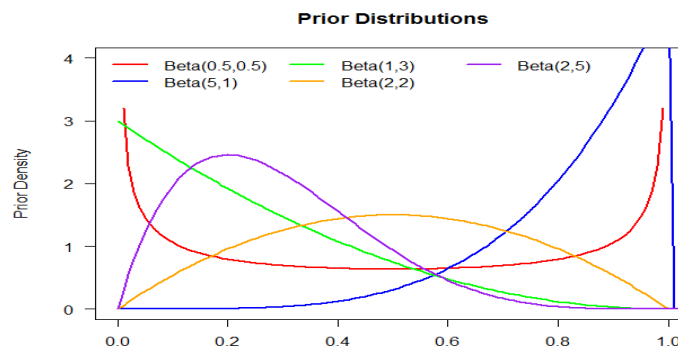


Figure 4.1: Plot of five different prior distributions

[  $\theta \sim \text{Beta}(.5, .5)$  :  $\theta \sim \text{Beta}(5, 1)$ :  $\theta \sim \text{Beta}(1, 3)$ :  $\theta \sim \text{Beta}(2, 2)$ :  $\theta \sim \text{Beta}(2, 5)$ ]

A  $\text{Beta}(1, 1)$  is equivalent to a uniform distribution in the unit interval and it is usually used as a non-informative prior.

(4.12).

### 4.1.3 Prior Specification in Linear Regression Model

For a Bayesian linear regression model (4.3), the following distributional specification can be assumed

$$\begin{cases} y | \sigma^2 \sim N(X\beta, 1/\sigma^2) \\ \beta_i | \nu \sim N_{k+1}(b_{0i}, B_{0ii} / \nu) \\ \nu \sim \Gamma(a, b) \\ \text{where } \nu = 1/\sigma^2 \end{cases} \quad (4.13)$$

where the variance parameter,  $B_{0ii}$  expresses the uncertainty about the parameter's location  $b_{0i}$ . If information about the process is rare, one chooses a large value for the variance parameter  $B_{0ii}$ . A prior distribution chosen this way is said to be flat or vague. In this situation, the regression coefficient values are far away from the mean  $b_{0i}$  and a reasonable prior and the exact distribution of  $b_{0i}$  may not have significant influence on the posterior. The natural normal –gamma conjugate prior chosen this way is mainly for mathematical

convenience and also has considerable flexibility to reflect the expert prior knowledge on the parameters.

On the other hand, if there is a reliable information about the coefficient  $\beta_i$ 's., then the choice of a small value for the variance parameter,  $B_{0ii}$  and a high probability is then assigned to values close to the mean  $b_{0i}$ . Sometimes, an estimate from the data can be used as hyperprior and then combined with likelihood to obtain a posterior mean far away from  $b_{0i}$ .

In the choice of the prior parameters,  $b_0$ ;  $B_0$  is the information matrix and the prior distribution has an impact on the posterior mean

$$b_n = (B_n)X^T y + B_0^{-1} b_0 \quad (4.14)$$

and the posterior covariance matrix

$$B_n = (X^T X + B_0^{-1})^{-1} \quad (4.15)$$

For a non-informative prior, the matrix  $B_0$  contains large values representing the uncertainty about the location,  $b_0$ . If information about the process is rare, then the posterior covariance matrix  $B_n$  is then approximated by  $\sigma^2 (X^T X)^{-1}$  and its mean by  $b_N \cong \sigma^2 (X^T X)^{-1} X^T y$ . This indicates that the vague prior information yields a posterior mean close to the OLS or ML estimator. On the other hand, if the substantial evidence about the coefficient vector is available, then the prior covariance matrix  $B_0$  contains small values, (that is the uncertainty is reduced). This yields the posterior covariance matrix  $B_n \approx \sigma^2 B_0$  and the mean  $b_N \approx b_0$ , i.e. the Bayesian estimator is similar to the prior mean (OLS estimates). More information can be found in Walter & Augustin (2010).

#### 4.1.4 The Posterior Distribution

The posterior density of a mean and variance for a normal sampling model is represented by the joint density of  $(\beta, \sigma^2)$  as the product of their densities:

$$g(\beta, \sigma^2 | y) = g(\beta | \sigma^2, y) g(\beta | \sigma^2) \quad (4.16)$$

The posterior distribution of the regression vector  $\beta$  conditional on the error variance  $\sigma^2$ ,  $g(\beta | \sigma^2)$  is the multivariate normal with mean  $\hat{\beta}_{OLS}$  and variance,  $V \sigma^2$ , where

$$\hat{\beta} = (X^T X)^{-1} X^T y, \quad V_\beta = (X^T X)^{-1} \quad (4.17)$$

Assume a prior with density distribution inverse Gamma  $IG(a, b)$ , this yields the marginal posterior distribution of  $\sigma^2$  equals  $IG\left(\frac{n-k}{2}, S/2\right)$ , where

$$S = (y - X\hat{\beta})^T (y - X\hat{\beta}) \quad (4.18)$$

#### 4.1.5 Bayesian Predictive Model

To predict a future observation  $\tilde{y}$  corresponding to a covariate vector  $x^*$  from the regression sampling model,  $\tilde{y}$  is conditioned on coefficient  $\beta$  and the variance  $\sigma^2$  is taken to be normally distributed  $N(x^*\beta, \sigma)$ . The posterior density of  $\tilde{y}$  is given as  $p(\tilde{y}|y)$  can be sampled from a mixture of joint densities  $p(\tilde{y}|\beta, \sigma^2)$ . This can be computed analytically by averaging over the posterior distribution of the parameters  $\beta$  and  $\sigma^2$  (i.e. integrating out with respect to  $\beta$  and  $\sigma^2$  as:

$$p(\tilde{y}|\beta, \sigma^2) = \int p(\tilde{y}|\sigma^2, y, \mathbf{X}, \mathbf{X}^*)p(\sigma^2|y, \boldsymbol{\beta}, \mathbf{X}^*)d\beta d\sigma^2 \quad (4.19)$$

#### Computations

The procedure to carry out the simulation is illustrated in Albert (2008). To carry out the simulation from the joint posterior distribution, the following steps are summarized from Albert (2008)

- The value of the error variance  $\sigma^2$  is first simulated from its marginal posterior density  $g(\sigma^2|y)$
- The coefficient  $\beta$  is then obtained from the conditional posterior density,  $g(\beta|\sigma^2, y)$

The algorithm is implemented in R environment with command function **blinreg** (see Appendix) to perform this simulation.

For a given vector  $x^*$  of covariates, suppose one intends to sample the mean response  $y$  at  $x^*$  given as

$$E(y|x^*) = x^*\beta \quad (4.20)$$

If  $\beta^*$  is a simulated draw from the marginal posterior of  $\beta$ , then  $x^*\beta^*$  will be a simulated draw from the marginal posterior,  $x^*\beta$ . The R function **blinregexpected** is implemented to simulate linear combinations of the regression coefficients. In the same manner, the posterior predictive distribution of future response values can be simulated using a similar algorithm.

Suppose  $\tilde{y}$  is a future response value corresponding to the row vector of covariates  $x^*$ . One simulates a single value of  $\tilde{y}$  by

- simulating  $(\beta, \sigma^2)$  from the joint posterior given the data  $y$
- simulating  $\tilde{y}$  from its sampling density given the simulated values of  $\beta$  and  $\sigma^2$

$$\tilde{y} \sim N(x^* \beta, \sigma^2) \quad (4.21)$$

This procedure can also be implemented in R with function **blinregpred** for set of covariate values of interest.

#### 4.1.6 Empirical Bayes Method

A statistical procedure, in which prior distribution is estimated from the data, is known as Empirical Bayes methods. Empirical Bayes is seen as an akin to perform a fully Bayesian analysis. In other hand, a hierarchical model facilitates the computation of the model parameters at the highest level of the hierarchy, instead of being integrated out. Empirical Bayes maximizes the marginal likelihood of the model, representing one approach for setting hyperparameters; see Bishop (2005) for more detail.

For example, in a two-stage hierarchical Bayes model with sample observations,  $y_1, \dots, y_N$ , one first generates an unobserved set of parameters  $\theta = (\theta_1, \dots, \theta_n)$  using its probability distribution,  $p(y|\theta)$  on the data. The parameters  $\theta$  can then be drawn from a population characterized by hyperparameters  $\eta$  according to a probability distribution  $p(\theta|\eta)$ . In such a case, Bayes model utilizes the parameter estimated at the first stage as the hyperparameters  $\eta$  and then used it to draw samples from a distribution  $p(\eta)$ , this is a direct opposite of the empirical Bayes approximation.

To implement the hierarchical procedure, the parameter of interest  $\theta_i$ , are obtained from both sample data and directly from the population of parameters  $\theta$ , inferring from the data as a whole, the hyperparameters  $\eta$  can be summarized using Bayes' theorem as

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)}{p(y)} \int p(\theta|\eta)p(\eta)d\eta \quad (4.22)$$

Since the integral (4.22) is not tractable analytically and must be evaluated by numerical methods. Stochastic approximations such as Markov Chain Monte Carlo sampling or deterministic approximations are commonly used. McCulloch et al. (2008) describes other techniques to obtain ML parameter estimates, including a Monte Carlo Newton-Raphson

method, a stochastic approximation (SA) algorithm, and a simulated maximum likelihood, where simulation is done to maximize the likelihood directly (as opposed to the log likelihood).

#### **4.1.8 Markov Chain Monte Carlo Sampling Method**

Berger, (2000) recently reported that the introduction of Markov chain Monte Carlo (MCMC) algorithm has greatly enhanced the computations of Bayesian analysis. The approach enables the user to integrate prior information into statistical models for decision making. The MCMC permits the analyst to combine prior information with the likelihood of the data resulting into posterior density of model parameters.

MCMC constructs a chain that is stationary, which converges to the target distribution. It constructs an irreducible Markov chain, which ensures that most of the Markov chains resulting from an MCMC algorithm are recurrent. The Markov chain converges to its stationary distribution for every starting value instead of almost every starting value. Thus, a recurrent state is needed to ensure that the MCMC algorithm converges. MCMC algorithms construct a transition kernel which results in a Markov chain which is recurrent and converges to the standard normal distribution. A general principle to do this is the Metropolis-Hastings (MH) algorithm. The Gibbs sampler is a special case of the MH algorithm. MCMC methods represent a type of bridge between traditional frequentist methods and Bayesian methods.

Geweke (1991) proposed a diagnostic test for assessing the convergence of the mean of each parameter. He considers the simulated Markov chain (obtained from the MCMC output) as a time series and applies a z-test to check whether the means from two different subsamples are equal. These subsamples come from the beginning and end of the generated chain. Typically, the first 10% of the chain is used as the beginning sample and the last 50% is used as the end sample. Using this z-test, parameters with  $|z| > 2$  indicates an evidence of significant differences between the means of the first and last set of iterations. This means a non convergence of the chains. Relevant literature related to convergence issues can be found in the following papers: (Geweke, 1992; Brooks and Robert (1998) and Cowles & Carlin (1996).

#### **4.1.9 Bayesian Credible Region Estimation**

The Bayesian approach to interval estimation allows the user to make probabilistic statements about how likely it is that the parameter  $\theta$  is in some intervals or a set. Suppose we want to

construct an interval  $S = (\theta_L, \theta_U)$ . The posterior probability that the parameter  $\theta$  is contained in that interval is

$$p(\theta \in S|y) = \int_{\theta_L}^{\theta_U} f_{\theta|y}(\theta|y)d\theta \quad (4.23)$$

The goal is to construct an interval  $S$  with a large (e.g. 95 %) percentage. In the literature, there are two most common methods for Bayesian credible interval estimation: Equal-tailed Intervals and Highest Posterior Density.

### Equal-tailed intervals:

A 95% equal tail probability interval uses the values that have cumulative probability of 0.025 and 0.975 as the endpoints. This is not preferred when the posterior is highly skewed or multimodal.

If we want a  $(1 - \alpha)100\%$  credible interval, then we select the bounds  $\theta_L$  and  $\theta_U$  such that

$$p(\theta \leq \theta_L) = p(\theta > \theta_U) = \frac{\alpha}{2} \quad (4.24)$$

### Highest Posterior Density (HPD) intervals:

The HPD credible region is familiar in the Bayesian literature; see for example Park & Casella (2008).. HPD sets (intervals) contain all values of the parameter  $\theta$  such that the posterior density  $f_{\theta|y}$  is larger than some constant  $C_\alpha$ , where  $C_\alpha$  ensures that the coverage probability will be  $(1 - \alpha)$  posterior probability.

The distribution function of HPD regions can be derived on sub-vectors of the parameter  $\beta$  from the marginal posterior distribution of  $\beta$ . For a single parameter,  $\beta_j$  has a Student t-distribution as

$$\beta_j | y, X \sim \mathcal{T} \left( n, \frac{c}{c+1} \left( \frac{\tilde{\beta}_i}{c} \right) + \hat{\beta}_i, \frac{c(s^2 + (\tilde{\beta}_i - \hat{\beta}_i)^T X^T X (\tilde{\beta}_i - \hat{\beta}_i) / (c+1))}{n(c+1)} h_{ii} \right) \quad (4.28)$$

where  $h_{ii}$  is the  $ii$  element of the matrix  $(X^T X)^{-1}$

$$\text{If } \tau = \frac{\tilde{\beta} + c\hat{\beta}}{c+1}$$

$$K = \frac{c(s^2 + (\tilde{\beta}_i - \hat{\beta}_i)^T (X^T X)^{-1} (\tilde{\beta}_i - \hat{\beta}_i))}{n(c+1)} = K_{(ij)}, \text{ then transform}$$

$\tau_i = \frac{\beta_i - \tau_i}{\sqrt{K_{(ij)}}}$  has a standard Student  $t$  –distribution with a  $n$  degree of freedom. A  $(1 - \alpha)$

HPD interval on  $\beta_j$  is given by

$$\tau_i - \sqrt{K_{(ii)}}F^{-1}\left(1 - \frac{\alpha}{2}\right) + \beta_j, \tau_i + \sqrt{K_{(ii)}}F^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (4.29)$$

## 4.2 Analytical Procedure for Split Data

Using the present farming practice and data at hand, we speculate that the farmer wants to move to a new virgin land or cultivate a new genetically modified crop. His goal is to produce at an optimum yield performance in the new location or by adopting a new farming technology when there is limited data available. Expert knowledge from the current location or the present farming practice has to be incorporated into the model at the new location or under the new farming technology. It is assumed that there would be a change in the farmer characteristics and soil management in the new location.

The farming practice in the present location is assumed to be exactly the same as in the new location. Therefore, the exact same variables would be considered to model crop yield production on the farmer's new field. To simulate this situation, the survey data set was split as follows:

- 50% of the observations were randomly selected and labelled as the set of observations that were old. These observations were deemed to come from the current or present farming region or location.
- The remaining 50% of the observations were randomly selected and labelled as the set of observations that were new. These observations are deemed to come from the new location representing agronomic characteristics of the new area. The estimates from the old data set were used as prior information and the new dataset was taken as data from the new farming practice. These observations were assumed to come from the new location and were used to assess the performance of the old which were fitted on the limited amount of data from the new location.

The following procedures were then undertaken:

- The data set was first checked and cleaned. This task involves removing outliers and estimating missing values (if any), multi-collinearity, influential observations etc.
- A linear regression model was fitted to the old data set. The coefficients here were used as prior information when the new farming practice was either adopted or the farmer intended to cultivate a new crop variety. Three models were applied:

- M1: A linear regression model was fitted to the new data.
- M2: A Bayesian linear regression model with non-informative prior fitted to the new data.
- M3: A Bayesian linear regression model fitted to the new data using the coefficients from the old data set as priors.
- The predictive performance of the models was assessed using Leave one out cross validation on the whole dataset.

### **4.3 Leave-One-Out-Cross-Validation (LOOCV)**

Nowadays, it is preferable to use Cross validation algorithm to evaluate a model than ordinary inspection of the residuals. Cross validation addresses the 'over-fitting' problem. Over-fitting arises from the same data being used to fit or specify the model, and assess that model's fit. The drawback of residual evaluations is that it does not give a clear indication of the predictive performance of the model. This problem can simply be surpassed to avoid using the whole data set when building a model. Before the model is built, one excludes some of the data. Once the model has been built, the data that was excluded is then used to validate the model to predict new values. This process of validating a model is known as cross validation.

The LOOCV ordinarily takes one observation out of the data and sets it aside as the 'testing set'. Then, the model is applied to the training set of  $(n - 1)$  cases (i.e. the data minus the single case) and the resulting coefficients are applied to the testing set case to produce a predicted value which in turn is then compared with the actual value of that single case. The data set is now separated into two sets, called the training set and the testing set. The model is fit using the training set only. Then the model is used to predict the outcome values for the data in the testing set. The errors are accumulated to give the mean absolute test error, which is used to evaluate the model. For more details on the cross validation techniques, see (Efron & Tibshirani, 1997; Harrell, 2001).

The procedure is implemented in Chapter 7.

### **4.4 Bayesian Model Selection and Model Average**

In this section, we introduce Bayesian model average and model selection with Zellner's  $g$ -prior approach. The methods were performed through MCMC sampling procedures for model selection. Bayesian variable selection methods come equipped with natural measures

of uncertainty, such as the posterior probability of each model and the marginal inclusion probabilities of the individual predictors.

#### 4.4.1 Bayesian Model Averaging (BMA)

Bayesian Model Averaging (BMA) is an approach to address model uncertainty which allows us to assess the robustness in a canonical regression problem. The approach provides alternative specifications by calculating posterior distributions over coefficients and models. BMA produces a posterior probability for each possible model in addition to the posterior probability for each predictor (Fridley, 2009).

This posterior can be used to simply select the “best” model (usually the one with highest posterior probability). BMA performs better than adopting a single best model for prediction. BMA combines all possible explanatory variables to sample models and computes the weight of each model as the posterior model probabilities. These methods enhance the predictive performance of the models and builds models empirically from the data. For details see Min and Zellner (1993); Raftery et al. (1997).

Suppose a vector of explanatory variable  $\alpha_\gamma$  is a constant,  $\beta_\gamma$  is the coefficients and  $\varepsilon$  is a normal identically independently error term with variance,  $\sigma^2$ :

$$y = \alpha_\gamma + X\beta_\gamma + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \quad (4.30)$$

A problem arises when there are many potential explanatory variables  $X$ : Which variables  $X_\gamma \in \{X\}$  should be then included in the model and how important are they? The direct approach which is to do inference on a single linear model that includes all variables will be inefficient with a limited number of observations.

BMA approach thus tackles the problem by estimating models for all possible combinations of  $X$ ; it results into a weighted average over all generated models. If  $X$  includes  $k$  potential variables, this means, we are estimating  $M = 2^k$  models as possible combinations of predictors. The model weights for this averaging stem from posterior model probabilities that arise from Bayes' theorem:

$$\begin{aligned} p(M_\gamma | y, X) &= \frac{p(y | M_\gamma) p(M_\gamma)}{p(y_\gamma | X)} \\ &\propto p(y | M_\gamma, y, X) p(M_\gamma) \end{aligned} \quad (4.31)$$

Here,  $p(y_\gamma|X)$  denotes the integrated likelihood which is constant over all models and is thus simply a multiplicative term. Therefore, the posterior model probability (PMP) is proportional to the marginal likelihood of the model  $p(y_\gamma|M_\gamma, y, X)$  times a prior model probability  $p(M_\gamma)$ . By normalizing this product, one can infer the PMPs and model weighted posterior distribution for any statistic  $\theta = \alpha_\gamma$  coefficients:

$$p(\theta|y, X) = \sum_{\gamma=1}^M p(\theta|M_\gamma, y, X) \frac{p(M_\gamma|y, X)p(M_\gamma)}{\sum_{j=1}^M p(M_j|y, X)p(M_j)} \quad (4.32)$$

The model prior  $p(M_\gamma)$  has to be elicited by the researcher or expert and should reflect prior beliefs. A common choice is to set a uniform prior probability for each model  $p(M_\gamma) \propto 1$  to represent the lack of prior knowledge, as in a similar coefficient estimation defined in earlier sections. A limitation of the Bayes factor approach is for large  $k$ , computing the BIC for all  $2^k$  models is impossible. For more information on BMA, consult Hoeting et al. (1999) and for alternative Bayesian variable selection, see Madigan and Raftery (1994).

#### 4.4.2 Bayes Factor

The Bayes factor is an essential measure for model comparison and variance selection in Bayesian regression analysis. Suppose that the observed data  $y$  could have been generated under one of two models  $M_1$  and  $M_2$ . It is natural to ask from the Bayesian perspective “what is the posterior probability that  $M_1$  is true (assuming either  $M_1$  or  $M_2$  is true)?”.

Using Bayes theorem, we can calculate by:

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{p(y|M_1)p(M_1) + p(y|M_2)p(M_2)} \quad k = 1, 2 \quad (4.33)$$

The key quantities here are  $p(y|M_1)$  and  $p(y|M_2)$  which are called marginal likelihoods or integrated likelihoods.

The extent to which the data supports model  $M_1$  over  $M_2$  is measured by the posterior odds of  $M_1$  against  $M_2$ . This yields the ratio of their posterior probabilities. From equation (4.33), this leads to :

$$B_{12} = \frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \times \frac{p(M_1)}{p(M_2)} \quad (4.34)$$

The first factor on the right-hand side of equation (4.33) is the ratio of the integrated likelihoods of the two models and is called the Bayes factor for  $M_1$  against  $M_2$ , denoted by  $B_{12}$ . The second factor on the right-hand side of (4.33) is the prior odds, and this often equals to 1, representing the absence of a prior preference for either model, that is  $p(M_1) = p(M_2) = 1/2$ . Hence, we can write equation (4.33) as:

$$\text{Posterior odds} = \text{Bayes factor} \times \text{prior odds} \quad (4.35)$$

It follows that the Bayes factor is equal to the posterior odds when prior odds are equal to 1. Equation (4.35) can be implemented in R using **MCMCpack** written by Martin and Quinn (2006). Since the posterior odds equal the Bayes factor when the models are equally likely *a priori*, the Bayes factor is a measure of how much support is available in the data for one model relative to another. When  $B_{12} > 1$ , the data favour  $M_1$  over  $M_2$ , and when  $B_{12} < 1$ , the data favour  $M_2$  over  $M_1$ .

For example, if  $B_{ij} = 5$ , then model  $M_i$  is five times more likely than model  $M_j$ , given the data.: Jeffreys recommended a scale of evidence for interpreting Bayes factors. An interpretation of the strength of evidence for Bayes factor B is given in Table 4.1

**Table 4.2: Interpretation of Evidence for Bayes factor**

Bayes Factor	Strength of Evidence
$-\infty < B_{12} \leq 0.1$	Strong evidence against $M_1$
$0.1 < B_{12} \leq (1/3)$	Moderate evidence against $M_1$
$(1/3) < B_{12} < 1$	Weak evidence against $M_2$
$1 < B_{12} < 3$	Weak evidence for $M_1$
$3 \leq B_{12} < 10$	Moderate evidence in favour of $M_1$
$10 \leq B_{12} < \infty$	Strong Evidence in favour of $M_1$

The above interpretation is provided by Jeffreys (1961), though he included two additional categories: very strong for  $B_{12} > 30$ , and decisive for  $B_{12} > 100$ . Other interpretations of Bayes factors have been proposed as well.

The Bayes factor plays a vital role in Bayesian model selection in regression analysis. The mathematical relation was introduced by Schwarz (1978). He showed that the Bayesian Information Criterion (BIC) can be approximated by Bayes factor for a large  $n$  as

$$-2\log(BF) \approx \Delta BIC = W - \log(n) (|M_1| - |M_2|) \quad (4.36)$$

where  $\Delta BIC$  is the change in BIC from model 1 to model 2, where

$$W = -2 \log \left[ \frac{\sup_{\gamma(1)} p(y|\beta_{\gamma(1)}\sigma^2)}{\sup_{\gamma(2)} p(y|\beta_{\gamma(2)}\sigma^2)} \right] \quad (4.37)$$

Equation (4.37) is the usual likelihood ratio statistic and

$$|\gamma| = \sum_{j=1}^p \gamma_j \quad (4.38)$$

Equation (4.38) is the number of parameters included in the model denoted by  $\gamma$ . The BIC approximation can be used to compute the posterior probability of each model.

Raftery (1995) shows that BIC approximation can be used in place of the Bayes factor  $B_{12}$ .

This comparison is expressed on the scale of twice the logarithm:

$$2 \log B_{12} = 2 \log p(y|\hat{\theta}_1, M_1) - \log p(y|\hat{\theta}_2, M_2) - (p_1 - p_2) \log(n) + O(n^{-1/2}) \quad (4.39)$$

Bayes factors can be computed using only the maximum of the likelihood function, and the comparisons often relate closely to classical frequentist tests. Thus, when  $M_2$  is nested within  $M_1$ , the log Bayes factor becomes

$$2 \log B_{12} \approx \chi^2_{12} - df_{12} \log(n) \quad (4.40)$$

$\chi^2_{12}$  is the standard likelihood-ratio tests for testing  $M_2$  against  $M_1$  and  $df_{12} = (p_1 - p_2)$  is the number of degrees of freedom associated with the test.

#### 4.4.3 AIC Weights for Model Comparison

In a general linear regression model, one assumes that there is no true model. In fact, the reality in the data can only be approximated by the models. That is one need to find the model that would best approximate the observed data. In other words, the best model can be constructed by minimizing the loss of information between the generic distribution and the proposed model. Kullback and Leibler (1951) addressed such issues by employing information theory. He developed a divergence measure known as the Kullback-Leibler information to represent the information loss when one approximates the data realization.

Burnham and Anderson (2001) proposed Akaike weights for model selection. The aim is to assess which of the candidate models are usually adequate for the data. Then, among the models, which one should be chosen as the basis for interpretation, prediction, or other subsequent use in the process of data realization. They established a functional link between

the maximum likelihood of the model and the information theory. They proposed the use of Kullback-Leibler information through Akaike's information criterion (AIC), which is defined as

$$AIC = -2 \log(\hat{\theta}) + 2p$$

where  $p$  is the number of estimated parameters in the model (include the intercept), and  $\hat{\theta}$  is the likelihood of the model given the data. Burnham and Anderson (2002) suggested that other penalized criteria such as corrected AICC<sub>c</sub>, BIC and their modifications can be chosen as suitable selection criteria.

From OLS regression output, AIC can easily be computed by

$$AIC = n * \log(\hat{\sigma}^2) + 2*p$$

where  $\hat{\sigma}^2 = \frac{RSS}{n}$  is the deviance ; RSS, is residual sum of squares and  $n$  is the sample size.

It is worthwhile to note here that because the variance is estimated, it must be added to the number of estimated parameters  $p$ .

In order to compare any two models with their associated AIC measures, one selected the model with smaller AIC values and compute the difference called the delta AIC or the corresponding Akaike weights. The delta AIC,  $\Delta_i$  measures a model relative to the best model, and is calculated as

$$\Delta(AIC) = \Delta_i = AIC_i - \min(AIC)$$

where  $AIC_i$  is the AIC value for model  $i$  and  $\min(AIC)$  is the AIC value of the best model. Burnham and Anderson (2002) gave the interpretation as relative measure of evidence between the models. For instance, for  $\Delta_i < 2$ , it suggests substantial evidence for the model, for values between 3 and 7, it indicates that the model has considerably less support, whereas for  $\Delta_i > 10$ , it indicates that the model is very unlikely. This is similar to Jeffery (1961) interpretation of Bayes factor, Table 4.1 above.

Alternatively, Akaike weights,  $w_i$ , which provides another measure of the strength of evidence for each model, and represent as the ratio of  $\Delta_i$  values for each model relative to the whole set of all  $R$  candidate models:

$$\text{Akaike Weight} = w_i = \frac{\exp(-0.5\Delta_i)}{\sum_{i=1}^R \exp(-0.5\Delta_i)}$$

By changing the scale of the  $\Delta_i$  enables the user to compare the model on a scale of 1 (i.e., so that the sum of the  $w_i$  equals 1. The Akaike weight  $w_i$  is easily interpreted as it indicates the probability that the model is the best among the whole set of candidate models given the data.

The use of information criteria in model selection and further extension on derivation of Akaike weights can be found in Bozdogan (1987); Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004).

#### 4.4.4 Bayesian Model Selection using Zellner's $g$ -prior

Zellner (1986) introduced an attractive method of implementing prior information into a regression model. He assumes  $(\beta|\sigma^2)$  is normal with mean  $\beta_o$  and variance-covariance matrix of the form

$$v = c\sigma^2(X'X)^{-1} \quad (4.41)$$

and then takes  $\sigma^2$  distributed according to the non-informative prior proportional to  $\sigma^2$ . This is called Zellner's  $g$  prior. The value  $c$  is a scale parameter for the variance covariance structure.

One nice thing in Zellner's  $g$  prior is that it requires only two prior inputs from the user: (1) a choice of the prior mean  $\beta_o$ , and (2)  $c$ , is a measure of uncertainty in the prior, which is the amount of information that can be attributed to the sample data.

To implement Zellner's  $g$  - prior for model selection in a linear regression analysis with  $k$  predictors, there are  $2^k$  possible models corresponding to the inclusion or exclusion of each predictor in the model. A  $g$ -prior can be placed on the full model that contains all parameters. By assuming  $\beta_o$  as concentrated toward zero vectors (sometime called zero mass point) and we choose  $c$  to be a large corresponding to a large variance representing prior beliefs. The approach provides a simple way of inputting subjective information in a regression model. The particular choice of distribution is called a  $g$ -prior.

What follows is an illustration on the use of Zellner's  $g$ -prior. In a linear regression model with many explanatory variables, Zellner's  $g$ -prior provides a convenient way of selecting the optimal model from all possible combination of predictors.

Given a  $g$ -prior, one assumes that the regression vector  $\beta$ , conditional on  $\sigma^2$ , has a multivariate normal prior distribution with mean  $\hat{\beta}$  and variance-covariance matrix  $c\sigma^2(X'X)^{-1}$ , and then assign  $\sigma^2$  a standard non- informative prior proportional to  $1/\sigma^2$ . A good guess of the prior is by choosing a small value for  $c$ . In contrast, choosing a large value of  $c$  would have an effect similar to choosing the standard non-informative prior for  $(\beta, \sigma^2)$ .

The core idea in the Zellner's  $g$ -prior model selection process is to allow the analyst or expert to introduce information about the location of the regression by bypassing the most difficult aspects of the prior correlation structure. The structure is fixed in Zellner's proposal since the prior corresponds to

$$\beta | \sigma^2, X \sim N(\beta^0, c\sigma^2(X'X)^{-1})$$

and

$$(\sigma^2 | X) \propto 1/\sigma^2$$

The prior relies on a conditional Gaussian prior for  $\beta$  and the improper (Jeffreys' prior) for  $\sigma^2$ . The experimenter thus restricts prior determination to the choice of  $\beta^0$  and constant  $c$ . It should be noted that  $c$  is interpreted as a measure of the amount of information available in the prior relative to the sample observation. For example, setting  $1/c = 0.5$  gives the prior the same weight as 50% of the sample.

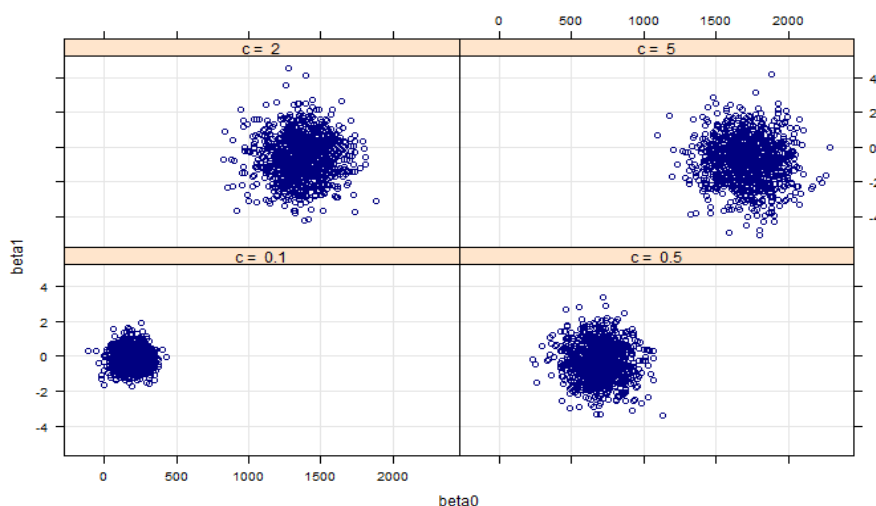


Figure 4.2: Effect of constant  $c$  on posterior estimates using Zellner's  $g$ - prior

Given a prior distribution, the posterior distribution is obtained as

$$p(\beta, \sigma^2 | y, X) \propto p(y | \beta, \sigma^2 X) p(\beta, \sigma^2 | X)$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{\sigma^2}(y - X\beta)^T(y - X\beta) - \frac{1}{\sigma^2}(\beta^0 - \hat{\beta})^T X^T X(\beta^0 - \hat{\beta})\right] (\sigma^2)^{-k/2}$$

$$\times \exp\left(-\frac{1}{\sigma^2}(\beta^0 - \beta)^T X^T X(\beta^0 - \beta)\right)$$

Thus, the posterior distribution of the regression vector  $\beta$  conditional on  $\sigma^2$ ,  $p(\beta | \sigma^2)$  is multivariate normal with mean  $\beta^{(1)}$  and variance-covariance matrix  $V_1$

$$\beta^{(1)}, \sigma^2 | y, X \sim N_{k+1} \left( \frac{c}{c+1} \left( \frac{\beta^0}{c} + \hat{\beta} \right), \frac{c\sigma^2}{c+1} (X^T X)^{-1} \right) \quad (4.42)$$

Hence, the posterior marginal density of  $N(a_1, b_1)$  is given as

$$a_1 = \frac{c}{c+1} \left( \frac{\beta^0}{c} + \hat{\beta} \right), \quad b_1 = \frac{c\sigma^2}{c+1} (X^T X)^{-1} \quad (4.43)$$

Similarly, the posterior of variance-covariance is inverse-Gamma,  $IG(a_n, b_n)$

$$\sigma^2 | y, X \sim IG \left( \frac{n}{2}, \frac{S^2}{2} + \frac{c}{2(c+1)} (\beta^0 - \hat{\beta})^T X^T X (\beta^0 - \hat{\beta}) \right) \quad (4.43)$$

From the joint posterior distribution of the two parameters, the variance is first simulated from the inverse gamma distribution. The regression coefficient can be simulated since  $\beta$  has a conditional multivariate normal density. By implementing a function **blinreg** in R environment with prior option, Zellner's  $g$ -prior produces marginal posterior estimates and log-likelihood of the models. The approach is demonstrated in Appendix G.

For more information on Bayesian variable selection built on Zellner's  $g$ -priors, see (George and Foster, 2000; Chipman, et al. 2001).

## 5. Foundation of Uncertainty Theory and Estimation Procedures

The chapter provides some useful conceptual definitions, axioms and notions in uncertainty statistics. The chapter defines the notion of an uncertain variable, uncertainty distribution and uncertainty measures (i.e. uncertain expected value and variance). The chapter establishes the fundamental concept of uncertainty theory and makes comparison with probability measure and credibility measure theory. Some uncertainty distribution functions and procedures for data collection from experts are discussed. The concept of uncertain distance is defined. The distance concept is then extended to construct of an uncertainty distribution for difference of two uncertainty distributions.

### 5.1 Review of Uncertainty Statistics

Uncertainty theory is a branch of mathematics based on normality, self-duality, countable sub-additivity and product measure axioms, the theory was founded by (Liu, 2007) and later refined (Liu, 2010) to model imprecise quantities specified by human beings, such as “approximately 50 kg”, “nearly 1000 km”, “small size”, “tall man”. These quantities cannot be described by random phenomena (probability theory) or fuzzy set theory (complete subjective axioms). Uncertainty usually arises in real world situations where data on a particular subject matter is insufficient to support probability frequency distribution, hence the researcher tends to rely on experts/ judgment or belief that particular events will occur. Human thinking tends to exaggerate the frequency of unlikely events; hence the belief degree tends to significantly differ from the real frequency as exhibited in probability theory. Such a situation can make researchers or modellers to jump into misleading conclusions. Hence, such situations are better dealt with by Liu’s uncertainty theory.

In classical statistics such as in probability theory, inputs are single values which are treated as random variables following a probability distribution. The outcome is a risk distribution. A distinction ought to be made between uncertainty and inherent variability. Variability represents heterogeneity or diversity, which is not reducible through further measurement or study. Uncertainty represents ignorance about a poorly characterized phenomenon which is only sometimes reducible through further measurement or study. Conventionally, a random variable is characterized by objective uncertainty and governed by a random phenomenon, which is a function or mapping from a probability space to the set of real numbers

(Kolmogorov, 1933). Liu and Xu (2010) stated that probability measure is a set function satisfying axioms of normality, non-negativity and countable additivity. The theory has successfully been applied to deal with objective uncertainty phenomena. The random uncertainty addressed by probability theory is merely one of those phenomena, so it is not able to handle other forms of uncertainties.

Due to the drawback of probability theory, Zadeh (1965) proposed the concept of fuzzy set theory which tends to model subjective uncertainty via fuzzy membership functions. However, fuzzy measure does not possess self-duality, thus the possibility of an event occurring or not can be assigned an equal possibility measure of one each. For example, fuzzy measure assigns equal probability of occurrence and non-occurrence to an event. Probability theory and fuzzy set theory are two extremes of uncertainty measure, thus probability theory requires strictly complete additivity while the fuzzy set theory is characterized by non-additivity.

Uncertainty theory is defined based on an axiomatic system which includes self-duality, thus uncertainty theory satisfies the law of contradiction and excluded middle that is consistent with the reality of human thinking. Furthermore, uncertainty theory allows neither a completely additive nor completely non-additive but has a sub-additivity property which is perceived as consistent with real world applications (Guo, 2010).

The next section seeks to describe these key concepts and the distributional properties of an uncertain variable.

## **5.2 Uncertainty Measures and Relations**

In measure theory, an uncertain random variable is characterized by chance measure (Liu, 2008). There are three concepts of uncertainty measure, are characterized by their respective variables as randomness, uncertainty (impreciseness) and fuzziness. A variable that exhibits randomness and fuzziness is special cases of hybrid variables and categories of uncertain variable.

Researchers nowadays encounter such imprecise quantities and try to develop a mathematical procedure to model such processes. How to deal with such subjectivity uncertainty has been a challenge for decades. Quantities that exhibit such characteristics of both subjective and objective uncertainty have not been clearly defined by human perception or judgment. An uncertain variable can be used to describe an imprecise quantity but not by probability nor by

fuzzy theory. Liu (2007) stated that a measurable function that maps from uncertainty space to a set of real numbers is an uncertain measure. Table 5.1 illustrates the relationship between probability and uncertainty theory, while Table 5.2 illustrates the relationship between credibility and uncertainty theory.

Table 5.1: Relationship between Probability theory versus Uncertainty theory

<b>Probability Axioms</b>	<b>Uncertainty Axioms</b>
Normality	Normality
Monotonicity	Monotonicity
Self- Duality	Self -Duality
<b>Additivity</b>	<b>Countable Sub-additivity</b>
$\Pr\{A \cup B\} = \Pr(A) + \Pr(B)$	$M\{A \cup B\} \leq M(A) + M(B)$
<b>Product Measure</b>	<b>Product Measure</b>
$\Pr = Pr_1 \times Pr_2 \times \dots \times Pr_n$	$M = M_1 \wedge M_2 \wedge \dots \wedge M_n$

Table 5.2: Relationship between Credibility theory and Uncertainty theory

<b>Credibility Axioms</b>	<b>Uncertainty Axioms</b>
Normality	Normality
Monotonicity	Monotonicity
Self Duality	Self- Duality
<b>Maximality</b>	<b>Countable Sub-additivity</b>
$\text{Cr}\{A \cup B\} = \text{Cr}(A) \vee \text{Cr}(B)$	$M\{A \cup B\} \leq M(A) + M(B)$
<b>Product Measure</b>	<b>Product Measure</b>
$\text{Cr} = Cr_1 \times Cr_2 \times \dots \times Cr_n$	$M = M_1 \wedge M_2 \wedge \dots \wedge M_n$

The fundamental difference between a random variable and an uncertain variable lies in the measures they require. Random variable is defined by probability measure while uncertain variable is defined by the uncertainty measure. In the associated triplets, the first two elements are similar in form: the set and the  $\sigma$ -algebra on the set. However, the third component in the triplets: the probability measure  $Pr$  obeys  $\sigma$ -additivity while the uncertain measure  $M$  obeys  $\sigma$ -subadditivity. The specification of the measure has impacts on the behaviour of any measurable function on the uncertainty space.

Uncertain measure is neither a completely additive measure nor a completely non-additive measure; it may be called a “partially additive measure” because of its self-duality. Uncertain measure preserves the law of truth conservation and is consistent with the measure of excluded middle and the law of contradiction.

### 5.3 Axioms and Definitions

Liu (2007) provided the foundation and the working environment for uncertainty measure. The concept of uncertainty theory has become a branch of mathematics with many recent applications. For more related work on uncertainty theory such as uncertain process, uncertain calculus, uncertain programming and its application to system reliability design see on uncertain distribution and for portfolio selection Liu (2008, 2009), Li and Liu (2009).

The core concept in uncertainty theory is the uncertain measure, which is a set function defined on the  $\sigma$ -algebra generated from a non-empty set. Suppose  $\Gamma$  is a nonempty set (space), and  $\mathcal{L}$  is the  $\sigma$ -algebra over  $\Gamma$ . Each element, say  $A \in \Gamma$  and hence  $A \in \mathcal{L}$ , is called an uncertain event. A number denoted as  $M(\cdot)$ :  $0 \leq M(A) \leq 1$  is assigned to event  $A \in \mathcal{L}$  which indicates an uncertain measuring grade with which event  $A \in \mathcal{L}$  occurs. In order to deal with uncertainty in human behaviours, Liu (2007) proposed an uncertain measure  $M(A)$  satisfying the following axioms:

1. Normality Axiom:  $M(\Gamma) = 1$  for the universal set  $\Gamma$ .
2. Self Duality Axiom :  $M(\cdot)$  is self dual, i.e. for any set  $A \in \Gamma$

$$M(A) + M(A^c) = 1 \quad (5.1)$$

3. Monotonicity Axiom:  $M(\cdot)$  is non-decreasing whenever  $A \subset B$ , then

$$M(A) \leq M(B) \quad (5.2)$$

4. Sub-additivity Axiom: for a countable sequence of events  $A_i, i = 1, \dots, \dots, \dots$ , then

$$M\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} M(A_i) \quad (5.3)$$

5. Product Axiom: Let  $(\Gamma_j, L_j, M_j)$  be uncertainty spaces for  $j = 1, \dots, m$ . Then the product uncertain measure  $M$  is an uncertain measure on the product  $\sigma$ -algebra  $L$  for sets  $A = A_1 \times A_2 \times \dots \times A_m$  satisfying

$$M\left(\prod_{j=1}^m \Pi\right) = \min_{1 \leq j \leq m} (M_j(A_j)) \quad (5.4)$$

for each event  $A_j \in L_j$  and  $\mathbf{A} \in \mathbf{L}$ .

**Remark 5.1:** A triplet  $(\Gamma, L, M)$  is called an uncertainty space. Any function that satisfies axioms (1) – (4) is an uncertain measure. In order to obtain an uncertain measure of an event, a product uncertain measure was defined by Liu (2009), leading to a fifth axiom of uncertainty theory.

#### 5.4 Uncertainty Measures

**Definition 5.1:** Uncertain Variable: an imprecise variable  $\xi$  is an uncertain variable and defines a measurable mapping, i.e.  $M: D \subseteq R$ . An observation of an imprecise variable is a real number or more broadly, a symbol, an interval, or a real valued vector or an expression from a sample space. Equivalently, it is an uncertainty distribution  $\Phi_\xi(\cdot)$  under a given scheme comprising a set and  $\sigma$ -algebra. The single value of a variable with imprecision should not be understood as an isolated real number but rather as a representative or realization from the uncertain population.

**Definition 5.2:** (Liu, 2007) An uncertain variable is a measurable function  $\xi$  from an uncertainty space  $(\Gamma, L, M)$  to the set of real numbers  $R$ , that is for any Borel set  $B$  of real numbers, there exists a pre-image  $\Gamma$  such that

$$\xi \subset B = \{ \gamma \in \Gamma \mid \xi(\gamma) \subset B \} \quad (5.5)$$

**Illustration 1:** Consider an uncertain space  $(\Gamma, L, M)$  to be  $\{\gamma_1, \gamma_2, \gamma_3\}$  with respect to set  $L$ . Then the set-valued function

$$\xi(r) = \begin{cases} [1, 3] & \text{if } \gamma = \gamma_1 \\ [2, 4] & \text{if } \gamma = \gamma_2 \\ [3, 5] & \text{if } \gamma = \gamma_3 \end{cases}$$

This is an uncertain set on  $\{\gamma_1, \gamma_2, \gamma_3\}$ .

**Illustration 2:** Consider an uncertain space  $(\Gamma, L, M)$  to be mapped on  $R$  with Borel algebra  $L$ . Then, the set value function is  $\xi(\gamma) \{ \gamma, \gamma + 1 \}$  for all  $\gamma \in \Gamma$  is also an uncertain set on  $(\Gamma, L, M)$ .

**Theorem 5.1:** Liu (2007). Let  $\xi$  be an uncertain variable with continuous uncertainty distribution  $\Phi(x)$ . Then for any real number  $x$  we have

$$M\{\xi \leq x\} = \Phi(x), \quad M\{\xi > x\} = 1 - \Phi(x) \quad (5.6)$$

**Proof:** The equation  $M\{\xi \leq x\} = \Phi(x)$  follows from the definition of uncertainty distribution. By using the duality axioms of uncertain measure and continuity of uncertainty distribution, we obtain the second part,

$$M\{\xi > x\} = 1 - M\{\xi \leq x\} = 1 - \Phi(x) \quad (5.7)$$

#### 5.4.1 Some Uncertainty Distributions

**Definition 5.3:** Let  $\xi$  be an uncertain variable or quantity of impreciseness on an uncertainty measure space  $(\Gamma, L, M)$ . The uncertainty measure  $M : \Phi(x) \rightarrow [0, 1]$  of uncertain variable  $\xi$  is defined by

$$\Phi_{\xi}(x) = M\{\tau \in \Gamma | \xi(\tau) \leq x\} = M\{\xi \leq x\} \quad (5.8)$$

For an uncertain measure, the axiomatic measure development and the set class in the  $\sigma$ -algebra play critical roles in defining the set function. The measure function  $\{M\}$  defines the measurability of an uncertain variable and the critical roles it plays in defining set function. Chung (1974) stated that a random variable on a probability space  $(\Omega, A, P)$  induces a probability measure  $(R, B, \mu)$  by means of the following correspondence for all  $B \in \mathbb{B}$ :  $\mu(B) = \Pr\{X^{-1}(B)\} = \Pr\{X \in B\}$ . Guo, et al. (2010) stated a procedure for determining an uncertainty grading measure  $M\{\cdot\}$  on an uncertain variable.

**Definition 5.4:** Let  $\xi$  be an uncertain variable with essential form, which takes values from ascending ordered set  $\mathbb{D} = \{c_0, c_2, \dots, c_n\}$  with the uncertainty distribution  $\Psi$  satisfying the following necessary and sufficient conditions:

1.  $\Psi(c_0) = M\{\xi \leq c_0\} = 0$
2. For  $i = 1, 2, \dots, n - 1$ 

$$\Psi(c_i) = M\{\xi < c_i\} = \Psi_{i-}$$

$$\Psi(c_{i-}) = M\{\xi = c_i\} = \pi_i$$

$$\Psi(c_{i+}) = M\{\xi > c_i\} = \Psi_{i+}$$
3.  $\Psi(c_{n-}) = M\{\xi < c_n\} = \Psi_{n-}$ 

$$\Psi(c_n) = M\{\xi = c_i\} = \pi_n$$

$$\Psi(c_{n+}) = M\{\xi \leq c_i\} = 1$$

4. The uncertain measure singleton  $\{c_i\}$  can be estimated with following uncertainty density functions  $\pi_i$

$$\pi_{i-} = M\{\xi = c_i - 0\}$$

$$\pi_i = M\{\xi = c_n\}$$

$$\pi_{i+} = M\{\xi = c_n + 0\}$$

**Illustration 3:** Let  $\xi$  be an uncertain variable, which takes values  $X = \{1, 2, 3, 4\}$ , with an uncertainty distribution  $\Phi(x)$  as shown in Table 5.3 below

**Table 5.3:** Example of uncertainty measures

$\xi$	$\Phi(x-) = M\{\xi < x\}$	$\Phi(x) = M\{\xi = x\}$	$\Phi(x) = M\{\xi > x\}$
0	0	0.25	0.75
1	0.45	0.125	0.425
2	0.70	0.07	0.23
3	0.84	0.01	0.15
4	0.95	0.05	0.00

We demonstrate the computations of induced uncertainty measure in uncertainty statistics theory similar to probability space for the data in Table (5.4) below

**Table 5.4:** Computation of Uncertainty Distribution and Density Measures

Uncertainty Distribution $\Phi(x)$	Uncertainty Density ( $\pi$ )
$\Phi(0_+) = M\{\xi \leq 0_+\} = 0.05$	$\pi_{0+} = M\{\xi = 0_+\} = 0.05$
$\Phi(1_-) = M\{\xi \leq 1 - 0\} = 0.45$	$\pi_{1-} = M\{\xi = 1 - 0\} = 0.40$
$\Phi(1) = M\{\xi \leq 1\} = 0.575$	$\pi_1 = M\{\xi = 1\} = 0.125$
$\Phi(1_+) = M\{\xi \leq 1 + 0\} = 0.59$	$\pi_{1+} = M\{\xi = 1 + 0\} = 0.015$
$\Phi(2_-) = M\{\xi \leq 2 - 0\} = 0.70$	$\pi_{2-} = M\{\xi = 2 - 0\} = 0.11$
$\Phi(2) = M\{\xi \leq 2\} = 0.77$	$\pi_2 = M\{\xi = 2\} = 0.07$
$\Phi(2_+) = M\{\xi \leq 2 + 0\} = 0.80$	$\pi_{2+} = M\{\xi = 2 + 0\} = 0.03$
$\Phi(3_-) = M\{\xi \leq 3 - 0\} = 0.84$	$\pi_{3-} = M\{\xi = 3 - 0\} = 0.04$
$\Phi(3) = M\{\xi \leq 3\} = 0.85$	$\pi_3 = M\{\xi = 3\} = 0.01$
$\Phi(3_+) = M\{\xi \leq 3 + 0\} = 0.93$	$\pi_{3+} = M\{\xi = 3 + 0\} = 0.08$
$\Phi(4_-) = M\{\xi \leq 4 - 0\} = 0.95$	$\pi_{4-} = M\{\xi = 4 - 0\} = 0.02$
$\Phi(4) = M\{\xi \leq 4\} = 1.00$	$\pi_4 = M\{\xi = 4\} = 0.05$

**Example 1:** The uncertainty distribution  $\Phi(x)$  of an uncertain variable  $\xi$  is defined  $\Phi(x) = M(\xi \leq x)$  for any real number  $x$ . For instance, the linear uncertain variable  $\xi \sim L(a, b)$  has an uncertainty distribution

$$\Phi(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b \end{cases} \quad (5.9)$$

where  $a$  and  $b$  are real numbers with  $0 < a < b$  and presented in Figure 5.1

**Example 2:** An uncertain variable  $\xi$  is called zigzag if it has a zigzag uncertainty distribution  $\Phi(x)$  defined by

$$\Phi(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{2(b-a)} & \text{if } a \leq x < b \\ \frac{x+c-2b}{2(c-b)} & \text{if } b \leq x < c \\ 1 & \text{if } x \geq c \end{cases} \quad (5.10)$$

Equation (5.10) is denoted by  $\mathcal{Z}(a, b, c)$  where  $a, b, c$  are real numbers with  $(0 < a < b < c)$  as shown in Figure 5.2

**Example 3:** An uncertain variable  $\xi$  is called normal if it has a normal uncertainty distribution  $\Phi(x)$  defined by

$$\Phi(x) = \left[ 1 + \exp\left(\frac{\pi(\mu - x)}{\sigma\sqrt{3}}\right) \right]^{-1}, \quad x \in R \quad (5.11)$$

Equation (5.11) is denoted by  $\xi \sim N(\mu, \sigma)$  where  $\mu$  and  $\sigma$  are real numbers with  $\sigma > 0$  which is presented in Figure (5.3). It should be noted that the uncertainty normal distribution is characterized by its mean and its variance which are respectively given by  $E[\xi] = \mu$  and  $Var[\xi] = \sigma^2$ .

**Example 4:** An uncertain variable is called  $Ln(\xi)$  if  $\xi$  is a normal uncertain variable  $Ln(\xi) \sim N(\mu, \sigma)$ . In other words, a lognormal uncertain variable has an uncertainty distribution

$$\Phi_{ln}(x) = \left[ 1 + \exp\left(\frac{\pi(\mu - \ln(x))}{\sigma\sqrt{3}}\right) \right]^{-1}, \quad x \geq 0 \quad (5.12)$$

Equation (5.12) is denoted by  $\xi = \text{In}(x) \sim \text{LOG}(\mu, \sigma)$  where  $\mu$  and  $\sigma$  are real numbers with scale parameter  $\sigma > 0$  and presented in Figure (5.4). The mean and variance of a lognormal uncertainty distribution are given as  $\mu$  and  $\sigma$  respectively.

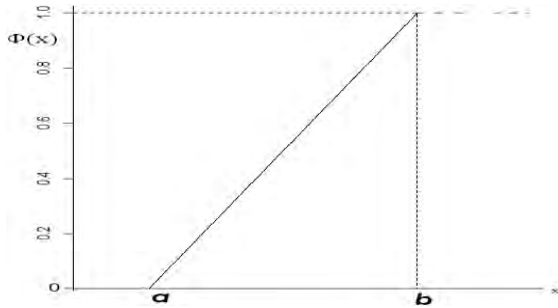


Figure 5.1: Uncertainty Linear Distribution

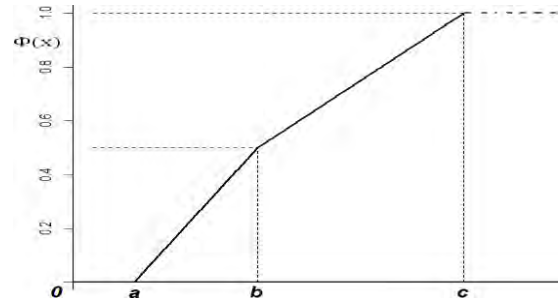


Figure 5.3: Uncertainty Zigzag Distribution

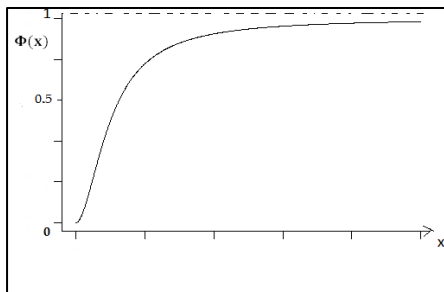


Figure 5.2: Uncertainty Normal Distribution

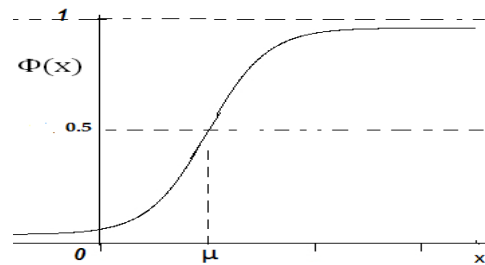


Figure 5.4: Uncertainty Lognormal Distribution

#### 5.4.2 Inverses of Uncertain Distribution

**Definition 5.5:** An uncertainty distribution  $\Phi(x)$  is said to be regular if its inverse function  $\Phi(x)^{-1}(\alpha)$  exists and is unique for each  $\alpha \in (0, 1)$ .

Suppose a regular uncertainty distribution is a continuous function and is strictly increasing at point  $x$  with  $0 < \Phi(x) < 1$ , then we have

$$\lim_{x \rightarrow -\infty} \Phi(x) = 0 \quad \lim_{x \rightarrow +\infty} \Phi(x) = 1 \quad (5.13)$$

**Definition 5.6:** Let  $\xi$  be an uncertain variable with regular uncertainty distribution  $\Phi(x)$ , the inverse function  $\Phi^{-1}(\alpha)$  is called the inverse uncertainty distribution of  $\xi$ .

It should be stated that the inverse uncertainty distribution  $\Phi^{-1}(\alpha)$  is defined on an open interval  $(0,1)$  while for a closed interval  $[0,1]$  the inverse is defined to assume the limiting value given as

$$\Phi^{-1}(0) = \lim_{\alpha \downarrow 0} \Phi^{-1}(\alpha) \quad , \quad \Phi^{-1}(1) = \lim_{\alpha \uparrow 1} \Phi^{-1}(\alpha) \quad , \quad (5.14)$$

**Example 5:** The inverse uncertainty distribution function  $\Phi(1)^{-1}$  of the linear uncertain variable  $\xi \sim L(a, b)$  corresponding to uncertainty linear distribution equation (5.9) is

$$\Phi^{-1}(\alpha) = (1 - \alpha)a + \alpha b \quad (5.15)$$

**Example 6:** The inverse uncertainty distribution of zigzag uncertain variable  $Z(a, b, c)$  corresponds to equation (5.10) is

$$\Phi(\alpha)^{-1} = \begin{cases} (1 - 2\alpha)a + 2\alpha b & \text{if } \alpha < 0.5 \\ (2 - 2\alpha)b + (2\alpha - 1)c & \text{if } \alpha \geq 0.5 \end{cases} \quad (5.16)$$

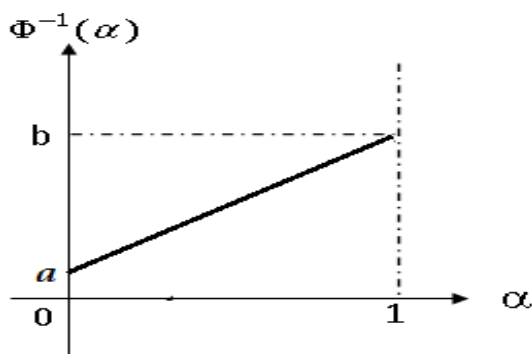
**Example 7:** The Inverse Uncertainty distribution of normal uncertain variable  $N(\mu, \sigma)$  corresponds to equation (5.11)

$$\Phi^{-1}(\alpha) = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \quad \alpha \in (0,1) \quad (5.17)$$

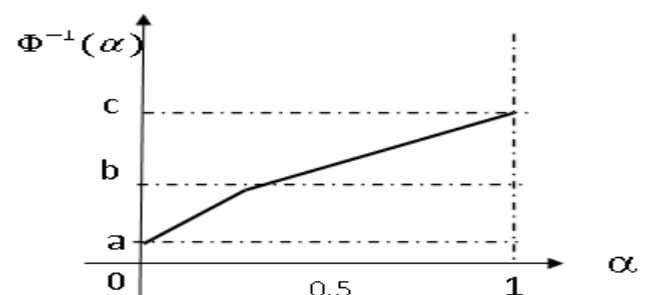
**Example 8:** The Inverse Uncertainty distribution of *lognormal* uncertain variable  $N(\mu, \sigma)$  corresponds to equation (5.12)

$$\Phi^{-1}(\alpha) = \exp(\mu) \left(\frac{\alpha}{1-\alpha}\right)^{\frac{\sigma\sqrt{3}}{\pi}} \quad (5.18)$$

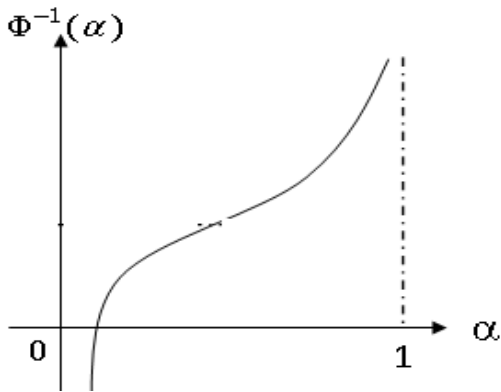
The inverses of uncertainty distributions are used to determine the values of their parameters. Where the inverse exists and it is easier to compute, it is preferable to use the inverse for estimation instead of the distributions. The corresponding graphs of the inverses are presented in Figures (5.5) - (5.8).



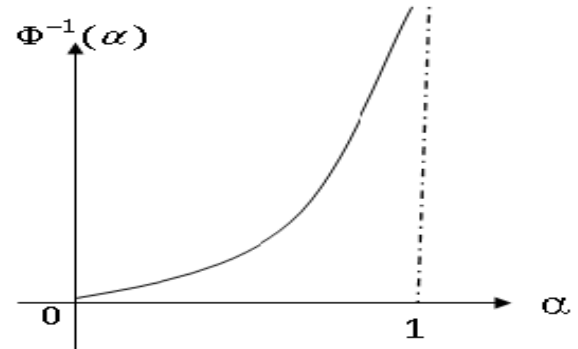
**Figure 5.5:** Inverse Uncertainty Linear Distribution



**Figure 5.6:** Inverse Uncertainty Zigzag Distribution



**Figure 5.7:** Inverse Uncertainty Normal Distribution



**Figure 5.8:** Inverse Uncertainty Lognormal Distribution

### 5.4.3 Uncertain Normal Distribution

**Definition 5.7:** Liu (2007). An uncertain variable  $\xi$  on uncertain space  $(\Gamma, L, M)$  is called normal if the uncertain distribution takes the form

$$\Phi_{\xi}(x) = \frac{1}{1 + e^{\left(\frac{-\pi}{\sigma\sqrt{3}}\right)(x-\mu)}}, \quad x \in R \quad (5.19)$$

**Definition 5.8:** Guo (2012). An uncertain normal distribution becomes a standard uncertain normal distribution if equation (5.19) takes the form

$$\Phi_{\xi}(x) = \frac{1}{1 + e^{\frac{-\pi}{\sigma\sqrt{3}}x}}, \quad x \in R \quad (5.20)$$

The standard uncertain normal distribution has parameters  $(\mu = 0, \sigma = 1)$ .

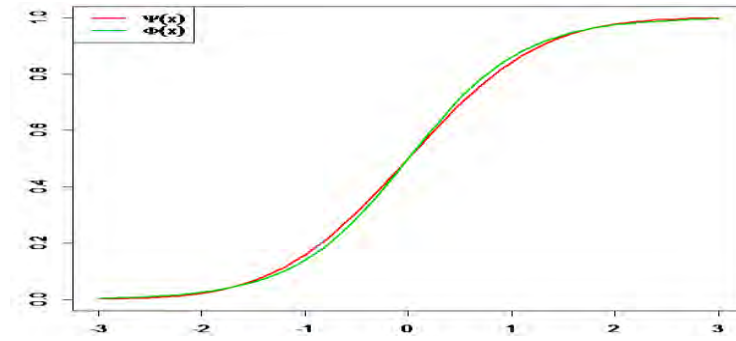
The emphasis on the functional form of an uncertain normal distribution arises seeing that it can be used to play the role of a tractable cumulative distribution, while its Gaussian distribution counterpart cannot be used. The basic reason is that Gaussian distribution is usually expressed by an integration, which in nature is  $\sigma$ -additive. The Gaussian distribution approximated this way is called logistic distribution in probability theory. The shape of the logistic is very close to the Gaussian cumulative function but with greater spread. The Kolmogorov- Smirnov distance between the logistic and the Gaussian  $(0, 1)$  distribution has the smallest value when the logistic scale parameter is approximated by  $1.7\sigma$  i.e. when  $\sigma = \frac{(1.7)^2\pi^2}{3}$  as shown in Figure 5.9. The green graph is the logistic distribution in probability theory and the red graph is the standard uncertainty normal distribution.

Shannon, et al., (2009) recently reviewed several studies of some families of systems of distribution, where the distributional parameters were replaced by a reasonable value to approximate the cumulative normal distribution. For example, lognormal distribution was replaced with small absolute less than 0.25 while the Weibull shape parameter chosen close to 3.25 would closely approximate the standard normal distribution. Others have concentrated on building complex distributions and their successful application to problems in areas such as engineering, lifetime, economics and biomedical sciences, among others. For more information see (Eugene, et al., (2002); Johnson, et al., (1995); Jones, M.C. (2004)).

The fundamental comparisons between uncertainty normal variable and Gaussian normal random variable is presented in Table 5.4. Other useful uncertainty distributions in such variables as uncertain Chi- statistic, uncertain F-statistic, and uncertain t-statistic and so on, have been constructed from the standard uncertain normal distribution. More information on derivations and moments from an uncertainty normal distribution can be obtained from Guo, et al., (2010).

Table 5.4: Comparison between Uncertainty normal variable and Gaussian random variable

	Uncertain normal variable ( $\xi$ )	Gaussian random variable ( $Z$ )
Standard	$\Psi_{\xi_0}(x) = \left(1 + e^{\frac{-\pi}{\sqrt{3}}x}\right)^{-1}$	$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$
General	$\Psi_{\xi}(x) = \left[1 + \exp\left(\frac{-\pi(x - \theta)}{\sigma\sqrt{3}}\right)\right]^{-1}$	$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$
Link	$\xi = \theta + \sigma\xi_0$	$X = \mu + \sigma Z$



**Figure 5.9:** Standard Uncertainty Normal Distribution ( $\Psi$ ) and Gaussian Distribution ( $\Phi$ )

#### 5.4.4 Uncertain Expected Value

Expected value is the average value of uncertain variable in the sense of uncertain measure and represents the size or location of an uncertain variable.

**Definition 5.9:** Liu (2007). Let  $\xi$  be an uncertain variable. Then the uncertain expected value of  $\xi$  is defined by

$$E[\xi] = \int_0^{+\infty} M\{\xi \geq r\} dr - \int_{-\infty}^0 M\{\xi \leq r\} dr \quad (5.21)$$

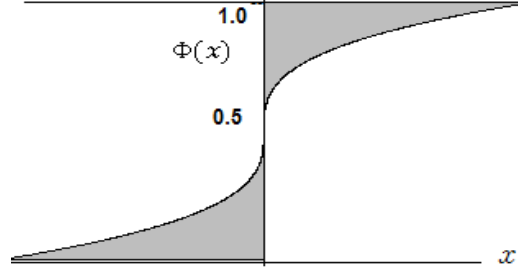
provided, at least, one of the integrals is finite.

**Theorem 5.2:** Liu (2007). Let  $\xi$  be an uncertain variable with uncertainty distribution  $\Phi$ . If the expected value exists, then

$$E[\xi] = \int_0^{+\infty} (1 - \Phi(x)) dx - \int_{-\infty}^0 \Phi(x) dx \quad (5.22)$$

**Proof:** Given an uncertainty distribution  $\Phi(\cdot)$  it follows from the definition of expected value operator and the Fubini theorem that

$$\begin{aligned} E[\xi] &= \int_0^{+\infty} M\{\xi \geq r\} dr - \int_{-\infty}^0 M\{\xi \leq r\} dr \\ &= \int_0^{+\infty} (1 - \Phi(x)) dx - \int_{-\infty}^0 \Phi(x) dx \end{aligned}$$



**Figure 5.10:** Lebesgue finite integral

**Theorem 5.3:** Let  $\xi$  be an uncertain variable with uncertainty distribution  $\Phi$ . If the expected value  $E[\xi]$  exists, then

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha \quad (5.23)$$

**Proof:** Suppose the inverse of uncertainty distribution can be easily obtained, it follows from the definition of expected value operator and uncertainty distribution that

$$\begin{aligned} E[\xi] &= \int_0^{+\infty} (1 - \Phi(x)) dx - \int_{-\infty}^0 \Phi(x) dx \\ &= \int_{\Phi(0)}^1 \Phi^{-1}(\alpha) d\alpha + \int_0^{\Phi(0)} \Phi^{-1}(\alpha) d\alpha \\ &= \int_0^1 \Phi^{-1}(\alpha) d\alpha \\ E[\xi] &= \int_0^1 \Phi^{-1}(\alpha) d\alpha \end{aligned}$$

**Example 5.9:** Let  $\xi \sim L(a, b)$  be an uncertain linear variable then its inverse uncertainty distribution  $\Phi^{-1}(\alpha) = (1 - \alpha)a + \alpha b$  and its expected value is given by

$$E[\xi] = \int_0^1 ((1 - \alpha)a + \alpha b) d\alpha = \frac{a + b}{2} \quad (5.24)$$

and its second moment is given by

$$E[\xi^2] = \int_0^1 ((1 - \alpha)a + \alpha b)^2 d\alpha = \frac{a^2 + ab + b^2}{3} \quad (5.25)$$

For an uncertain zigzag variable  $\xi \sim Z(a, b, c)$ , then the expected value is given by

$$E[\xi] = \int_0^1 ((1 - \alpha)a + \alpha b) d\alpha = \frac{a + 2b + c}{4} \quad (5.26)$$

The normal uncertain variable  $\xi \sim N(\mu, \sigma)$ , has an expected value  $\mu$ , i.e.,  $E[\xi] = \mu$

For uncertain lognormal variable  $\xi \sim LOGN(\mu, \sigma)$  with  $\sigma < \frac{\pi}{\sqrt{3}}$ , Liu (2012) proposed that the expected value is given by

$$E[\xi] = \sigma \sqrt{3} \exp(\mu) \operatorname{Cosec}(\sigma \sqrt{3}) \quad (5.27)$$

Otherwise,  $E[\xi] = +\infty$ .

For detailed information about the expectation of uncertainty functions, see Liu (2012, pp51-52).

#### 5.4.5 Expectation of a function of an uncertain variable

In probability theory, the expected value of a function of a random variable is the Lebesgue-Stieltjes integral of the function with respect to its probability distribution. In the same manner, Xue et al. (2008) proposed an expected value of a function of a fuzzy variable, when the function is monotone, analogous to the probability concept. In uncertain statistics, the following theorem will facilitate the computation of expected value of a function of an uncertain variable with respect to its uncertainty distribution. A function of an uncertain variable can be said to be an uncertain variable itself and share any property of an uncertain variable.

**Theorem 5.4:** Liu (2011). Let  $\xi_1, \dots, \xi_n$  be independent uncertain variables with regular uncertainty distributions  $\Phi_1, \dots, \Phi_n$  respectively. If the function  $f(x_1, \dots, x_n)$  is strictly increasing with respect to  $x_1, \dots, x_m$  and strictly decreasing with respect to  $x_{m+1}, \dots, x_n$ , then  $\xi = f(\xi_1, \dots, \xi_n)$  is an uncertain variable with inverse uncertainty distribution given by

$$\Psi^{-1}(\alpha) = f\{\Phi_1^{-1}(\alpha), \dots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)\} \quad (5.28)$$

**Proof:** We prove by induction. Let  $n = 2$  and  $m = 1$ . We have

$$M\{\Psi^{-1}(\alpha)\} = M\{f(\xi_1, \xi_2) \leq f\{\Phi_1^{-1}(\alpha), \Phi_2^{-1}(1-\alpha)\}\}$$

Since the function  $f(x_1, x_2)$  is strictly increasing with respect to  $x_1$  and strictly decreasing with  $x_2$ , we obtain

$$\{\xi \leq \Psi^{-1}(\alpha)\} \supset \{\xi_1 \leq \Phi_1^{-1}(\alpha)\} \cap \{\xi \leq \Phi_2^{-1}(1-\alpha)\}$$

By using the independence of  $\xi_1$  and  $\xi_2$ , we get

$$M\{\xi \leq \Psi^{-1}(\alpha)\} \geq M\{\xi_1 \leq \Phi_1^{-1}(\alpha) \wedge \xi_2 \leq \Phi_2^{-1}(1 - \alpha)\} = \alpha \wedge \alpha$$

On the other hand, since the function  $f(x_1, x_2)$  is strictly increasing with respect to  $x_1$  and strictly decreasing with  $x_2$  we obtain

$$M\{\xi \leq \Psi^{-1}(\alpha)\} \supset M\{\xi_1 \geq \Phi_1^{-1}(\alpha)\} \cap M\{\xi_2 \leq \Phi_2^{-1}(1 - \alpha)\}.$$

By using the independence of  $\xi_1$  and  $\xi_2$ , we get

$$M\{\xi \leq \Psi^{-1}(\alpha)\} \supset M\{\xi_1 \geq \Phi_1^{-1}(\alpha)\} \cap M\{\xi_2 \geq \Phi_2^{-1}(1 - \alpha)\} = \alpha \wedge \alpha = \alpha$$

In other words,  $M\{\xi \leq \Psi^{-1}(\alpha)\} = \alpha$  and  $\Psi$  is just the uncertainty distribution of  $\xi$

The theorem is proved.

**Definition 5.10:** Liu and Ha (2010). Assume  $\xi_1, \dots, \xi_n$  are independent uncertain variables with regular uncertainty distributions  $\Phi_1, \dots, \Phi_n$  respectively. If  $(x_1, \dots, x_n)$  is strictly increasing with respect to  $x_1, \dots, x_m$  and strictly decreasing with respect to  $x_{m+1}, \dots, x_n$ , then the uncertain variable  $\xi = (\xi_1, \dots, \xi_n)$  has an expected value

$$E[\xi] = \int_0^1 f\{\Phi_1^{-1}(\alpha), \dots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1 - \alpha), \dots, \Phi_n^{-1}(1 - \alpha)\} d\alpha \quad (5.28)$$

provided that  $E[\xi]$  exists.

**Theorem 5.5:** Let  $\xi$  be an uncertain variable with uncertainty distribution  $\Phi(x)$ . If  $f(x)$  is a monotone function such that the the expected value  $f(\xi)$  exists, then

$$E[f(\xi)] = \int_{-\infty}^{+\infty} f(x) d\Phi(x) \quad (5.29)$$

**Proof:** Suppose  $f(x)$  is a monotone increasing function and the expected value  $E(f(\xi))$  is finite, we immediately have

$$\lim_{y \rightarrow +\infty} M\{\xi \geq y\} f(y) = \lim_{y \rightarrow +\infty} (1 - \Phi(y)) f(y) = 0 \quad (5.30)$$

and

$$\lim_{y \rightarrow -\infty} M\{\xi \leq y\} f(y) = \lim_{y \rightarrow -\infty} (\Phi(y)) f(y) = 0 \quad (5.31)$$

Let  $a$  and  $b$  be two real numbers such that  $a < 0 < b$ , suppose a function  $f(\xi)$  is integrable on the interval  $[a, b]$  and from definition of expected value in equation (5.21)

$$E[f(\xi)] = \int_0^b M\{\xi \geq r\} dr - \int_a^0 M\{\xi \leq r\} dr$$

$$\begin{aligned}
&= \int_0^b M\{\xi \geq f^{-1}(r)\}dr - \int_a^0 M\{\xi \leq f^{-1}(r)\}dr \\
&= \int_{f^{-1}(0)}^{f^{-1}(b)} M\{\xi \geq y\}df(y) - \int_{f^{-1}(a)}^{f^{-1}(0)} M\{\xi \leq y\}df(y)
\end{aligned}$$

As  $a \rightarrow -\infty$  and  $b \rightarrow +\infty$  and applying equations (5.30) and (5.31) above, we obtain

$$\begin{aligned}
E[f(\xi)] &= \int_{f^{-1}(0)}^{f^{-1}(+\infty)} M\{\xi \geq y\}df(y) - \int_{f^{-1}(-\infty)}^{f^{-1}(0)} M\{\xi \leq y\}df(y) \\
E[f(\xi)] &= \int_{f^{-1}(0)}^{+\infty} M\{\xi \geq r\}dr - \int_{-\infty}^{f^{-1}(0)} M\{\xi \leq r\}dr \\
E[f(\xi)] &= \int_{f^{-1}(0)}^{+\infty} f(y)d\Phi(y) - \int_{-\infty}^{f^{-1}(0)} f(y)d\Phi(y) \\
E[f(\xi)] &= \int_{-\infty}^{+\infty} f(y)d\Phi(y)
\end{aligned}$$

**Remark 5.2:** Let  $\xi$  is an uncertain variable with regular uncertainty distribution  $\Phi(x)$ , and  $f(x)$  is a strictly monotone (increasing or decreasing) function such that the expected value  $E[f(\xi)]$  exists, then by the above theorem

$$E[f(\xi)] = \int_0^1 \Phi^{-1}(\alpha) d\alpha \quad (5.32)$$

Suppose  $f(\xi) = \Phi^{-1}(\alpha)$ , then  $\alpha = \Phi(x)$ . It follows from the change of variables of integration that

$$E[f(\xi)] = \int_{-\infty}^{+\infty} f(x) d\Phi(x) \quad (5.33)$$

Generally, under any arithmetic operations (+, -, /, x), the theorem holds for a function of an uncertain variable with respect to the uncertainty distribution.

**Remark 5.3:** Assuming that  $f$  and  $g$  are two co-monotonic functions of an uncertain variable  $\xi$ , then we write

$$\begin{aligned}
E[f(\xi) + g(\xi)] &= \int_0^1 \{f(\Phi^{-1}(\alpha)) + g(\Psi^{-1}(\alpha))\} d\alpha \\
&= \int_0^1 f(\Phi^{-1}(\alpha)) d\alpha + \int_0^1 g(\Psi^{-1}(\alpha)) d\alpha \\
&= E[f(\xi) + g(\xi)] \tag{5.34}
\end{aligned}$$

The preceding section plays a vital role in estimation of parameters of a function of uncertain variable with respect to uncertainty distribution. We shall explore the results from the remarks of equations (5.32) and (5.34).

#### 5.4.6 Uncertain Variance

The variance of an uncertain variable is an important criterion for ranking uncertain variables. It suggests a degree for the spread of the distribution around its expected value.

**Definition 5.11:** Let  $\xi$  be an uncertain variable with finite expected value  $\mu$ , then the variance of  $\xi$  is defined by  $V(\xi) = E[(\xi - \mu)^2]$ .

Thus, the variance is just the expected value of  $(\xi - \mu)^2$ . Since  $(\xi - \mu)^2$  is a nonnegative uncertain variable, we also have

$$V[\xi] = \int_0^{+\infty} M\{(\xi - \mu)^2 \geq x\} dx \tag{5.35}$$

**Theorem 5.6:** Liu (2007). Let  $\xi$  be an uncertain variable on an uncertainty measure space  $(\Gamma, L, M)$  with uncertainty distribution  $\Phi$  and a known finite expectation  $\mu$ . Then the variance is constrained by

$$V[\xi] = E[(\xi - \mu)^2] \leq 2 \int_0^{+\infty} (r(1 - \Phi(\mu + r) + \Phi(\mu - r))) dr \tag{5.36}$$

**Proof (Variance):** Let  $\xi$  be an uncertain variable with expected value  $\mu$ , suppose we know its uncertainty distribution  $\Phi$  then the variance

$$\begin{aligned}
V[\xi] &= E[(\xi - \mu)^2] = \int_0^{+\infty} M\{(\xi - \mu)^2 \geq x\} dx \\
&= \int_0^{+\infty} M\{(\xi \geq \mu + \sqrt{x}) \cup (\xi \leq \mu - \sqrt{x})\} dx
\end{aligned}$$

$$\begin{aligned} &\leq \int_0^{+\infty} M\{(\xi \geq \mu + \sqrt{x}) + (\xi \geq \mu - \sqrt{x})\}dx \\ &= \int_0^{+\infty} 1 - \Phi(\mu + \sqrt{x}) + \Phi(\mu - \sqrt{x})\}dx \end{aligned}$$

By change of variable  $r = \sqrt{x}$ , we obtain the expression for the variance as

$$V[\xi] = 2 \int_0^{+\infty} (r(1 - \Phi(\mu + r) + \Phi(\mu - r))) dr \quad (5.37)$$

The theorem is proved.

### 5.5 Uncertain Distance

Distance is used to measure the degree of the separation between any two variables in a measure space. For example, the distance in a  $n$ -dimensional Euclidean space is  $d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ .

In an uncertainty space  $(\Gamma, L, M)$ , the distance between two uncertain variables is defined by uncertain distance. An uncertain variable describes an imprecise quantity; therefore, the distance between two uncertain variables is to be specified in terms of expectations of the norms of the difference.

**Definition 5.21:** The distance between two uncertain variables  $\xi$  and  $\eta$  is defined as

$$d(\xi, \eta) = E|\xi - \eta| \quad (5.38)$$

Thus, the distance between  $\xi$  and  $\eta$  is just the expected value of  $|\xi - \eta|$ . It is obvious the  $|\xi - \eta|$  is a nonnegative uncertain variable, then distance is defined by

$$E\{|\xi - \eta|\} = \int_0^{+\infty} M\{|\xi - \eta| \geq r\}dr \quad (5.39)$$

The basic properties of uncertain distance are given by the theorem that follows.

**Theorem 5.11:** Liu (2010). Let  $\xi, \eta$  and  $\tau$  be uncertain variables, and let  $d(\xi, \eta)$  be the distance function. Then we have

- (a) (Non-negativity):  $d(\xi, \eta) \geq 0$ .
- (b) (Identification):  $d(\xi, \eta) = 0$  if and only if  $\xi = \eta$
- (c) (Symmetry):  $d(\xi, \eta) = d(\eta, \xi)$

(d) (Triangle Inequality):  $\xi d(\xi, \eta) \leq d(\xi, \tau) + d(\tau, \eta)$ .

**Proof:** The parts (a), (b) and (c) follow immediately from definition (5.21). It follows from the sub-additivity axioms in measure theory that

$$\begin{aligned}
 d(\xi, \eta) &= E|\xi - \eta| = \int_0^{+\infty} M\{|\xi - \eta| \geq r\} dr \\
 &\leq \int_0^{+\infty} M\{|\xi - \tau| + |\tau - \eta| \geq r\} dr \\
 &= \int_0^{+\infty} M\{|\xi - \tau| \geq r - a\} \cup \{|\tau - \eta| \geq a\} dr \quad \text{for all } 0 \leq a \leq r \\
 &= \int_0^{+\infty} M\{|\xi - \tau| \geq r_1\} dr_1 + \int_0^{+\infty} M\{|\tau - \eta| \geq r_2\} dr_2 \quad \text{for } r_1 + r_2 = r \\
 &= E(|\xi - \tau|) + E|\tau - \eta| = d(\xi, \tau) + d(\tau, \eta)
 \end{aligned}$$

The theorem is verified.

Using equation (5.38), we obtain the distance as

$$\begin{aligned}
 d(\xi, \eta) &= E|\xi - \eta| = \int_0^{+\infty} M\{|\xi - \eta| \geq x\} dx \\
 &= \int_0^{+\infty} M\{|\xi - \tau| \geq x\} \cup \{|\tau - \eta| \leq -x\} dx \\
 &= \int_0^{+\infty} M\{|\xi - \tau| \geq x\} dx + \int_0^{+\infty} M\{|\tau - \eta| \leq -x\} dx
 \end{aligned}$$

(Note that we apply symmetry property of distance here, or self-duality of axioms definition)

$$= \int_0^{+\infty} (1 - Y(x) + Y(-x)) dx$$

Hence, the distance between  $\xi$  and  $\eta$  is given by

$$d(\xi, \eta) = \int_0^{+\infty} (1 - Y(x) + Y(-x)) dx \quad (5.40)$$

We therefore conclude that equation (5.40) is a stipulation expression rather than precise using the result of equation(5.38). However, if it is easier to express equation (5.40) as its inverse of their distributions, then distance between two uncertain variables becomes

$$d(\xi, \eta) = \int_{Y(0)}^1 (Y^{-1}(\alpha) d\alpha - \int_0^{Y(0)} Y^{-1}(-\alpha) d\alpha) \quad (5.41)$$

where the inverse uncertainty distribution of the difference between two uncertain variables  $\xi$  and  $\eta$  having inverses of independent distributions  $\Phi$  and  $\Psi$  respectively given as

$$Y^{-1}(\alpha) = \Phi^{-1}(\alpha) - \Psi^{-1}(\alpha) \quad (5.42)$$

### 5.5.1 Distance for Uncertainty Normal Distributions

**Theorem 3.12:** Guo (2012) stated that if two uncertain variables  $\xi_1$  and  $\xi_2$  have two independent uncertainty normal distributions  $N(\mu_1, \sigma_1)$  and  $N(\mu_2, \sigma_2)$  respectively then the stipulated distance  $d(\xi_1, \xi_2)$  is also an uncertain normal variable and its distribution is given as

$$d(\xi_1, \xi_2) = (\mu_1 - \mu_2)(1 - 2Y(0)) + \frac{(\sigma_1 + \sigma_2)2\sqrt{3}}{\pi}(2Y(0) - 1 - 3Y(0)\ln 2Y(0))$$

**Proof:** From the definition of uncertain distance

$$d(\xi_1, \xi_2) = E(|\xi_1 - \xi_2|) = \int_0^{+\infty} M\{|\xi_1 - \xi_2| \geq r\} dr$$

The key step is to find the uncertainty distribution of  $\xi_1 - \xi_2$  denoted by  $Y$  then

$$Y^{-1}(\alpha) = \Phi^{-1}(\alpha) - \Psi^{-1}(\alpha)$$

$$\Psi_1^{-1}(\alpha) = \mu_1 + \frac{\sigma_1\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \text{ for a strictly increasing } \xi_1$$

$$\Psi_2^{-1}(1-\alpha) = \mu_2 - \frac{\sigma_2\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \text{ for a strictly decreasing } \xi_2$$

Using inverse of the difference, we have

$$Y^{-1}(\alpha) = \Psi_{\xi_1}^{-1}(\alpha) - \Psi_{\xi_2}^{-1}(1-\alpha)$$

$$d(\xi_1, \xi_2) = (\mu_1 - \mu_2)(1 - 2Y(0)) + \frac{(\sigma_1 + \sigma_2)\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right)$$

Thus, the uncertainty distribution of  $\xi_1 - \xi_2$  is  $N(\xi_1 - \xi_2, \sigma_1 + \sigma_2)$ . Again, we determine

$$d(\xi_1, \xi_2) = Y(0) = \left(1 + \exp\left(\frac{\pi(\mu_1 - \mu_2)}{(\sigma_1 + \sigma_2)\sqrt{3}}\right)\right)^{-1}$$

We note that

$$\int_{Y(0)}^1 \ln\left(\frac{\alpha}{1-\alpha}\right) d\alpha = \int_{Y(0)}^1 \ln\alpha - \ln(1-\alpha) d\alpha$$

$$= 2 \int_{Y(0)}^1 \ln\alpha d\alpha = 2(\alpha \ln\alpha - \int_{Y(0)}^1 \alpha d \ln\alpha) = 2(\alpha \ln\alpha - \alpha)$$

First integrand

$$\int_{Y(0)}^1 \ln\left(\frac{\alpha}{1-\alpha}\right) d\alpha = [2(\alpha \ln \alpha - \alpha)]_0^{Y(0)} = 2(Y(0) - 1 - Y(0) \ln Y(0)) \quad (5.43)$$

Second integrand

$$- \int_0^{Y(0)} \ln\left(\frac{\alpha}{1-\alpha}\right) d\alpha = [2(\alpha \ln \alpha - \alpha)]_0^{Y(0)} = 2(Y(0) - 2Y(0) \ln Y(0)) \quad (5.44)$$

Now, uncertain distance

$$\begin{aligned} d(\xi_1, \xi_2) &= \int_0^{+\infty} M\{|\xi_1 - \xi_2| \geq r\} dr \\ &\leq \int_0^{+\infty} (1 - Y(x) + Y(-x)) dx \\ &= \int_{Y(0)}^1 (\mu_1 - \mu_2) + \frac{(\sigma_1 + \sigma_2)\sqrt{3}}{\pi} d\alpha - \int_0^{Y(0)} (\mu_1 - \mu_2) + \frac{(\sigma_1 + \sigma_2)\sqrt{3}}{\pi} d\alpha \\ &= (\mu_1 - \mu_2)(1 - 2Y(0)) + \frac{(\sigma_1 + \sigma_2)2\sqrt{3}}{\pi} \left[ \int_{Y(0)}^1 \ln\left(\frac{\alpha}{1-\alpha}\right) d\alpha - \int_0^{Y(0)} \ln\left(\frac{\alpha}{1-\alpha}\right) d\alpha \right] \end{aligned}$$

Substituting and collecting the terms in equations (5.43) and (5.44), the distance between two uncertainty normal distributions is obtained as

$$d(\xi_1, \xi_2) = (\mu_1 - \mu_2)(1 - 2Y(0)) + \frac{(\sigma_1 + \sigma_2)2\sqrt{3}}{\pi} (2Y(0) - 1 - 3Y(0) \ln 2Y(0)) \quad (5.45)$$

The formula is proved.

### 5.5.2 Distance between Two Uncertainty Linear Distributions

In this section, we adopt the approach of Guo (2012) above to construct the difference between two uncertainty linear distributions.

Suppose  $\xi$  and  $\eta$  are two independent uncertain variables with uncertainty linear distributions  $\xi \sim L(a_1, b_1)$  and  $\eta \sim L(a_2, b_2)$  respectively. Then distance  $d(\xi, \eta)$  is also an uncertain variable and has an uncertainty distribution is denoted by  $Y$ .

To determine the distance,  $d(\xi, \eta)$  between two uncertain variables having independent distributions, we assume that the difference is regular, so that its inverse distribution exists and is strictly increasing in  $\xi$  and strictly decreasing in  $\eta$ .

**Proof:** Using the corresponding inverses of their uncertainty distributions

$$\Phi^{-1}(\alpha) = (1 - \alpha)a_1 + \alpha b_1 \text{ strictly increasing in } \xi$$

$\Psi^{-1}(1 - \alpha) = \alpha a_2 + (1 - \alpha)b_2$  strictly decreasing in  $\eta$

The inverse distribution for the difference is obtained as

$$\begin{aligned} Y^{-1}(\alpha) &= \Phi^{-1}(\alpha) - \Psi^{-1}(1 - \alpha) \\ &= [(1 - \alpha)a_1 + \alpha b_1] - [\alpha a_2 + (1 - \alpha)b_2] \\ Y^{-1}(\alpha) &= (1 - \alpha)(a_1 - b_2) + \alpha(b_1 - a_2) \end{aligned}$$

The uncertainty distribution of  $|\xi - \eta| \sim L(a_1 - a_2, b_1 - b_2)$  is determined by

$$\begin{aligned} d(\xi, \eta) &= \int_{Y(0)}^1 (1 - \alpha)(a_1 - b_2) + \alpha(b_1 - a_2) d\alpha \\ &\quad - \int_0^{Y(0)} (1 - \alpha)(a_1 - b_2) + \alpha(b_1 - a_2) d\alpha \end{aligned}$$

For the first part of the integrand

$$\begin{aligned} &\int_{Y(0)}^1 (1 - \alpha)(a_1 - b_2) + \alpha(b_1 - a_2) d\alpha \\ &- \frac{1}{2} [(a_1 - b_2)\{2 - Y^2(0)\} + (b_1 - a_2) Y^2(0) - 2(a_1 - b_2)] \\ &= -\frac{1}{2} [(a_1 - b_2)\{-Y^2(0)\} + (b_1 - a_2) Y^2(0)] \\ &= -\frac{1}{2} (Y^2(0) - 1)[(a_1 + b_1) - (a_2 + b_2)] \end{aligned}$$

Similarly for the second part of the integrand

$$\begin{aligned} &-\int_0^{Y(0)} (1 - \alpha)(a_1 - b_2) + \alpha(b_1 - a_2) d\alpha \\ &= -\frac{1}{2} [(a_1 - b_2)\{2 - Y^2(0)\} + (b_1 - a_2) Y^2(0) - 2(a_1 - b_2)] \\ &= -\frac{1}{2} [(a_1 - b_2)\{-Y^2(0)\} + (b_1 - a_2) Y^2(0)] \\ &= -\frac{1}{2} Y^2(0)\{(b_2 - a_1) + (b_1 - a_2)\} \\ &= -\frac{1}{2} Y^2(0)\{(b_1 - a_1) + (b_2 - a_2)\} \end{aligned}$$

Hence, the stipulated distance between two uncertain variables having independent linear uncertainty distributions is obtained as

$$d(\xi, \eta) = \frac{1}{2}(\Upsilon^2(0) - 1)\{(a_1 + b_1) - (a_2 + b_2)\} + \frac{1}{2}\Upsilon^2(0)\{(b_1 - a_1) + (b_2 - a_2)\}.$$

Simplifying and collecting the like terms, we obtain

$$d(\xi, \eta) = \frac{1}{2}\{(b_1 - a_1) - ((b_2 - a_2))\} \quad (5.46)$$

### 5.5.3 Distance for Uncertainty Linear and Uncertainty Normal Distributions

Suppose  $\xi$  and  $\eta$  are two independent uncertain variables with uncertainty linear distributions  $\xi \sim L(a, b)$  and  $\eta \sim N(\mu, \sigma)$  respectively. Then difference  $|\xi - \eta|$  is also an uncertain variable and has an uncertainty distribution given by  $\Upsilon$ .

To determine the stipulated distance  $d(\xi, \eta)$  between two uncertain variables having independent uncertainty distributions, it is assumed that the difference is regular, strictly increasing in  $\xi$  and strictly decreasing in  $\eta$ .

**Proof:** We use the corresponding inverses of independent uncertainty distributions given as

$\Phi^{-1}(\alpha) = (1 - \alpha)a + \alpha b$  is strictly increasing linear in  $\xi$ .

$\Psi_2^{-1}(1 - \alpha) = \mu_2 - \frac{\sigma_2\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right)$  strictly decreasing normal in  $\eta$

The inverse distribution of the difference is expressed as

$$\begin{aligned} \Upsilon^{-1}(\alpha) &= \Phi^{-1}(\alpha) - \Psi^{-1}(1 - \alpha) \\ &= [(1 - \alpha)a + \alpha b] - \left[\mu - \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right)\right] \\ &= (a - \mu) + (b - a)\alpha + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \end{aligned}$$

By substituting the difference inverse, the distance is given by the expression

$$\begin{aligned} d(\xi, \eta) &= \int_{\Upsilon(0)}^1 \left\{ (a - \mu) + (b - a)\alpha + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \right\} d\alpha \\ &\quad - \int_0^{\Upsilon(0)} \left\{ (a - \mu) + (b - a)\alpha + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \right\} d\alpha \end{aligned} \quad (5.47)$$

First part of the integrand

$$\int_{\Upsilon(0)}^1 \left\{ (a - \mu) + (b - a)\alpha + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \right\} d\alpha$$

$$= \left[ (a - \mu)\alpha + (b - a)\frac{\alpha^2}{2} + \frac{\sigma\sqrt{3}}{\pi} 2(\alpha \ln \alpha - \alpha) \right]_{Y(0)}^1$$

Substituting the limits, we get

$$\begin{aligned} &= \left[ (a - \mu) \cdot 1 + (b - a)\frac{1^2}{2} + \frac{\sigma\sqrt{3}}{\pi} 2(1 \cdot \ln(1) - 1) \right] \\ &- \left[ (a - \mu)Y(0) + (b - a)\frac{Y(0)^2}{2} + \frac{\sigma\sqrt{3}}{\pi} 2(Y(0)\ln Y(0) - Y(0)) \right] \\ &= \left[ \frac{1}{2}(a + b) - \mu - 2\frac{\sigma\sqrt{3}}{\pi} \right] \\ &\quad - \left[ \frac{Y(0)^2}{2}(a + b) - Y(0)\mu + \frac{\sigma\sqrt{3}}{\pi} 2(Y(0)\ln Y(0) - Y(0)) \right] \quad (5.48) \\ &= \left[ \frac{1}{2}(a + b)(1 - Y(0)^2) + \{Y(0) - 1\}\mu - \frac{2\sigma\sqrt{3}}{\pi} \{1 + (Y(0)\ln Y(0) - Y(0))\} \right] \end{aligned}$$

Similarly, the second integrand of equation (5.47)

$$\begin{aligned} &- \int_0^{Y(0)} \left\{ (a - \mu) + (b - a)\alpha + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1 - \alpha}\right) \right\} d\alpha \\ &= - \left[ (a - \mu)\alpha + (b - a)\frac{\alpha^2}{2} + \frac{\sigma\sqrt{3}}{\pi} 2(\alpha \ln \alpha - \alpha) \right]_0^{Y(0)} \end{aligned}$$

Collecting terms, we obtain

$$= - \left[ (a - \mu)Y(0) + (b - a)\frac{\alpha Y(0)^2}{2} + \frac{2\sigma\sqrt{3}}{\pi} (Y(0)\ln Y(0) - Y(0)) \right] \quad (5.49)$$

By collecting the like terms, (3.48) and (3.49), we obtained the uncertainty distance  $d(\xi, \eta)$  is given as

$$\begin{aligned} &= \left[ \frac{1}{2}(a + b)(1 - Y(0)^2) + \{Y(0) - 1\}\mu \right. \\ &\quad \left. - 2\frac{\sigma\sqrt{3}}{\pi} - \left\{ (a - \mu)Y(0) + (b - a)\frac{Y(0)^2}{2} \right\} \right] \quad (5.50) \end{aligned}$$

## 5.7 Uncertainty Inference and Estimation

In this section, an uncertainty estimation method of expert based information is introduced. An uncertainty distribution is regarded as the population and plays a vital role in describing a

process, i.e. the target of uncertainty statistical inference. Hence, it is important to discuss the estimation procedures to evaluate uncertainty distributions. The procedure for survey questionnaire design and interpretation of experts' data is described in this section.

### 5.7.1 Expert experimental data

In classical statistics, data refers typically to randomly sampled observations. The notion of data includes natural observational data, survey data, and other observational data (e.g. experimental data). In classical statistics, data are repeatedly observed and objectively collected. Randomness is intrinsic to all data in statistics. Data may be classified according to the degree of abstraction into three levels: (raw data, information, and knowledge). However, in uncertainty modelling, impreciseness is intrinsic to uncertainty data, (e.g. expert knowledge) due to variations and imperfections in human observation and abstraction abilities, as well as to the complexity of the real world. In other words, data in uncertainty statistics may be narrowed to only the knowledge level, i.e., expert knowledge, no matter what its original character, objective or subjective.

The expert response knowledge elements are associated with degree of belief; hence, it is unrealistic to assume the uncertain variables are independent of expert knowledge. In other words, the expert opinions are recorded in real number format. However the expert knowledge responses should never be regarded as an ordinary number. Instead, the response is treated as a representative of an expert knowledge of the population parameter or equivalently of an uncertainty distribution of an expert knowledge response. Liu (2010) proposed a questionnaire survey for collecting expert data. One or more domain experts are invited independently to complete a questionnaire about knowledge of the value of an uncertain variable, say  $\xi$ .

In the first round, the interviewer asks "what was a likely quantity of production output per hectare of sesame last growing season? The expert may choose a possible value for  $x$  (say about 250 kg seed per hectare). The uncertain variable  $\xi$  is to be elicited. The expert is also asked: "How likely is  $\xi$  to be less than or equal to  $x$ ?" This degree of belief is denoted by  $\alpha$  (say 0.6). It should be noted that the degree of expert's belief on  $\xi$  is greater than  $x$  should be  $1 - \alpha$  due to the self-duality of uncertain measure. The data is thus acquired from the domain expert and quoted as an ordered pair  $(x, \alpha) = (250, 0.6)$ .

Repeating the above process, the experimental expert data are obtained by several such questioning and recorded as ordered pairs as

$$(x_1, \alpha_1), (x_2, \alpha_2) \dots \dots (x_n, \alpha_n)$$

In general, for a given set of one expert's experimental data, the following lemma must be satisfied:

**Lemma 5:** Let  $(x_{ij}, \alpha_{ij})$ ,  $i = 1, \dots, n_j$ ,  $j = 1, \dots, m$  be  $j$  expert's experimental data follow an orderly arranged pair for both sample data and associated degree of belief satisfy the condition for each  $i$  expert:

$$0 < x_{i1} < x_{i2} < \dots \dots < x_{in_j}, 0 < \alpha_{i1} < \alpha_{i2} < \dots \dots < \alpha_{in_j} \quad (5.51)$$

$$\text{for } j = 1, \dots \dots, m$$

### 5.7.2 Nonparametric Estimation Method

Nonparametric uncertain statistics does not rely on the experimental expert data belonging to any particular uncertainty distribution. In order to determine the uncertainty distributions, Liu (2010) proposed a linear interpolation method for estimating an empirical uncertainty distribution, while Chen and Ralescu (2010) proposed a series of spline interpolation methods. Given an experimental data of one expert, a linear interpolation method can be constructed to obtain the empirical uncertainty distribution defined by

$$\Phi(x) = \begin{cases} 0, & \text{if } x < x_i \\ \alpha_i + \frac{(\alpha_{i+1} - \alpha_i)(x - \alpha_i)}{x_{i+1} - x_i}, & \text{if } x_i \leq x < x_{i+1}, 1 \leq i < n \\ 1, & \text{if } x \geq x_n \end{cases} \quad (5.52)$$

In order to determine the expected value of the uncertain variable  $\xi$  of an empirical distribution defined by equation (5.52), this expected value explicitly expressed as

$$E(\xi) = \frac{\alpha_1 + \alpha_2}{2} x_1 + \sum_{i=2}^{n-1} \frac{\alpha_{i+1} - \alpha_{i-1}}{2} x_i + \left\{ 1 - \frac{\alpha_{n-1} + \alpha_n}{2} \right\} x_n \quad (5.53)$$

The experimental data of the expert is monotonically arranged according to equation (5.51) and by substituting into equation (5.53), we obtain an expected value of the empirical uncertainty distribution described by linear equation (5.52).

In addition, Wang and Peng (2010) suggested that if a known functional form for an empirical distribution can be established then the unknown parameters of the function can be determined by fitting the empirical data for the given function. They proposed a method of moments for estimating the unknown parameters of an uncertainty distribution. However, for multiple domain experts, Liu (2010) proposed a combined (pooled) uncertainty distribution given as

$$\Phi(x) = c_1\Phi_1(x) + c_2\Phi_2(x) + \dots + c_m\Phi_m(x) \quad (5.54)$$

where  $c_1, c_2, \dots, c_m$  are convex combination coefficients representing the weights of domain experts. For such a dataset from multiple domain experts, Wang *et al.* (2012) proposed the Delphi method to determine the uncertainty distribution.

### 5.7.3 Method of Least Squares

Liu (2010) first introduced the use of the principle of least squares to estimate the unknown parameters of an uncertainty distribution  $\Phi(x)$ , minimizing the sum of the squares of the distance of the experimental expert data to the uncertainty distribution.

Suppose the value of the uncertain variable  $\xi$  can be located in an open interval  $(a, b)$  with the same possibility, which means that  $\xi \sim L(a, b)$ . The uncertainty linear distribution of  $\xi$  can be expressed as

$$\Phi(x | \theta_1, \theta_2) = \begin{cases} 0 & \text{if } x < a \\ \theta_1 x + \theta_2 & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b \end{cases} \quad (5.55)$$

where  $\hat{\theta}_1 = \frac{1}{b-a}$ ,  $\hat{\theta}_2 = -\frac{a}{b-a}$ .

The principle of least squares to estimate the unknown parameters  $\theta$  given the expert data  $\{(x_1, \alpha_1), (x_2, \alpha_2) \dots \dots (x_n, \alpha_n)\}$ , the optimal solution  $(\hat{\theta}_1, \hat{\theta}_2)$  of equation (5.55) called the least squares estimates of  $(\theta_1, \theta_2)$ , is defined by

$$\min_{\theta_1, \theta_2} \sum_{i=1}^n (\Phi(x_i | \theta_1, \theta_2) - \alpha_i)^2 \quad (5.56)$$

where  $\Phi(x | \theta_1, \theta_2)$  is the uncertainty distribution.

For a linear distribution expression (5.55), we find

$$\hat{\theta}_1 = \frac{n\bar{x}\bar{\alpha} - \sum_{i=1}^n x_i \alpha_i}{n\bar{x}^2 - \sum_{i=1}^n x_i^2}$$

$$\hat{\theta}_2 = \bar{\alpha} - \hat{\theta}_1 \bar{x}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{\alpha} = \frac{1}{n} \sum_{i=1}^n \alpha_i$$

Alternatively, suppose  $\xi$  is an uncertain variable with uncertainty distribution  $\Phi(x_i|\theta_1 \theta_2 \dots \theta_p)$ , where  $(\theta_1 \theta_2 \dots \theta_p)$  are unknown parameters. Also, the uncertainty distribution  $\Phi(x_i|\theta_1 \theta_2 \dots \theta_p)$  is assumed to be regular, so that the inverse exists, i.e.  $\Phi^{-1}(x)$ . The principle of least squares equation (5.56) can be rewritten in terms of its inverse distribution as

$$\min_{\theta_1, \theta_2, \dots, \theta_p} \sum_{i=1}^n (\Phi^{-1}(\alpha: \theta_1, \theta_2, \dots, \theta_p) - x_i)^2 \quad (5.57)$$

where the unknown parameters  $\theta_i, i = 1, \dots, p$  are the solution of the minimization problem. If the inverse uncertainty distribution of the uncertain variable  $\xi$  is simple and easy to calculate, we should solve the optimization problem (5.57) rather than the optimization problem (5.56) for the unknown parameters.

#### 5.7.4 Method of Moments

**Theorem 5.10:** Liu (2010). Let  $\xi$  be a non-negative uncertain variable with uncertainty distribution  $\Phi$  and  $k^{th}$  be positive integer. Then, the  $k^{th}$  moment  $\xi$  is

$$E(\xi^k) = k \int_0^{+\infty} x^{k-1} (1 - \Phi(x)) dx \quad (5.58)$$

**Proof:** Since  $\xi$  is an uncertain variable, we can write

$$\begin{aligned} E(\xi^k) &= \int_0^{+\infty} \mathcal{M}\{\xi^k \geq x\} dx &= \int_0^{+\infty} \mathcal{M}\{\xi \geq x\} dx^k \\ E(\xi^k) &= k \int_0^{+\infty} x^{k-1} \mathcal{M}\{\xi \geq x\} dx &= k \int_0^{+\infty} x^{k-1} (1 - \Phi(x)) dx \end{aligned} \quad (5.59)$$

The theorem is proved.

Applying this theorem, if  $k$  is an odd number it follows from the definition of expected value operator that

$$E(\xi^k) = k \int_0^{+\infty} x^{k-1} (1 - \Phi(x)) dx - k \int_{-\infty}^0 x^{k-1} (\Phi(x)) dx \quad (5.60)$$

When the uncertain vector is nonnegative and is described by a distribution function  $\Phi(x|\boldsymbol{\theta})$  where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  is an unknown parameter vector, Wang and Peng (2010) proposed a method of moments as a method to estimate  $\boldsymbol{\theta}$ .

Suppose  $E(\xi^k)$  and  $E(\bar{\xi}^k)$ ,  $k = 1, 2, \dots, p$  be respectively the  $k^{th}$  theoretical moments for the distribution and  $k^{th}$  sample moments of the empirical uncertainty distribution for sample data satisfying (5.62).

A general procedure for estimating unknown parameters  $(\theta_1, \theta_2, \dots, \theta_p)$ , using  $k^{th}$  moments of empirical uncertainty distribution can be expressed by a system of equations

$$E(\xi^k) = \bar{\xi}^k, \quad k = 1, 2, \dots, p \quad (5.61)$$

where

$$\bar{\xi}^k = \alpha_1 x_1^k + \frac{1}{k+1} \sum_{i=2}^{n-1} \sum_{j=0}^k (\alpha_{i+1} - \alpha_{i-1}) x_k^j x_{i+1}^{k-j} + \{1 - \alpha_n\} x_n^k \quad (5.62)$$

Now, we find the roots of the above equations (5.61), by setting (5.60) to the systems of equations (5.62) of its equivalent theoretical moments

$$E(\xi^k) = k \int_0^{+\infty} x^{k-1} (1 - \Phi(x : \theta_1, \theta_2, \dots, \theta_n)) dx \quad (5.63)$$

A solution of the equation systems  $(\theta_1, \theta_2, \dots, \theta_p)$  obtained this way is called the vector of moment estimates.

### 5.7.5 Methods of Delphi

In this section, the Delphi method for uncertainty statistics is discussed in detail. The Delphi method was originally developed in the 1950's by the RAND Corporation, based on the assumption that group experience is more valid than individual experience. This approach is similar to probability theory which says that the greater the number of elements sampled, the higher the consistency of the sample estimate to the true source parameter value.

The Delphi method is characterized as an anonymous investigation which comprises several rounds of interactions of a group of experts in related fields. The same set of experts is repeatedly interviewed on the subject matter until their opinions tend to be consistent, and a composite survey result is compiled. The Delphi method had earlier been reported and successfully applied in different disciplines: (cf. Ludwig 1994).

The participants are asked to first assess the matters independently and make individual judgment according to individual knowledge and experience. For the second round, the participants are provided with the feedback on the previous submission so that they can

assess the same matter again and make a new judgment about possibly altering their opinions. The final consensus is used to estimate the uncertainty distribution  $\Phi(x)$  of an uncertain variable  $\xi$  from a group of domain experts. In uncertain statistics, Wang *et al.* (2012) introduced the Delphi method and provided a step by step procedure. The procedure is summarized below.

Suppose  $m$  unreplicated data points are selected from experimental data of  $m$  experts. Without loss of generality, all provided  $m$  data values are carefully arranged in ascending order of size such that  $0 < x_1 < \dots < x_m$  according to the condition in equation (5.51).

**Step 1:** The  $m$  data from  $n$  experts to form a distinct data set  $(x_{ij}, \alpha_{ij})$ , where  $x_{ij}$  denote the  $ij^{th}$  possible values provided by the  $i^{th}$  expert on the  $j^{th}$  value and  $\alpha_{ij}$  denote the  $ij^{th}$  degree of belief attached to uncertain variable  $\xi_i$  is less than  $x_{ij}$ ,  $i = 1, \dots, n_j$ ,  $j = 1, \dots, m$  respectively.

**Step 2:** Use the  $i^{th}$  expert data  $(x_i, \alpha_i)$  to generate the uncertainty distributions  $\Phi(x)$  of the  $i^{th}$  domain expert  $j = 1, \dots, m$ , respectively.

**Step 3:** List the number  $n$  of all possible values of  $\xi$  presented by the experts, say 5 experts in order of magnitude, where there are ties in values presented by different experts; the tied value is treated as one value. Then the possible values are  $0 < x_1 < \dots < x_{n_j}$ ,

$j = 1, \dots, m$ , are tabulated along the columns and the corresponding degree of belief  $\alpha_n$  as provided by each expert as expert distribution on the first column against  $m$  distinct data set. We score the  $i^{th}$  expert zero if he did not list that value. The following computations are necessarily performed:

$$\hat{\alpha}_j = \frac{1}{m} \sum_{i=1}^m \Phi_i(x_j) \quad j = 1, 2, \dots, m, i = 1, 2, \dots, 5 \quad (5.64)$$

and

$$d_j = \frac{1}{m} \sum_{i=1}^m (\Phi_i(x_j) - \hat{\alpha}_j)^2; \quad j = 1, 2, \dots, m, i = 1, 2, \dots, 5 \quad (5.65)$$

**Step 4:** If  $d_j$  is less than a given level  $\varepsilon_j$  for all  $j$ , then we go to Step 5. Otherwise, the  $j^{th}$  data receive the current summary of  $\hat{\alpha}_i$  and then provide the feedback to the experts for possible adjustments, say  $(x_{i1}, \hat{\alpha}_{i1}); (x_{i2}, \hat{\alpha}_{i2}); \dots, (x_{in}, \hat{\alpha}_{i5})$  for  $j = 1, 2, \dots, 5$  and repeat Step 2.

**Step 5:** The aggregated data  $(x_{i1}, \hat{\alpha}_{i1}); (x_{i2}, \hat{\alpha}_{i2}); \dots, (x_{in}, \hat{\alpha}_{i5})$  is then used to compute the empirical distribution  $\Phi_j(x), j = 1, \dots, 5$ .

## 5.8 Summary

This chapter discussed the uncertainty theory and relevant concepts. It established the fundamental concepts of uncertainty statistics and made comparison between probability, credibility and uncertainty theory using measure theory. The chapter established a relationship between a standard uncertainty normal distribution and logistic probability distribution as a transformed Gaussian probability distribution. The uncertainty theory can be interpreted as the distribution of degree of belief, while a probability theory represents a frequency distribution. A technique for elicitation of information from expert domain knowledge was described and an estimation method for uncertainty distributions proposed. We also introduced a Delphi method for multiple experts and proposed distance measure for difference between two uncertainty distributions.

## 6. Results - Analysis of Crop Yields Data

### Using Uncertainty Statistics

This chapter presents the results of analysis carried out on two datasets from expert survey described in Section 2.6, in Chapter 2. Uncertainty empirical distributions for the five cereal crops yields were first estimated and descriptive estimates computed. The estimation method discussed in chapter 5 was then applied to analyze and fit some uncertainty distributions and a deterministic function for the maize yield. The Delphi method was implemented on sesame seed data as it combined multiple expert data and the determination of the distribution for the composite data. Finally, the distance between the expert based model and a standard uncertainty normal distribution was evaluated.

#### 6.1 Estimation of Uncertainty Distributions

In this section, the parameters of empirical uncertainty distributions are estimated by applying the linear interpolation method and the method of moments.

##### 6.1.1 Empirical Uncertainty Distributions

We present the results of empirical estimates on the crop yields. The linear interpolation method of equation (5.63) was employed to generate an empirical uncertainty distribution for the data of each crop listed in Table 2.2.

##### Maize yield:

The empirical uncertainty distribution for maize was

$$\Phi(x)_{maize} = \begin{cases} 0 & \text{if } x < 1.8 \\ 0.1 + 2(x - 1.8) & \text{if } 1.8 \leq x < 2.0 \\ 0.50 + 0.5(x - 2.0) & \text{if } 2.0 \leq x < 2.5 \\ 0.75 + 1.25(x - 2.5) & \text{if } 2.5 \leq x < 2.7 \\ 1 & \text{if } x \geq 2.7 \end{cases} \quad (6.1)$$

Hence, the empirical uncertainty distribution (6.1) for maize had expected value,  $E(\xi) = 2.153 \text{ ton/ha}$  and variance  $0.0977 \text{ (ton / ha)}^2$ .

Rice yield: The empirical uncertainty distribution for rice was

$$\Phi(x)_{\text{rice}} = \begin{cases} 0 & \text{if } x < 0.8 \\ 2(x - 0.8) & \text{if } 0.8 \leq x < 1.0 \\ 0.40 + 3/2(x - 1.0) & \text{if } 1.0 \leq x < 1.2 \\ 0.70 + 2/3(x - 1.2) & \text{if } 1.2 \leq x < 1.5 \\ 0.90 + 1/5(x - 1.5) & \text{if } 1.5 \leq x < 2.0 \\ 1 & \text{if } x \geq 2.0 \end{cases} \quad (6.2)$$

The empirical uncertainty distribution (6.2) for rice had expected value  $E(\xi) = 1.135 \text{ ton/ha}$  and variance  $0.0754 \text{ (ton/ha)}^2$ .

**Sorghum yield:** The empirical uncertainty distribution for sorghum was

$$\Phi(x)_{\text{sorghum}} = \begin{cases} 0 & \text{if } x < 0.8 \\ 0.2 + 3/2(x - 0.8) & \text{if } 0.8 \leq x < 1.0 \\ 0.5 + 2(x - 1.0) & \text{if } 1.0 \leq x < 1.2 \\ 0.9 + 1/3(x - 1.2) & \text{if } 1.2 \leq x < 1.5 \\ 1 & \text{if } x \geq 1.5 \end{cases} \quad (6.3)$$

The empirical uncertainty distribution (6.3) for sorghum had expected value  $E(\xi) = 1.005 \text{ ton/ha}$  and variance  $0.0303 \text{ (ton/ha)}^2$ .

**Soybean yield:** The empirical uncertainty distribution for soybean was

$$\Phi(x)_{\text{soybean}} = \begin{cases} 0 & \text{if } x < 1.0 \\ 0.2 + 3/5(x - 1.0) & \text{if } 1.0 \leq x < 1.5 \\ 0.50 + 5/6(x - 1.5) & \text{if } 1.5 \leq x < 1.8 \\ 0.75 + 5/4(x - 1.8) & \text{if } 1.8 \leq x < 2.0 \\ 1 & \text{if } x \geq 2.0 \end{cases} \quad (6.4)$$

The empirical uncertainty distribution (6.4) for soybean had expected value  $E(\xi) = 1.4625 \text{ ton/ha}$  and variance  $0.1219 \text{ (ton/ha)}^2$ .

**Cowpea yield:** The empirical uncertainty distribution for cowpea was

$$\Phi(x)_{cowpea} = \begin{cases} 0 & \text{if } x < 0.8 \\ 2(x - 0.8) & \text{if } 0.8 \leq x < 1.0 \\ 0.5 + 3/2(x - 1.0) & \text{if } 1.0 \leq x < 1.2 \\ 0.8 + 1/3(x - 1.2) & \text{if } 1.2 \leq x < 1.5 \\ 0.9 + 1/3(x - 1.5) & \text{if } 1.5 \leq x < 1.8 \\ 1 & \text{if } x \geq 1.8 \end{cases} \quad (6.5)$$

The empirical uncertainty distribution (6.5) for soybean had expected value  $E(\xi) = 1.100 \text{ ton/ha}$  and variance  $0.0567 (\text{ton/ha})^2$ .

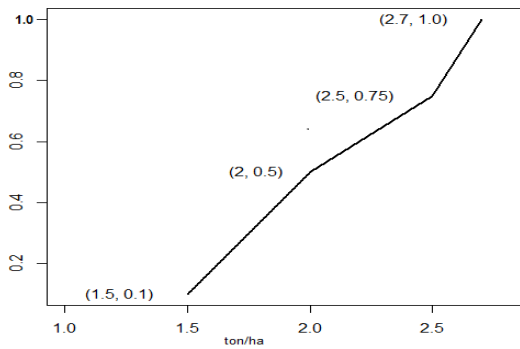
### 6.1.2 Results of Analysis of Uncertainty Empirical Estimation

The crop yields data in Table 2.2 were analysed in R environment to compute the first and the second moments of the uncertainty empirical distributions. The results are presented in Table 6.1. The table shows the actual crop yields and the expectations of crop yield as described by uncertainty empirical distribution formula equation (5.63). It can be seen from Table 6.1 that the percentage relative deviations between the farmer's actual yields and estimated yields ranged between 0.50% and 10.0%. The minimum percent yield deviation (0.50%) was recorded under sorghum and the maximum percent deviation (10.0%) recorded for cowpea.

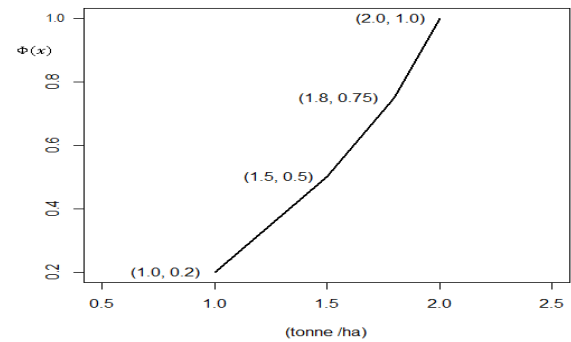
**Table 6.1:** Summary of Observed and Expected values of Crop yields

	Crop yield (ton/ha)				
	Maize	Rice	Sorghum	Soybean	Cowpea
Actual value	2.200	1.200	1.010	1.500	1.000
Estimated value	2.153	1.135	1.005	1.463	1.100
Estimated Error	2.13%	5.04 %	0.50%	2.5 0%	10.0%

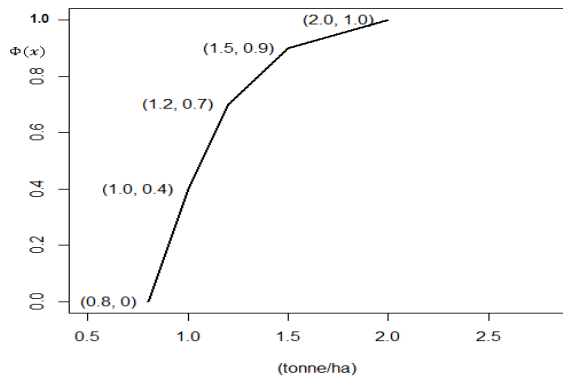
The graphs of empirical distributions of cereal crop yields are displayed in Figures 6.1to 6.5 as provided for maize, rice, sorghum, soybean and cowpea respectively.



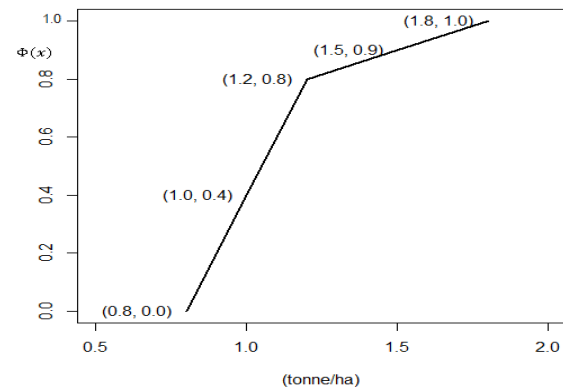
**Figure 6.1:** Expert Distribution for Maize Yields



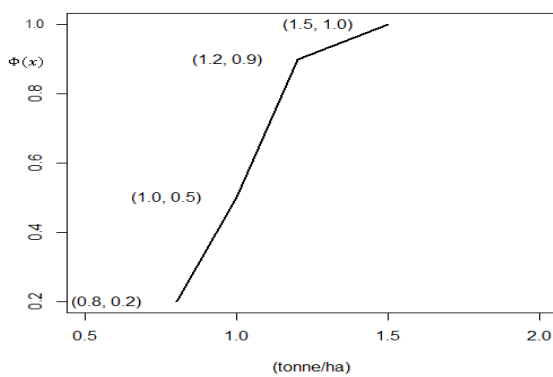
**Figure 6.4:** Expert Distribution for Soybean Yields



**Figure 6.2:** Expert Distribution for Rice Yields



**Figure 6.5:** Expert Distribution for Cowpea Yields



**Figure 6.3:** Expert Distribution for Sorghum Yields

## 6.2 Results of Estimation of Uncertainty Distributions

In this section, the method of moments and principle of least squares are applied to fit some uncertainty distributions (Linear and Normal) on the crop yield data as provided by the experts.

### **Uncertainty Linear Distribution**

We estimate the parameters of the uncertainty linear distribution defined on  $\xi \sim L(a, b)$  as defined by

$$\Phi(x | \theta_1, \theta_2) = \begin{cases} 0 & \text{if } x < a \\ \theta_1 x + \theta_2 & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b \end{cases} \quad (6.6)$$

where  $\hat{\theta}_1 = \frac{1}{b-a}$ ,  $\hat{\theta}_2 = -\frac{a}{b-a}$  and  $a < b$ .

To solve for the parameters of the linear distribution (6.6), we apply equations (5.73) through (5.75) to estimate first and second moments for the experts' data and setting the values to the right-hand side terms of the following equations

$$E(\xi) = \int_0^{+\infty} (1 - \Phi(x | \theta_1, \theta_2)) dx = \frac{a+b}{2}$$

and

$$E(\xi^2) = \int_0^{+\infty} (1 - \Phi(x | \theta_1, \theta_2)) dx = \frac{a^2 + ab + b^2}{3}$$

### **Uncertainty Normal Distribution**

We also fit the crop yield data to the uncertainty normal distribution

$$\Phi(x) = \left[ 1 + \exp\left(\frac{\pi(\mu - x)}{\sigma\sqrt{3}}\right) \right]^{-1}, \quad x \in R$$

where  $\mu$  and  $\sigma$  are the unknown parameters to be estimated. The corresponding inverse uncertainty distribution  $N(\mu, \sigma)$  is given as

$$\Phi^{-1}(\alpha) = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{\alpha}{1-\alpha}\right) \quad \alpha \in (0,1)$$

By applying our estimation method, we obtain the following estimates

$$\hat{\mu} = \bar{x} - \frac{\sigma\sqrt{3}}{n\pi} \sum_{i=1}^n G_i$$

and

$$\hat{\sigma} = \frac{n\pi\sqrt{3}}{3} \left\{ \frac{\bar{x} \sum_{i=1}^n G_i - \sum_{i=1}^n x_i G_i}{(\sum_{i=1}^n G_i)^2 - n \sum_{i=1}^n G_i^2} \right\}$$

where  $\bar{x} = (x_1 + \dots + x_n)/n$  and  $G_i = \left(\frac{\alpha_i}{1-\alpha_i}\right)$

#### **6.2.1 Results of Uncertainty Linear Distribution analysis**

We fitted the uncertainty linear uncertainty and normal distributions to the five cereal crop yields as provided by the experts, the results are presented in this section. The analysis involved the computation of the first and second moments, uncertainty mean and variance of the linear model and normal distributions.

**Maize:** The first and second moment respectively were  $\bar{\xi} = 2.1525$  and  $\bar{\xi}^2 = 4.731$

The maize yield had an uncertainty linear distribution on  $L(1.611, 2.694)$  with linear function determined as  $\Phi(x) = 0.9235x - 1.4875$ . This uncertainty distribution of maize yield could be expressed by the equation

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 1.611 \\ 0.9235x - 1.4875 & \text{if } 1.611 \leq x < 2.694 \\ 1 & \text{if } x \geq 2.694 \end{cases} \quad (6.7)$$

In contrast, the observed yield was 2.200 *ton/ha* and  $Var[\Phi(x)] = Var(aX + b) = a^2 Var(X) = 0.08332 \text{ (ton / ha)}^2$

**Rice:** The first and second moments are estimated respectively as  $\bar{\xi} = 1.135$  and  $\bar{\xi}^2 = 1.364$ . The rice yield had an uncertainty linear distribution on  $L(0.682, 1.6118)$ , and a linear function given as  $\Phi(x) = 1.04875x - 0.6903$ . The uncertainty distribution for rice yield could be described by the equation

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 0.682 \\ 1.0487x - 0.6903 & \text{if } 0.682 \leq x < 1.6118 \\ 1 & \text{if } x \geq 1.6118 \end{cases} \quad (6.8)$$

The mean yield is 1.200 *ton/ha* and variance is 0.0829  $\text{(ton / ha)}^2$

**Sorghum:** The first and second moment respectively were  $\bar{\xi} = 1.005$  and  $\bar{\xi}^2 = 1.0403$ .

The sorghum yield had an uncertainty linear distribution on  $L(0.7035, 1.3065)$  with linear function given as  $\Phi(x) = 1.6593x - 1.667$ . The uncertainty distribution was

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 0.7035 \\ 1.6593x - 1.667 & \text{if } 0.7035 \leq x < 1.3065 \\ 1 & \text{if } x \geq 1.3065 \end{cases} \quad (6.9)$$

In contrast, the mean yield was 1.010 *ton/ha* and variance 0.0834  $\text{(ton / ha)}^2$

**Soybean:** The first and second moment respectively were  $\bar{\xi} = 1.4625$  and  $\bar{\xi}^2 = 2.26083$ .

The soybean yield had an uncertainty linear distribution on  $L(0.8577, 2.0673)$  with linear function determined as  $\Phi(x) = 0.8267x - 0.7091$ . The uncertainty distribution was given by

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 0.8577 \\ 0.8267x - 0.7091 & \text{if } 0.8577 \leq x < 2.0673 \\ 1 & \text{if } x \geq 2.0673 \end{cases} \quad (6.10)$$

The observed yield was 1.500 *ton/ha* and variance 0.0833  $\text{(ton / ha)}^2$ .

**Cowpea:** The first and second moment respectively were  $\bar{\xi} = 1.021$  and  $\bar{\xi}^2 = 1.26667$ . The cowpea yield had an uncertainty linear distribution on  $L(0.2008, 1.8412)$  with linear function determined as  $\Phi(x) = 0.6096x - 1.224$ . This uncertainty distribution of cowpea yield could be described by the equation

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 0.2008 \\ 0.6096x - 0.1224 & \text{if } 0.2008 \leq x < 1.8412 \\ 1 & \text{if } x \geq 1.8412 \end{cases} \quad (6.11)$$

The mean yield was 1.000 *ton/ha* and variance 0.02107 (*ton / ha*)<sup>2</sup>

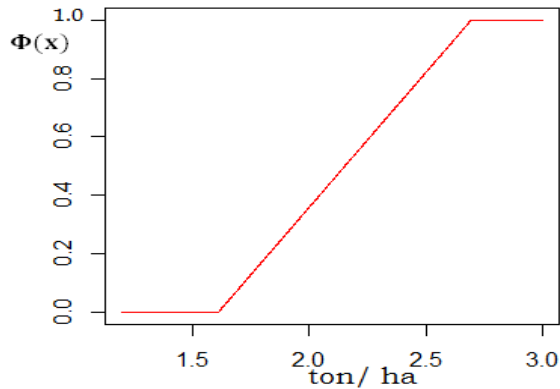
### Results Summary

The data in Table 2.2 was subjected to R programming (code shown in Appendix C.2) to compute the first and second moments for an uncertainty linear distribution. The resulting estimates of parameters of uncertainty linear distributions are presented in Table 6.2 alongside the mean and variance of the uncertainty linear distribution. The estimated values for respective crops are also tabulated along with first moment and variance, linear function parameters, lower and upper bounds.

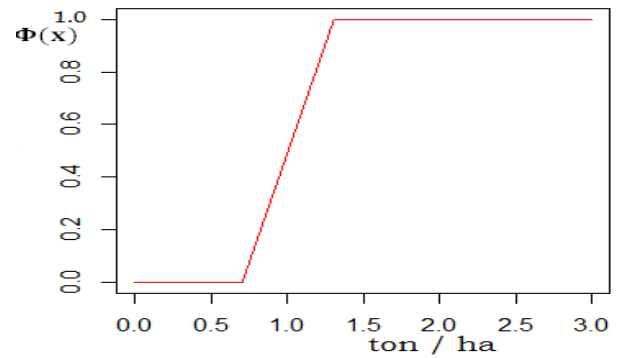
**Table 6.2:** Estimates of Uncertainty Linear Distribution of Crop yields

	Crop yield ( <i>tons per hectare</i> )				
	First moment	Variance	Linear function	Lower bound	Upper bound
Maize	2.153	0.0833	$0.9235x - 1.4875$	1.611	2.694
Rice	1.135	0.0829	$1.0488x - 0.6903$	0.658	1.6118
Sorghum	1.005	0.0834	$1.6593x - 1.1667$	0.7035	1.3065
Soybean	1.463	0.0833	$0.8267x - 0.7091$	0.8577	2.0673
Cowpea	1.021	0.0211	$0.6096x - 1.224$	0.2008	1.8412

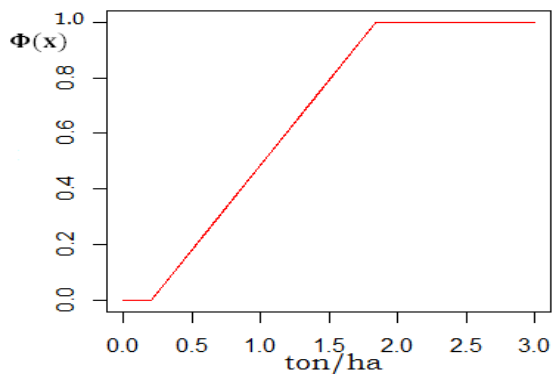
The graphs of the linear distributions are presented in Figures 6.6 to 6.9 for maize, rice, sorghum, soybean and cowpea respectively.



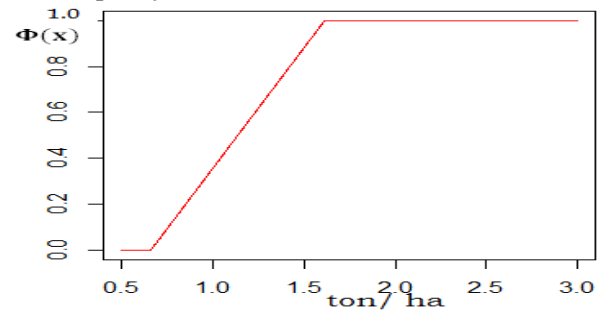
**Figure 6.6:** Uncertainty Linear Distribution for Maize yield



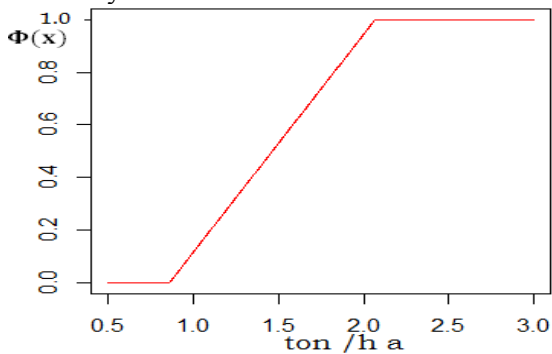
**Figure 6.9:** Uncertainty Linear Distribution for Cowpea yield.



**Figure 6.7:** Uncertainty Linear Distribution for Rice yield



**Figure 6.10:** Uncertainty Linear Distribution for Cowpea Yield



**Figure 6.8:** Uncertainty Linear Distribution for Sorghum yield

## 6.2.2 Results of Uncertainty Normal Distribution analysis

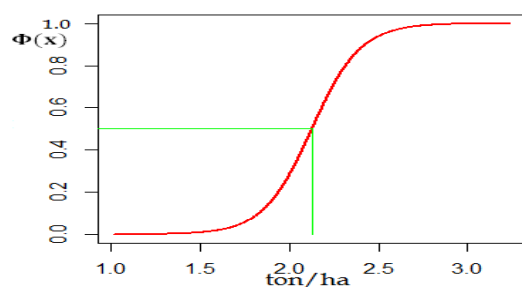
The data in Table 2.2 was analyzed in R code presented in the Appendix C.3 to compute the location and spread of uncertainty normal distribution. In order to perform the computation of these estimates, recognizing that the graphs of uncertainty normal distributions and their inverses have asymptotes at zero and one, that is  $\alpha \in (0,1)$  is open, the lower and upper limit values were adjusted by replacing  $\alpha = 0$  by 0.01 and  $\alpha = 1$  by 0.99. The estimates are presented in Table 6.3 for respective crop yields.

**Table 6.3:** Summary estimates of Uncertainty Normal Distribution of Crop yields (*ton/ha*)

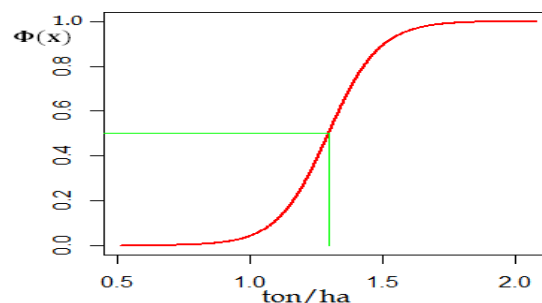
Crops	Data Point	Mean $\mu$	variance $\sigma^2$	Normal function	observed value
Maize	4	2.1277	0.24621	N(2.128, 0.246)	2.200
Rice	5	1.2937	0.17397	N(1.294, 0.174)	1.200
Sorghum	4	0.9713	0.20632	N(0.971, 0.206)	1.010
Soybean	4	1.4104	0.27729	N(1.410, 0.277)	1.500
Cowpea	5	1.1906	0.19801	N(1.191, 0.198)	1.000

The graphs of the fitted uncertainty normal distributions are displayed in the Figures

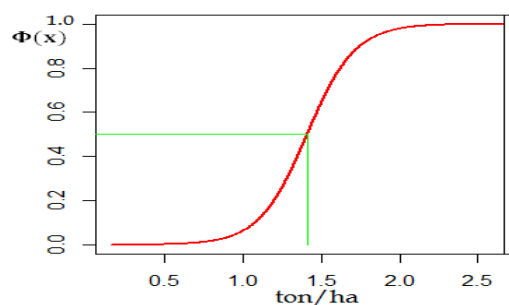
6.11 to 6.17



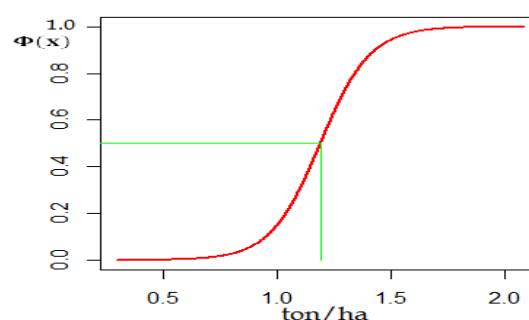
**Figure 6.11:** Uncertainty Normal Distribution for maize



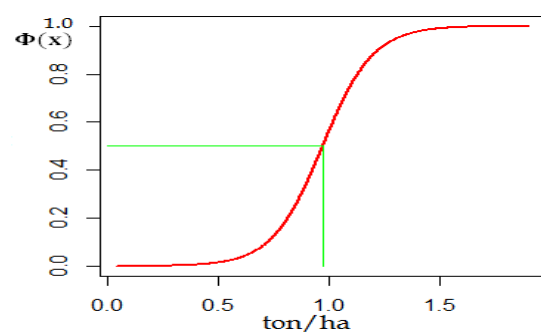
**Figure 6.12:** Uncertainty Normal Distribution for Rice yield



**Figure 6.13:** Uncertainty Normal for Sorghum yield.



**Figure 6.15:** Uncertainty Normal for Cowpea yield



**Figure 6.14:** Uncertainty normal distributions for Soybean yield

### 6.3 Fitting data to a Deterministic Production function

The method of uncertainty statistics is applied to fit a deterministic function to the maize yield response to NPK fertilizer.

Let  $\xi$  be an uncertain variable (maize yield per plot) with regular uncertainty distribution  $\Phi(x)$ . Assume  $f(x) = Ax^{1/2}$  a strictly monotone function of  $x$ , where  $x$  is quantity ( $kg/ha$ ) of nitrogen fertilizer.

Suppose  $(x_i, \alpha_i), i = 1, \dots, n; i = 1; \dots; n$  are the expert's experimental data satisfying condition (5.62) and the uncertainty distribution of the uncertain variable  $\xi$  has a functional form with an unknown parameter  $k$  can be defined as

$$\Phi(x; k) = Ax^{\frac{1}{2}} \quad (6.12)$$

By applying uncertainty expectation of a function equation (5.29) with respect to the uncertainty distribution,  $E[f(x)] = \int_{-\infty}^{+\infty} f(x) d\Phi(x)$ , we determine the value of the unknown parameter  $A$  by setting the empirical expectation  $E[f(\xi)] = \bar{\xi}$ .

Hence,  $(\xi) = \frac{1}{3A^2}$ , we generate a single equation with one unknown as

$$\frac{1}{3A^2} = \frac{\alpha_1 + \alpha_2}{2} x_1 + \sum_{i=2}^{n-1} \frac{\alpha_{i+1} - \alpha_{i-1}}{2} x_i + \left\{ 1 - \frac{\alpha_{n-1} + \alpha_n}{2} \right\} x_n \quad (6.13)$$

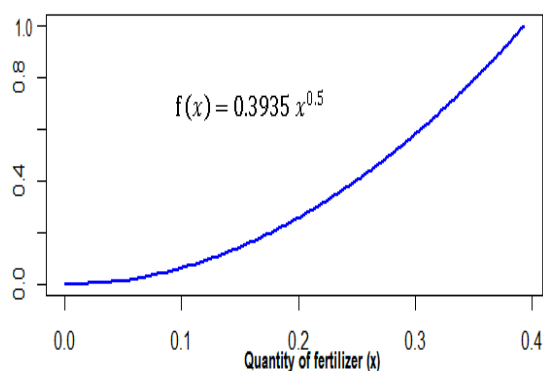
We solve for parameter  $A$  as

$$\frac{1}{3A^2} = \left[ 3 \left\{ \frac{\alpha_1 + \alpha_2}{2} x_1 + \sum_{i=2}^{n-1} \frac{\alpha_{i+1} - \alpha_{i-1}}{2} x_i + \left\{ 1 - \frac{\alpha_{n-1} + \alpha_n}{2} \right\} x_n \right\} \right]^{-1/2} \quad (6.14)$$

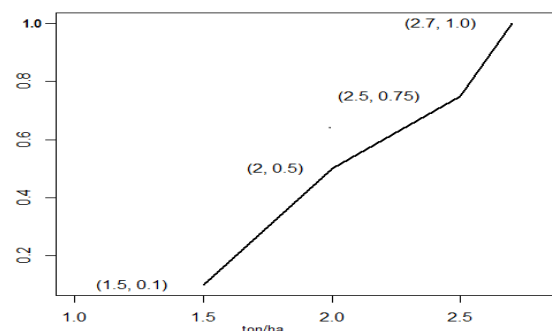
Referring to the maize data in Table (5.1) with first sample moment for the maize data

$\xi = 2.153$  the value of the unknown parameter was  $\hat{A} = 0.3935$  and hence, the particular function became  $f(x) = 0.3935 x^{1/2}$ .

The graph of the fitted function and expert distribution is presented in Figure 6.16 and 6.17



**Figure 6.16:** Fitted Deterministic function on Maize yield response to NPK fertilizer



**Figure 6.17:** Expert Distribution for the Maize yields

## 6.4 Estimating Uncertainty Distribution from Multiple Expert Data

In this section, an aggregation method and a Delphi method is applied to sesame yield reported by multiple experts.

### 6.4.1 Aggregate Method

All possible values provided by the five agriculturists were listed: 230, 240, 250, 260, 270, 275, 280, 290, 300 un-replicated and the weighted aggregate were determined. Then, the empirical distribution was constructed according to equation (5.63). From the composite data, five empirical uncertainty distributions  $\Phi_i^{(1)}(x)$ ,  $i = 1, \dots, 5$  could be generated. A combined (pooled) uncertainty distribution was generated by applying equation (5.54) given as

$$\Phi(x) = \frac{1}{5} \{ \Phi_1(x) + \Phi_2(x) + \dots + \Phi_5(x) \} \quad (6.15)$$

where  $c_1, \dots, c_m = \frac{1}{5}$  are convex combination coefficients representing equal weights of the each expert, indicating that each expert has equal knowledge of the subject matter.

The estimated value of  $\alpha_j$  was the averaged distribution over the number of all experts for every listed uncertain variable. The resulting composite data is presented in Table 6.4.

**Table 6.4:** Composite data from multiple agricultural experts

$j$	1	2	3	4	5	6	7	8	9
$x_{ij}$	230	240	250	260	270	275	280	290	300
$\alpha_j$	0.15	0.40	0.45	0.50	0.60	0.64	0.76	0.90	1

The empirical uncertainty distribution for composite sesame data was determined as

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 230 \\ 0.15 + 1/40(x - 230) & \text{if } 230 \leq x < 240 \\ 0.40 + 1/200(x - 240) & \text{if } 240 \leq x < 250 \\ 0.45 + 1/200(x - 250) & \text{if } 250 \leq x < 260 \\ 0.50 + 1/100(x - 260) & \text{if } 260 \leq x < 270 \\ 0.60 + 1/100(x - 270) & \text{if } 270 \leq x < 275 \\ 0.64 + 1/125(x - 275) & \text{if } 275 \leq x < 280 \\ 0.76 + 7/50(x - 280) & \text{if } 280 \leq x < 290 \\ 0.90 + 1/100(x - 290) & \text{if } 290 \leq x < 300 \\ 1 & \text{if } x \geq 300 \end{cases} \quad (6.16)$$

The empirical uncertainty distribution (6.16) for sesame had expected value  $E(\xi) = 258.35 \text{ kg/ha}$  and variance  $562.611 \text{ kg/ha}$

### 6.4.2 Analysis of Multiple Experts' Data Using Delphi Method

In this section, data provided by the multiple agricultural experts are fitted by the Delphi method and results are presented in Table 6.5. The Delphi method employs the method of moments on the composite data. At first, all possible values of the uncertain variable  $\xi$  and their associated degree of belief (intuition grade) are listed as ordered pairs  $(x_{ij}, \alpha_j)$ .

**Table 6.5:** First round Data Computation for multiple agricultural experts

$j$	1	2	3	4	5	6	7	8	9
$x$	230	240	250	260	270	275	280	290	300
$\Phi_1^{(1)}(x)$	0	0	0	0.4	0.6	0.80	0.95	1.0	1.0
$\Phi_2^{(1)}(x)$	0	0	0	0.0	0.4	0.5	0.60	0.80	0.9
$\Phi_3^{(1)}(x)$	0	0.1	0.3	0.5	0.7	0.9	1.0	1.00	1.0
$\Phi_4^{(1)}(x)$	0	0.0	0.2	0.4	0.6	0.8	0.9	1.0	1.0
$\Phi_5^{(1)}(x)$	0.15	0.4	0.8	0.99	1.0	1.0	1.0	1.0	1.0

Equations (5.64) and (5.65) are applied to compute the aggregated degree of belief and the corresponding mean square deviation

$$\hat{\alpha}_j = \frac{1}{5} \sum_{i=1}^5 \Phi_i(x_j), \quad j = 1, \dots, 9 \quad (6.17)$$

and

$$d_j = \frac{1}{5} \sum_{i=1}^5 (\Phi_i(x_j) - \alpha_j)^2, \quad j = 1, \dots, 9 \quad (6.18)$$

Using equations (6.17) and (6.18) yielded the results in Table 6.6, finding  $d_3^{(1)} = 0.086 > 0.05$  and  $d_4^{(1)} = 0.1003 > 0.05$ . Each agriculturist was provided with the feedback result in Table 6.6.

**Table 6.6:** Decision Table

$j$	1	2	3	4	5	6	7	8	9
$x_{ij}^{(1)}$	230	240	250	260	270	275	280	290	300
$\alpha_j$	0.03	0.1	0.26	0.458	0.68	0.80	0.89	0.96	0.98
$d_j^{(1)}$	0.0036	0.024	0.086	0.1003	0.038	0.028	0.0224	0.006	0.002

On the basis of the outcome of Table 6.6, each expert was expected to adjust the earlier information provided. For instance,  $A_5$  had scored value 250 as 0.8 and  $A_4$  scored the same

value as 0.2. These series indicated that  $A_5$  gave the higher scores above the pooled average while the  $A_4$  score was too low. The experts are expected to compromise using the group mean, while  $A_3$  could retain the current score 0.3 for the same value 250.

An expert may have special intuition on some values based on degree of belief and knowledge. He may decide to provide grading much lower for a value than others. It is also evident that experts will provide different data values for the same intuition. The experts discarded the subjective data for 230 and 240, and brought in data for 295.

The revised information for the sesame seed yield production was then re-evaluated by the five agricultural experts and the new scores are presented in Table 6.7 below. The agronomist  $A_5$  had realized that he had reported a very least values 230 *kg/ha* and 240 *kg/ha* among all the experts and at the same time rated these values higher 15% and 40% respectively. After the experts were given the feedback of their first estimates,  $A_5$  needed to compromise with others and dropped these values and brought in 250 *kg/ha*, while  $A_3$  dropped 240 and brought 250. The revised data points come down to 8 instead of 9 in previous session. Equation (6.17) and criterion (6.18) were applied again on the revised data from the experts and we obtained Table 6.7.

$$\hat{\alpha}_j^{(2)} = \frac{1}{5} \sum_{i=1}^5 \Phi_i(x_j), \quad j = 1, \dots, 8 \quad (6.19)$$

and

$$d_j^{(2)} = \frac{1}{5} \sum_{i=1}^5 (\Phi_i(x_j) - \alpha_j)^2, \quad j = 1, \dots, 8 \quad (6.20)$$

**Table 6.7:** Adjusted data after First round computation and feedback

A1 :	(260, 0.4)	(270, 0.70)	(275, 0.8)	(280, 0.95)	(290, 1.0)
A2 :	(270, 0.6)	(280, 0.75)	(290, 0.8)	(300, 0.90)	
A3 :	(250, 0.3)	(260, 0.50)	(270, 0.6)	(280, 0.80)	(295, 1.0)
A4 :	(260, 0.35)	(270, 0.50)	(275, 0.6)	(280, 0.80)	(290, 0.9)
A5 :	(250, 0.4)	(260, 0.60)	(270, 0.8)	(275, 0.90)	

**Table 6.8:** Second round iteration computation for multiple agricultural experts

$j$	1	2	3	4	5	6	7	8
$x$	250	260	270	275	280	290	295	300
$\Phi_1^{(2)}(x)$	0.40	0.70	0.80	0.95	1.00	1.00	1.00	1.00
$\Phi_2^{(2)}(x)$	0.00	0.00	0.60	0.70	0.75	0.95	0.99	1.00
$\Phi_3^{(2)}(x)$	0.30	0.50	0.60	0.80	0.90	1.00	1.00	1.00
$\Phi_4^{(2)}(x)$	0.00	0.35	0.50	0.60	0.80	0.90	1.00	1.00
$\Phi_5^{(2)}(x)$	0.40	0.60	0.80	0.90	1.00	1.00	1.00	1.00

**Table 6.9:** Decision Table

$j$	1	2	3	4	5	6	7	8
$x_{ij}^{(2)}$	250	260	270	275	280	290	295	300
$\alpha_j^2$	0.22	0.43	0.66	0.81	0.89	0.96	0.975	1
$d_j^{(2)}$	0.034	0.013	0.014	0.022	0.0104	0.0024	0.001	0

From Table (6.9),  $\max_{1 \leq i \leq n} \{d_j^{(2)}\} = d_1^{(2)} = 0.034 < 0.05 = \varepsilon$

and  $\alpha_j^2 = \frac{1}{5} \sum_{i=1}^5 \Phi_i(x_j)$ ,  $j = 1, \dots, 8$

Thus, the process was terminated and integrated expert data for the five agriculturists generated as: (250, 0.22); (260, 0.43); (270, 0.66); (275, 0.81); (280, 0.89); (290, 0.96); (295, 0.975); (300, 1).

By applying equation (5.64), the empirical uncertainty distribution for integrated expert data for the five agriculturists was obtained as:

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 250 \\ 0.22 + 0.21/10(x - 250) & \text{if } 240 \leq x < 250 \\ 0.43 + 0.23/10(x - 260) & \text{if } 250 \leq x < 260 \\ 0.66 + 0.15/5(x - 270) & \text{if } 260 \leq x < 270 \\ 0.81 + 0.08/5(x - 275) & \text{if } 270 \leq x < 275 \\ 0.89 + 0.07/10(x - 280) & \text{if } 275 \leq x < 280 \\ 0.96 + 0.015/5(x - 290) & \text{if } 290 \leq x < 295 \\ 0.975 + 0.025/5(x - 295) & \text{if } 295 \leq x < 300 \\ 1 & \text{if } x \geq 300 \end{cases} \quad (6.21)$$

The empirical uncertainty distribution is given by equation (6.21) after collection of terms

$$\Phi(x) = \begin{cases} 0 & \text{if } x < 230 \\ 0.021x - 5.03 & \text{if } 250 \leq x < 260 \\ 0.023x - 5.55 & \text{if } 260 \leq x < 270 \\ 0.030x - 7.44 & \text{if } 270 \leq x < 275 \\ 0.016x - 3.59 & \text{if } 275 \leq x < 280 \\ 0.007x - 1.07 & \text{if } 280 \leq x < 290 \\ 0.003x + 0.09 & \text{if } 290 \leq x < 295 \\ 0.005x - 0.50 & \text{if } 295 \leq x < 300 \\ 1 & \text{if } x \geq 300 \end{cases} \quad (6.22)$$

Hence, the empirical uncertainty distribution for the integrated expert data on sesame yield was defined by (6.22) and had expected value of 264.35 kg per hectare and variance 163.28 (ton/ha)

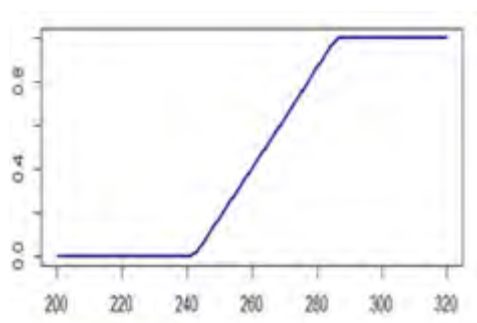
### 6.5 Fitted Uncertainty Distributions for Sesame Seed Yield

The Delphi procedure is implemented on the multiple experts' data to obtain a composite data of sesame seed yield. The resulting aggregated data was fitted into uncertainty models. In order to determine the parameter estimates, since the graphs of uncertainty normal distributions and their inverses have asymptotes at zero and one, that is  $\alpha \in (0,1)$  is open, the lower and upper limit values were adjusted by replacing  $\alpha = 0$  by 0.01 and  $\alpha = 1$  by 0.99. The result is presented in Table 6.10.

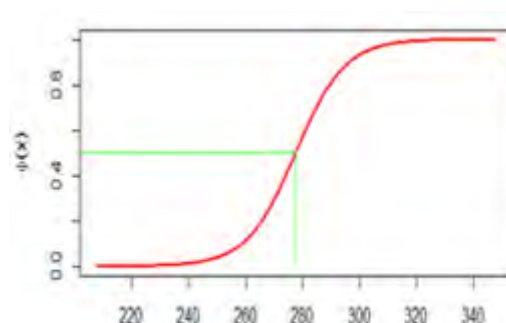
**Table 6.10:** Estimates of Uncertainty Linear and Uncertainty Normal Distribution on Sesame

	Data point	Observed Mean	Expected value	Spread	Functional form
Normal model	8	280	277.500	15.556	N(262.39, 15.56)
Linear model	8	280	264.375	12.778	L(242.24, 286.51)

The fitted uncertainty linear and normal distribution graphs are displayed in the Figures 6.18 and 19 below:



**Figure 6.18:** Fitted Uncertainty Linear Distribution for Sesame seed yield



**Figure 6.9:** Fitted Uncertainty Normal Distribution for Sesame seed yield

## 6.7 Distance Estimates between Expert Distributions and Uncertainty Normal Distribution

In this section, the distance measure derived in chapter 5 is applied to compute the distance between two independent distributions. The distance measure is used to check the closeness of the empirical model and the realization of data (expert distribution). The uncertainty distance is a non-negative and symmetry measure. The distance estimates between a normal distribution and the expert elicited distributions was computed for the crop yield data. The MATLAB tool kit online developed by Liu (2011) was used to compute the distance between the expert distribution and a standard uncertainty normal distribution  $N(0,1)$ . The distance estimates are presented in Table 6. 11.

**Table 6.11:** Distance between Standard Uncertainty Normal Distribution and estimated Distributions for the Expert Crop data

Crops	Standard Uncertainty Distribution	Estimated Uncertainty Distribution	Distance value	Rank
Maize	N(0,1)	L(1.611, 2.694)	1.2109	a
Rice	N(0,1)	L(0.658, 1.611)	1.3869	a
Sorghum	N(0,1)	L(0.704, 1.307)	1.2521	b
Soybean	N(0,1)	L(0.858, 2.067)	1.6453	b
Cowpea	N(0,1)	L(0.201, 1.841)	1.4381	b
Maize	N(0,1)	N(2.128, 0.246)	2.1887	b
Rice	N(0,1)	N(1.294, 0.174)	1.4584	b
Sorghum	N(0,1)	N(0.971, 0.206)	1.2486	a
Soybean	N(0,1)	N(1.410, 0.277)	1.5883	a
Cowpea	N(0,1)	N(1.191, 0.198)	1.3925	a

Table 6.11 presents the distance estimates to assess the divergence of uncertainty models of the expert data for cereal crop yields from the standard uncertainty normal distribution. For each crop, we have fitted two uncertainty distributions for the crop yield as provided by the experts (uncertainty linear and normal function). The two estimated models are then compared with the standard uncertainty normal distribution.

For maize and rice yields, the estimated uncertainty linear distribution is better than the corresponding uncertainty normal distribution, as it is indicated in the rank column as “a” for linear distribution means a smaller distance than the estimated uncertainty normal distributions.

For sorghum, soybean and cowpea, it can be seen that the uncertainty normal distribution had a smaller distance than the corresponding estimated uncertainty linear model. The rank “a” for the estimated normal function indicates would be preferred distributions than their corresponding linear distributions for sorghum, soybean and cowpea.

## **6.8 Summary and Discussion**

In this chapter, uncertainty statistics was extensively employed to analyse the two datasets provided by the agricultural experts. The first dataset, which consisted of five cereal crop yields, was subjected to uncertainty analytical procedures. Linear interpolation was first performed on the five cereal crops to obtain the empirical distributions, the expected value and variances. The uncertainty distributions were then fitted on the crop yields. Lastly, the parameters of a deterministic function for the maize yield on the NPK fertilizer are estimated using the uncertainty procedure. We later analyzed the sesame seed yield obtained from multiple experts by employing the Delphi method, which resulted in a composite aggregated data. An empirical distribution was fitted to the aggregate data. We then fitted uncertainty linear and normal function to the composite data. Further, the estimated expert distributions were compared with the standard uncertainty normal distribution using the distance measure.

The model comparison here is to assess the divergence of the two models from the standard uncertainty distribution. The proposed uncertainty distance measure was computed to compare the closeness of the two elicited experts’ distributions (linear and normal distributions) and the standard uncertainty normal distribution. The distance comparison between distributions is as a discrepancy measure similar to the Kullback-Leibler (KL) divergence measure in probability theory. This provided a simple and cost saving approach where sampling data is not available or expensive to acquire.

The use of expert information is recently growing as demonstrated by an extensive survey of the literature on the use of expert opinion in scientific inquiry and policy making. This work has further emphasized the usefulness of expert’s opinion in projection of crop yields. The use of expert knowledge has been successful in several area agricultural oriented researches; e.g. crop insurance (Sherrick, B.J. 2002), agricultural ecology (Martin et al. 2005), birds survey (Kuhnert et al.(2005) and safety in dairy production (Valeeva, et al. (2005)

## 7. Results of Linear Regression Analysis on Cereal Crop Yields

This chapter presents the results of the OLS and Bayesian analysis of the linear regression model fitted to the agricultural household surveys. The “full” data is first analysed using *lm* command for ordinary least squares estimation and the **MCMCregress** command under **MCMCpack** package for direct Bayesian estimation, *R2WinBUGS* package (with uniform prior) all in *R* environment.

In the second part, we analyse the split data by implementing the procedure discussed in Chapter 4, Section 4.5. The whole data set was split into 2 sets: the old data and the new data. The prior was estimated from the old dataset and OLS estimates are then incorporated into the Bayesian analysis. Three models were generated: (GLM: M1), Bayesian model with non-informative prior (M2) and Bayesian model with informative prior (M3). The predictive performance of the models is compared using the leave-one-out cross validation method.

### 7.1 Descriptive Statistics of Agricultural survey data

In this section, summary statistics are presented for the agricultural survey data in Table 7.1. The second column, N is the number of respondent farmers; column 3 is the observed minimum and maximum values for the respective variables, followed by the mean and standard deviations. Table 7.1 also presents the summary statistics for the categorical variables: soil supplements and educational levels on the crop yields. The continuous variables are actual measures of the variables listed in rows 1 to 4. The raw data is presented in Appendices E.

For sorghum, the number of respondent farmers was 135; it is interesting to note that the majority of the respondents are in their forties and probably with no formal education. Farming is the primary occupation of the sample population as it can be seen that a respondent had the highest farming experience of 57 years and the average years in farming operation of 24 years. It is possible that a smaller percentage of the farmers had less than 5 years in farming activities, probably after retirement from civil service. This group of farmers cultivated less than 5 hectares of land; this shows that the majority were subsistent farmers. Large percentages of the farming population had access to extension service, while the reverse is the case for accessibility to credit facilities such as government loans and

subsidies. This might be due to lack of information available to this group of farmers because with low level of education and awareness.

The level of agricultural productivity of the sampled farmers was quite low compared to their counterparts in other developed economies. The low yield can be attributed to reduction in soil fertility and the depletion in top soil. The descriptive statistics of the survey data is presented in Table 7.1 on sorghum crop yields.

**Table 7.1: Descriptive Statistics of variables on Sorghum yield**

<b>Variables</b>	<b>N</b>	<b>Min</b>	<b>Max</b>	<b>mean</b>	<b>SD</b>
Farm size (ha)	135	0.30	5	2.64	1.120
Quantity of NPK fertilizer(kg/ha)	135	0.00	1000	100.1	133.5
Quantity of seed sow (kg/ha)	135	10.0	80	21.99	8.710
Farming experience (years)	135	0.00	57	23.96	12.86
<b>Soil supplements category on crop yields</b>					
No soil supplement	20.0	434.8	7000	2415	428.3
Organic supplement	3.00	400.0	2000	1006.7	471.2
Inorganic supplement	82.0	208.3	12000	1934.4	192.3
Integrated soil supplement	30.0	250.0	16200	2222.4	524.4
<b>Education Levels category on crop yields</b>					
No formal Education	79.0	208.30	12000	2232.7	217.3
Primary Education	15.0	48.00	1300	3820.4	61.20
Secondary Certificate	18.0	73.30	1144	3166.4	72.40
Degree/ HND Diploma	21.0	12.00	2200	3726.2	85.29
Postgraduate /Professional Diploma	2.00	132.0	2588.4	1058.8	574.2
<b>Extension service on crop yields</b>					
Access to Extension service	130	250.0	16200	2063.5	108.9
No access to Ext. service	5	208.3	5000	1731.7	864.9
<b>Credit facility on crop yields</b>					
Access to credit facility	13	434.7	6307.7	1878.2	467.8
No access to credit facility	122	208.3	16200	2069.6	180.9

Min.= minimum, Max.= maximum, SD= Standard deviation, N= number of farmers. The values provided in the table are in *kg/ha*.

This summary presented is the same across the farming population when cultivating other cereal crops as can be seen in Tables 7.2 for millet.

**Table 7.2:** Descriptive Statistics of the predictors of model variables on Millet yields

Variables	N	Min	Max	Mean	SD
Farm size (ha)	95.0	0.60	5.00	2.33	0.81
Quantity of NPK fertilizer(kg/ha)	95.0	0.00	500	101	102.1
Quantity of seed (kg/ha)	95.0	6.00	76.0	26.1	11.1
Farming experience(years)	95.0	0.00	70.0	25.9	12.5
<b>Soil supplements category on crop yields</b>					
No soil supplement	14.0	384.2	4166.7	1627.8	294.8
Organic manure supplement	3.00	1500	3466.7	2374.9	578.0
Inorganic supplement	63.0	210.0	3000	940.10	71.90
Integrated soil supplement	15.0	333.3	2000	1135.8	146.9
<b>Education Levels category on crop yields</b>					
No formal education	40.0	210	4166.7	1142.6	121.6
Primary education	18.0	333.3	2500	1113.5	180.8
Secondary Education	13.0	600	1700	1085.4	102.4
Degree/NCE/HND Diploma	20.0	250	3467	1169.5	187.5
Postgraduate/Professional	4.00	764.7	3000	1581.2	498.3
<b>Extension service on crop yields</b>					
Access to Extension Services	88	210	4166.7	1171.6	81.46
No access to Extension Services	7	500	1500	924.5	118.0
<b>Credit facility on crop yields</b>					
Access to credit facility	17	210	4166.7	1184.6	82.90
No credit facility	78	48.9	2200	510.7	141.1

Min.= minimum, Max.= maximum, SD= Standard deviation, N= number of farmers. The values provided in the table are in *kg/ha*.

Tables 7.1 and 7.2 present summary statistics for two groups of farmers (sorghum and millet). The results on millet showed a substantial reduction in yield under inorganic farming relative to organic practice. The demographics of this group of farmers in this study show that they are poor, low-resourced and subsistent farmers. One reason for reduction in crop yields under inorganic soil supplement can be attributed to, perhaps, they did not apply enough dose of NPK fertilizer or they did not complete the sequence of fertilizer applications at different

stages of crop growth development. Due to the financial insolvency of these farmers, perhaps they only applied fertilizer at the crop seedling stage to boost the crop vigour, but were unable to apply at a stage in yield component formation. Another reason that can be adduced to yield reduction under inorganic supplement may be the inappropriate timing of fertilizer application.

## 7.2 Model Specification

The model specification for the data was the general linear model. This is a multiple linear regression model which allows the user to analyse the effects of explanatory variables that involve continuous covariates and categorical variables. The continuous variables are: farm size, quantity of NPK fertilizer (kg/ha), quantity of seed sown (known as crop density in agronomy) (kg/ha), farmer experience (years) while categorical factors are: soil (4 levels of supplements), Education (5 levels) and binary variables were extension availability (yes/no) and credit facility (yes/no).

The model was first fitted to full data using Maximum likelihood Estimation (MLE) and direct MCMC approach. The data analyzed consisted of 13 predictors with crop yields (kg/ha) as the target variable. The raw data is tabulated in Appendix E: farm size effect  $x_{i,1}$ , NPK fertilizer effect  $x_{i,2}$ , the Quantity of seed sow effect  $x_{i,3}$ , and farmers' experience effect  $x_{i,11}$ . The definitions of other binary variables are as follows:

Soil:

$$x_{i,4} = \begin{cases} 1 & (\text{Organic manure soil supplement}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$x_{i,5} = \begin{cases} 1 & (\text{Inorganic manure supplement (synthetic fertilizers)}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$x_{i,6} = \begin{cases} 1 & (\text{Integrated soil sup.}) \\ 0 & (\text{Otherwise}) \end{cases}$$

Education:

$$x_{i,7} = \begin{cases} 1 & (\text{primary school (spent between 2–7 years)}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$x_{i,8} = \begin{cases} 1 & (\text{Secondary school (spent 8–14 years)}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$x_{i,9} = \begin{cases} 1 & (\text{Degree honour / NCE / HND diploma (spent 14–18 years)}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$x_{i,10} = \begin{cases} 1 & (\text{Postgraduate / professional diploma (spent } \geq 18 \text{ years)}) \\ 0 & (\text{Otherwise}) \end{cases}$$

$$\text{Extension Services: } x_{i,12} = \begin{cases} 1 & (\text{Access to Extension services}) \\ 0 & (\text{No Access}) \end{cases}$$

$$\text{Credit facility: } x_{i,13} = \begin{cases} 1 & (\text{Access to Credit facility}) \\ 0 & (\text{Non Access}) \end{cases}$$

The general linear regression model is

$$y_i = \beta_0 + \beta_1 * x_{i1} + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \dots + \beta_p * x_{ip} + \varepsilon \quad (7.1)$$

In the formula above, we consider  $n$  observations (respondents) of one dependent variable (crop yields) and  $p$  independent variables. Thus,  $Y_i$  is the  $i^{\text{th}}$  observation of the dependent variable,  $x_{i,j}$  is the  $i^{\text{th}}$  observation of the  $j^{\text{th}}$  explanatory variable,  $j = 1, 2, \dots, 13$ . The values  $\beta_j$  represents parameter to be estimated,  $\beta_j$  and is the  $i^{\text{th}}$  independent identically distributed normal error.

It is worthwhile to note that  $\beta_0$  represents the intercept. That is, no soil supplements, no formal education, no access to extension services and no access to credit facility. More importantly, due to the mixture of continuous and categorical variables as predictors, the interpretation of the effects of predictor variables on the response variable will be different. The continuous variables were  $(X_1, X_2, X_3, X_{11})$  denoting farm size, NPK, seed and farmer's year of farming; these would have unit effect on the response. The categorical variables were soil supplements  $((X_4, X_5, X_6))$  and education  $(X_7, X_8, X_9, X_{10})$  would be interpreted as differential effects in levels with respect to the reference intercept, which usually assumes the value of the specific group intercept.

### 7.3 Results Regression Analysis on full Data

The results of the analysis (R code in Appendix C) of the regression model in equation (7.1) are presented in Tables 7.4 and 7.5 for sorghum and millet respectively. The summary estimates were computed from the maximum likelihood method and direct MCMC posterior estimates. A summary of posterior estimates from **WinBUGS** analysis is also presented and

corresponding density plots of the posterior estimates are displayed for each crop immediately after their tables.

### 7.3.1 Regression Analysis for Full Data on Sorghum Yield

The results presented in Table 7.3 utilize the statistical inferences described in Chapters 3 and 4. The mean estimates of ordinary least squares and direct MCMC estimation results are close estimated values and intervals. However, their interpretations are completely different, for example, MLE 95% confidence interval in classical statistical analysis means that for a repeated sampling, the 95% confidence interval would include the true value all the time. Bayesian credible intervals on the other hand are directly interpreted as the probability distributions of the parameters.

Generally, the assessments of the outputs in terms of mean estimates were similar, but their confidence intervals differed significantly. The direct **MCMC** estimates under a vague prior, thus the credible intervals from **MCMC** analysis are very close OLS values. The output of WinBUGS utilized a uniform prior resulted into highest posterior density (HPD) intervals, where the sets would likely contain (intervals) of all values of the parameter such that the posterior density is larger than some constant  $C_\alpha$ . The interval estimates include the coverage probability will be  $(1-\alpha)\%$ . Hence, HPD credible resulted into higher posterior mean and generate narrower intervals than the corresponding OLS and direct MCMC intervals. Thus, Bayesian approach chooses the shortest possible interval enclosing  $(1-\alpha)\%$  of the posterior mass. The variability is therefore reduced in **R2WinBUGS** simulated due to 30,000 simulations being performed to attain convergence

Occasionally credible sets can have strange ‘shapes’. In this research, it was found that the posterior distribution of the categorical predictors (e.g. Soil supplements, education levels) in the model could often be multimodal, and that the HPD credible intervals would include all the estimates in the form  $[\theta_{i_1}, \theta_{u_1}] \cup [\theta_{i_2}, \theta_{u_2}] \cup [\theta_{i_3}, \theta_{u_3}]$ .

The multidimensional HPD credible sets can be defined in the same way as in the univariate sets. With MCMC output, equal-tailed intervals are very easy to approximate. However, approximating HPD intervals from MCMC output can be extremely difficult, because it requires several simulations for convergence to approximations (e.g. the density or the solution to be attained for a specific model). It can be observed from **WinBUGS** column in

Table 7.3. The posterior intervals included no zero for all predictor coefficients; this absence indicates that the variables are significant at 5% probability level.

However, the MLE analysis indicated that the confidence intervals included zero, indicating that those variables are not significant at the 5% probability level. The MLE intervals included zero except for the intercept ( $\beta_0$ ) and farm size parameter ( $\beta_1$ ).

**Table 7.3:** Summary of estimates of MLE and Posterior of Bayesian Estimation on Sorghum

Parameters	MLE (CI 95%)	MCMC	R2WinBUGS
Intercept	3470.5 (1227, 5671)	3473 (1281.0, 5689.0)	3192 (3180, 3204)
Farm Size ( $x_1$ )	-754.73 (-1047, -462.3)	-753.1 (-1044, -464.4)	-776.0 (-778, -774)
Fert. ( $x_2$ )	1.402 (-1.350, 4.150)	1.403 (- 1.349, 4.105)	1.113 (1.098, 1.127)
Seed ( $x_3$ )	36.01 ( -2.250, 74.30)	36.09 ( - 2.100,74.34)	37.64 (37.42, 37.86)
SOIL1 ( $x_4$ )	-840.89 (-3032, 1350)	-831.8 (- 3024,1395.0)	-567.3 (-580, -555)
SOIL2 ( $x_5$ )	-153.31 (-1104, 797.6)	-160.2 (- 1105 , 797.2)	175.2 (170.7, 179.7)
SOIL3 ( $x_6$ )	191.53 (-919.5, 1303)	183.2 (- 923.3,1318.0)	449.2 (443.8, 454.6)
EDU1 ( $x_7$ )	-11.25 (-1026, 1004)	-18.25 (-1053 ,1013)	113.9 (108.0, 120.0)
EDU2 ( $x_8$ )	-418.78 (-1334, 496.1)	-415.8 (-1338 ,502.7)	-407.0 (-412, -402)
EDU3 ( $x_9$ )	-764.58 (-1640, 110.4)	-764.0 (-1647, 123.4)	-755.0 (-760,-751)
EDU4 ( $x_{10}$ )	932.17 ( -1603, 3467)	936.6 (-1547, 3458)	811.7 (797.5 , 825.8)
Exper. ( $x_{11}$ )	-0.5878 (-24.78, 23.60)	-0.632 (-24.79 , 23.80)	-0.394 (-0.532, -0.257)
Ext. ( $x_{12}$ )	-137.85 (-1780, 1504)	-141.2 (-1806 , 1516)	-81.04 (-90.19, -71.89)
Credit ( $x_{13}$ )	172.82 (858.0, 1231)	173.3 (-1806 , 502.7)	180.4 (174.4, 186.4)
Sigma	1733	3.061*10E-6	10

(95% CI in the parenthesis)

\*Note:  $\beta_0$  is the intercept and it assumes the reference value of no soil supplement, no formal education, no extension contact and no access to credit.

### **Interpretations**

Generally, it can be seen from Table 7.3 that the results showed negative prediction for increased farm size. An increased farm size may not necessarily lead to higher yields. An important economic argument in favour of equitable distribution of farm land is that smaller farms are more efficient and more productive in the management of the resources. Masterson and Rao (1999) asserted that small farms have both higher land productivity and equal or better technical efficiency.

The continuous predictors ( $X_1, X_2, X_3, X_{11}$ ) can be interpreted simply, as a unit change in any of these variables would result into relative coefficient effect on the yield response. The estimates ( $\beta_2$ ) and ( $\beta_3$ ) are respectively coefficients of NPK fertilizer and seed quantity which had positive effect on sorghum yield output. This fact means that a unit increase of each of these variables keeping other variables unchanged, would yield a corresponding coefficient increase on the sorghum yield. However, a unit change in either farm size (ha),  $\beta_3$  or farmer experience (years),  $\beta_{11}$ , would yield a negative impact on sorghum output.

The categorical variable effects can be described in term of differential effects with respect to the specific group intercept (reference level 0). The parameters for the categorical variables, say, soil supplements with dummy effects  $\beta_4 = -840.89$ ,  $\beta_5 = -153.31$  and  $\beta_6 = 191.53$ . These values are interpreted as differential effects with respect to no additional soil supplement. The  $\beta_4$  is the parameter of the use of organic manure supplement on sorghum yield, meaning a 840.69 decrease in sorghum yield compared to group reference no additional soil supplement on the field. While a farmer, who adopted integrated soil supplements had 191.53 kg/ha predicted increase from his field in relation to no additional soil supplements. The detailed interpretation of the coefficient estimate of each level effect is given below.

For a categorical variable, say, soil supplements, the differential effects of the level coefficients are described with respect to the reference level of no soil supplement, which assumes the intercept value  $\beta_0 = 3470.5$ . For a factor effect model, we illustrate two cases below.

**Case 1:** one categorical variable, soil supplements

For a farmer who did not add any soil supplement to his field, the expected yield (or likelihood estimates) would be equal to  $\beta_0 = 3470.5$ , and the following coefficients for other levels can be computed

Organic supplement,  $E(Y_4) = \beta_0 - 840.89(1) = 2529.61$ . This value reflects that the mean yield of this level would be less than the reference yield by 840.89.

Inorganic supplement,  $E(Y_5) = \beta_0 - 153.31(1) = 3216.19$ . This value indicates that the mean yield of this level would be less than the reference yield by 153.3

Integrated soil supplement,  $E(Y_6) = \beta_0 + 191.53(1) = 3662.03$ . This value implies that the mean yield at that level would produce 191.53 *kg/ha* predicted increase than the reference intercept.

Hence, the submodel of sorghum yield on the soil supplements can be given as

**Case 2:** Two categorical variables, say, soil,  $X = (0, 1, 2, 3)$  and Education,  $Z = (0, 1, 2, 3, 4)$

We construct dummy variables  $(x_1, x_2, x_3)$  for category values  $X = (1, 2, 3)$  and similarly four dummies  $(z_1, z_2, z_3, z_4)$  for  $Z = (1, 2, 3, 4)$  and consider the regression model:

$$y = \beta_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 z_1 + \alpha_5 z_2 + \alpha_6 z_3 + \alpha_7 z_4 + \varepsilon$$

To interpret the meaning of these coefficients, we consider the following cases:

$X=0$

- (1)  $(X = 0, Z = 0)$ ,  $y = \beta_0 + \varepsilon$
- (2)  $(X = 0, Z = 1)$ ,  $y = \beta_0 + \alpha_4 + \varepsilon$
- (3)  $(X = 0, Z = 2)$ ,  $y = \beta_0 + \alpha_5 + \varepsilon$

$X=1$

- (1)  $(X = 1, Z = 0)$ ,  $y = \beta_0 + \alpha_1 + \varepsilon$
- (2)  $(X = 1, Z = 1)$ ,  $y = \beta_0 + \alpha_1 + \alpha_4 + \varepsilon$
- (3)  $(X = 1, Z = 2)$ ,  $y = \beta_0 + \alpha_1 + \alpha_5 + \varepsilon$  and so on.

To interpret the results in Table 7.3, in terms of significance test at 5% probability, we refer to columns MLE and MCMC confidence intervals which agree in their results as against **WinBUGS** outputs, which generally indicate that all the predictors are significant (i.e. credible intervals all excludes zero).

The farm size had a negative effect ( $\beta_0$ ) on the sorghum yield which was significant at 5% probability. This value indicates that a unit increase in farm size would lead to 754.73 decreases on the response output. Likewise, farmer experience (in years) had a negative coefficient on the response output, but this was not significant at 5% probability. Both NPK fertilizer and seed quantity sown per hectare had positive contribution to sorghum yield, but the effect was not significantly different from zero.

Soil supplement levels did not indicate any significant difference from the reference level ( i.e. no soil supplement). Although organic and inorganic factor effects gave a relative decrease or less than reference group (no supplement), while an integrated soil supplement had 172.82 kg/ha predicted increase than the reference group value.

The educational levels did not indicate any differential difference between their levels and the reference group (no formal education) at 5% probability, as all confidence intervals include zeros. The predictor variables( x7, x8, x9 ) had negative coefficients, indicating relative decrease on the yield output for a farmer with higher level of education than the no formal education group , while binary variable (x10) had a positive coefficient indicating a relative higher sorghum yield than the non educated group (reference intercept value). Perhaps, this group of farmers who had higher education were retired civil servants with few years of farming experience but who had enough money to hire experienced farmers as labour to manage their farms and possibly had large farms.

The extension service contact had a negative coefficient which not significantly different from those farmers, who did not have any contact with an extension officer. The effect coded (1) would produce 137.85 less than the reference level (no contact to extension worker).

Access to a credit facility had a positive coefficient which not significantly different from the reference level (no access). This value indicates that an access to a loan/ credit facility would impact positively on productivity level compared with farmers without credit.

The density plot of the posterior estimates is displayed in Figure 7.1

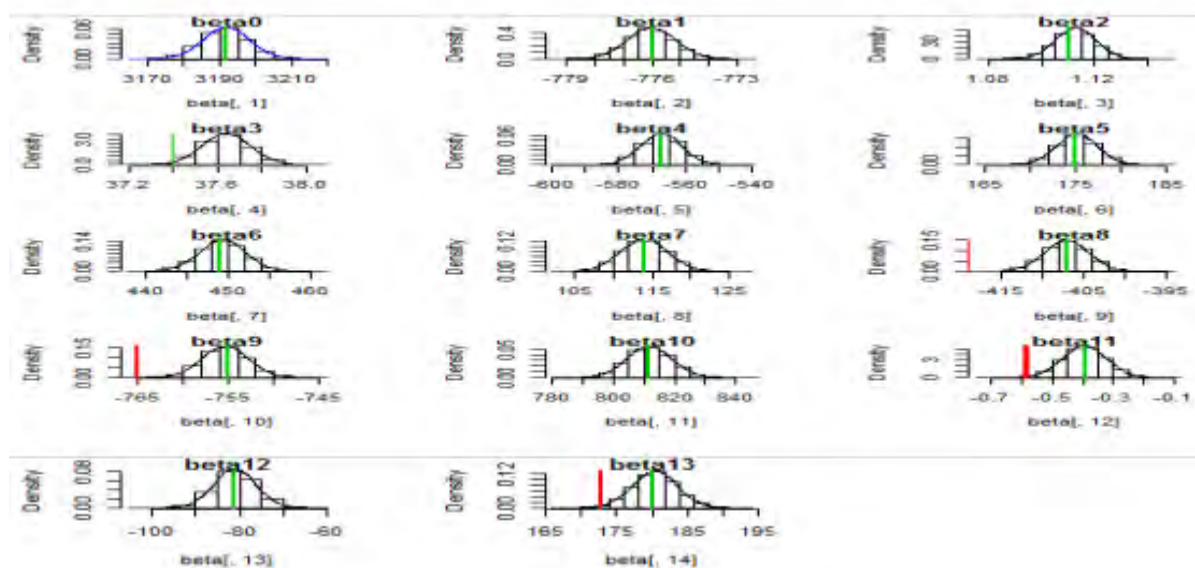


Figure 7.1: Density Plots of Bayesian posterior estimates indicated by green bar and MLE coefficient estimates by red bar.

From Figure 7.1, it can be observed that the mean estimates of MLE were lesser than the Bayesian posterior means called maximum *à priori*. The density plots seem to be normal distributions, indicating that the simulation converged. The trace plots are provided in Appendix C.

### 7.3.2 Regression Analysis on Millet Yield for Full Data

Table 7.5 presents the results of regression analysis on millet yields. The corresponding density plots of the estimates on millet yield *kg/ha* is presented in Figure 7.2. The graphs of Trace plots and density distributions are presented in Appendix C.4.

**Table 7.4:** Summary of estimates of MLE and Posterior of Bayesian Estimation on Millet

<i>Variables</i>	<i>MLE (CI 95%)</i>	<i>MCMC</i>	<i>R2WinBUGS</i>
Intercept	2204 (1239, 3169)	2206 (1241, 3183)	2113.8 (2109.0, 2128.0)
Farm Size ( $x_1$ )	-307.8 (-512.6, -103.2)	-307 (-512, -103)	-292.7 (-295.6, -289.6)
Fert. ( $x_2$ )	0.660 (-0.954, 2.275)	0.660 (-95.8, 2.251)	0.6500 (0.6300, 0.6800)
Seed ( $x_3$ )	-13.80 (-27.78, 0.1763)	-13.77 (-27.59, 0.212)	-12.76 (-12.97, -12.55)
SOIL1 ( $x_4$ )	885.9 (36.81, 1735)	889.0 (39.34, 1748)	869.60 (857.10, 882.0)
SOIL2 ( $x_5$ )	-545.4 (-1008, -82.94)	-548.7 (-1007, -80.44)	-540.5 (-547.3, -533.6)
SOIL3 ( $x_6$ )	-625.5	-629.1	-616.1

	(-1187, -63.73)	(-1186, -59.33)	(-624.5, -607.8)
EDU1 ( $x_7$ )	162.6 (-254.5, 579.6)	159.3 (-266.0, 586.2)	161.90 (155.80, 168.00)
EDU2 ( $x_8$ )	-92.97 (-553.3, 367.4)	-94.4 (-580.0, 371.0)	-117.6 (-124.4, -110.7)
EDU3 ( $x_9$ )	168.5 (-233.6, 570.6)	166.1 (-233, 569.4)	168.20 (162.20, 174.10)
EDU4 ( $x_{10}$ )	650.6 (-51.93, 1353)	648.6 (-43.5, 1367)	628.90 (618.60, 639.40)
Exper. ( $x_{11}$ )	11.31 (-2.191, 24.80)	11.24 (-2.341, 24.87)	11.150 (10.960, 11.350)
Ext. ( $x_{12}$ )	93.62 (-449.1, 636.3)	93.79 (-457, 642.1)	123.50 (115.5, 131.6)
Credit ( $x_{13}$ )	-547 (-1064, -29.53)	-544 (-1070, -30.16)	-534.5 (-542.1, -526.9)

(95% CI in parenthesis).

\*Note:  $\beta_0$  is the intercept that takes the reference value of no soil supplement, no formal education, no extension contact and no access to credit.

### *Interpretations*

Table 7.4 presents the results of the linear regression model fitted by the three analytical procedures (MLE, **MCMC**, and **WinBUGS**). The **WinBUGS** results in column (4) used a uniform prior. The estimated intervals do not included zero. This indicates that the predictor's effects are significantly different, although most of the coefficients were negative. Some predictors had positive coefficients on the millet yield output, i.e. fertilizer effect ( $\beta_2$ ), organic farm manure effect ( $\beta_4$ ) and credit facility effect ( $\beta_{11}$ ) although they are not significant at the 5 % probability level.

Farm size indicated significant but negative effect on the millet yield. It indicates that a unit increase in cultivated farm size, keeping all other predictors unchanged would result in a 307.8kg/ha predicted decrease in the yield of millet.

The parameter for NPK fertilizer was not significant at the 5% level of significance, A unit increase in the quantity of NPK fertilizer applied, other variables been held fixed, would lead to a 0.660 kg/ha predicted increase in the yield of millet.

The farmers experience was negative and insignificant at the 5% probability level. A unit increase in the years of farmers experience, other variables been held fixed, would lead to a 11.31kg/ha predicted increase in millet yield.

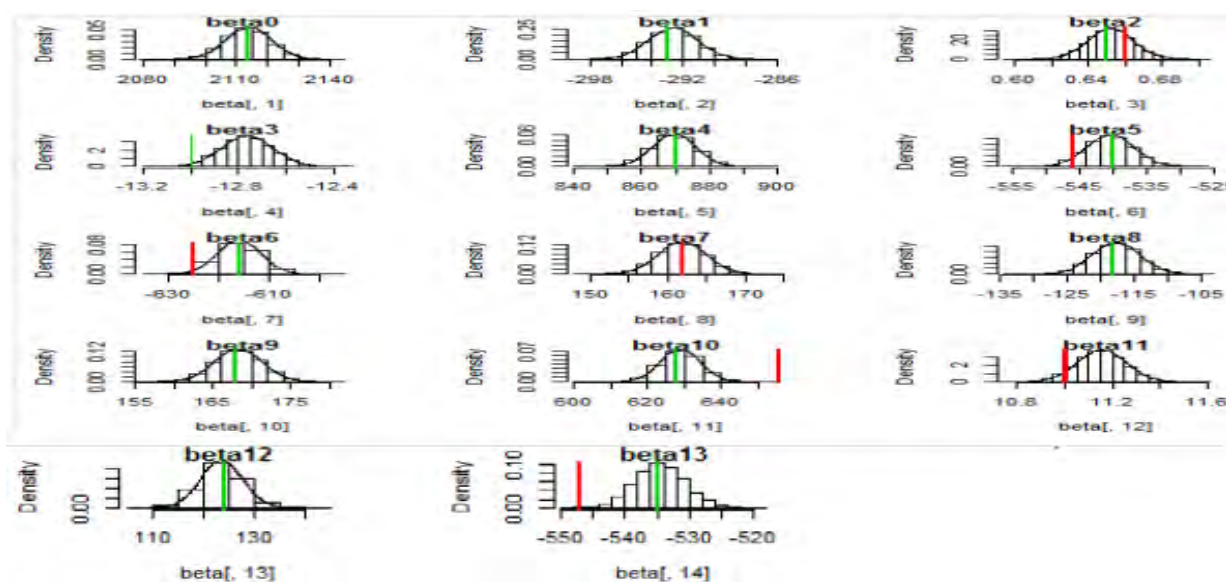
The soil supplements had no significant difference from the reference value, except integrated soil supplement. All soil supplement regimes showed relative decrease (negative

coefficient) on the millet output when compared to the reference supplement except the organic supplement, which gave 885.9 *kg/ha* predicted increase than the group reference.

The effects of higher educational levels were negative when compared to the no formal education reference, and the result would lead to a reduction in yield compared to the group reference value.

The extension service contact had a negative coefficient, while access to credit facility had a positive coefficient. Although, these effects were not significantly different from their reference level values.

The density plot of the posterior distribution is displayed in Figure 7.2.



**Figure 7.2:** Density Plots of Bayesian Posterior mean indicated by green bar and MLE coefficient estimates by red bar for Millet yield.

In conclusion, Tables 7.3 and 7.4 results show that the model had predicted the yield negatively. This shortcoming can be addressed by adopting constrained regression where the negative coefficients of the model could be set to zero. Such model parameters could be constrained to only positive values. In other hand, it may be appropriate that the response variable of interest (crop yield) could assume some probability distributions, where the parameters are restricted to positive values. Distributions such as the Truncated Normal, log-normal, gamma and Weibull can be used as alternative solutions to redeem the negative prediction of the yield. The commonly used probability distributions for crop yield density have been discussed in many empirical studies (Ker and Goodwin, 2000; Nelson and Preckel (2001; Sherrick, et al.2004).

## 7.4 Bayesian Prediction Analyses

In this section, we summarize the results (R code Appendix C.3) for the predicted and expected yields for a given set of covariates using equation(4.20) on the full data set. The model (4.20) was implemented in *R* environment using *LearnBayes* given a set of covariates. The results of a simulated sample of 5000 draws from the joint posterior are presented in Table 7.5.

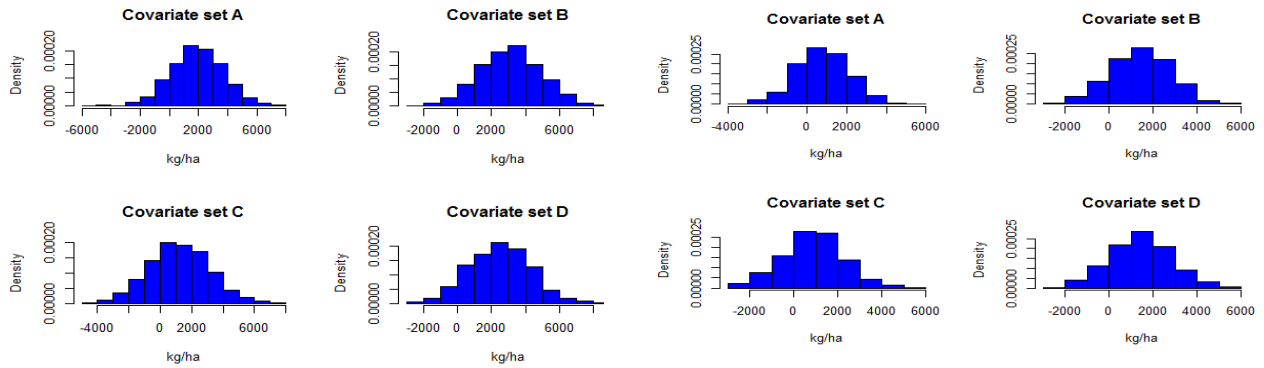
The table presents the predicted crop yields (*kg/ha*) and credible regions for sorghum and millet. It can be seen that most of the predicted regions contained zero, indicating no significant effects of the predictor. Therefore, the predicted results are similar to the fitted model in the previous tables. The predicted values are close to the expected and the credible regions almost overlapped, particularly for millet.

The output generated the corresponding expected values for a given set of covariates. A simulated sample of 5000 draws from the joint posterior was performed and the results presented in Table 7.5. The corresponding histograms of posterior densities of these predicted means for the crops: sorghum and millet are presented in Figures 7.3.and 7.4.

**Table 7.5** Predicted and expected values and 90% credible regions of crop yields (*kg/ha*) for sorghum and millet for sets of covariates

Covariates	Sorghum		Millet	
	Predicted	Expected	Predicted	Expected
<i>A</i> *	1911.1 (-1006.4, 4886.9)	1944. (1476.6, 2375.3)	1019.2 (821.3, 1256.4)	1035.2 (821.3, 1256.4)
<i>B</i> *	3125.2 (181.9, 6095.5)	3119.4 (2176.1, 4054.5)	863.0 (471.8, 3132.3)	890.62 (471.8, 1312.3)
<i>C</i> *	1201.3 (-1922.4, 4332.6)	1156.2 (-363.5, 2744.5)	1046.9 (555.5, 1529.8)	1060.3 (555.5, 1529.8)
<i>D</i> *	2485.9 (-352.4, 5356.3)	2485.4 (1498.6, 3522.1)	267.7 (-370.7, 899.3)	267.24 (-370.7, 899.4)

(Credible intervals indicated in the *parentheses*).



**Figure 7.3:** Histogram of simulated draws of the predictive distribution for Sorghum yield for sets of covariate values.

**Figure 7.4:** Histogram of simulated draws of the predictive distribution for millet yield for sets of covariate values.

## 7.5 Results of Regression Analysis on Split Data

The data used for the analysis is described in Chapter 2 and the detailed survey design. We applied the procedure described in Chapter 4, Section 4.5. The predictive performance of the models was compared through the Leave one- out cross validation method.

The dataset consists of 135 small-scale farmers (sorghum) and 95 farmers (millet) with no missing value. The target (dependent) variable of interest was a continuous variable indicating the quantity of crop yield per hectare from the individual farm. The data set is also consisted of 13 input (independent) variables. The variable types are summarized in Table 7.6.

The analytical procedure on split data discussed in section 4.4 and leave one-out cross validation (LOOCV) was used to compare the predictive performance of the 3 models. The Gaussian prior was estimated from the old data and incorporated as expert prior in the Bayesian model (Model 3). Assuming a non- informative prior, diffuse prior, flat prior in the Bayesian model (M2) and GLM (M1) all fitted on the new data.

**Model 1:** Linear regression model (*MLE*)

**Model 2:** Non- informative prior  $p(\beta) \propto \text{constant}$  (Derivation in Appendix C.4)

**Model 3:** Informative Prior  $p(\beta | u, \sigma^2) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\beta - u)^2}{2\sigma^2}\right)$

(Derivation of posterior in Appendix C.4)

**Table 7.6** Variable type and description for each variable in the data set.

Variable	Model Role	Variables Type	Description
yield	Target	Numerical - Continuous	harvest crop yield in kilogram per ha
SIZE	input	Numerical - Continuous	hectare of land cultivated
NPK	input	Numerical - Continuous	Quantity of NPK fertilizer applied per ha
SEED	input	Numerical - Continuous	Quantity of seed sow kg per ha
SOIL	input	Categorical-Nominal	Four categories of Soil supplements
EDU	input	Categorical-Nominal	Five categories of Educational level
EXP	input	Numerical –discrete/con	Number of years in farming activities
EXT	input	Binary-Nominal	Contact to Extension worker during planting
CREDIT	input	Binary-Nominal	Access to credit facility

The summary statistics has been presented earlier in Table 7.1 and 7.2. The binary variable consists of zeros and ones. In both data sets, there appeared to be outliers towards the right tail. The outliers are later treated when the models had been fitted.

Initially, the data set was randomly spit into 2 sets (50: 50). For sorghum,

The old data set contained 68 observations out of 135 sorghum observations

The new data contained 68 observations

The validation test comprised of the whole data which was used for a validation test.

Three models were generated and their predictive performance compared using LOOCV

**Model 1:** Linear regression model (*GLM*)

**Model 2:** Non- informative prior  $p(\beta) \propto \text{constant}$  (see the derivation in Appendix A.1)

**Model 3:** Informative Prior  $p(\beta | u, \sigma^2) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\beta-u)^2}{2\sigma^2}\right)$ , (see the derivation in Appendix A.2)

### 7.5.1 Linear Regression Model on old Data for Sorghum

A linear regression model was fitted on the old available data on sorghum yield. Two Fisher scoring iterations were needed for the algorithm to converge. The estimated parameters of the model are given in Table 7.7.

**Table 7.7:** Linear regression model fitted on the old data for Sorghum

Variables	Estimate	Std. Err.	t value	Pr(> t )	significant
Intercept)	4.39E+03	2.65E+03	1.65E+00	0.1040	significant
SIZE	-9.19E+02	2.98E+02	-3.08E+00	0.0032	significant
NPK	4.00E+00	2.47E+00	1.62E+00	0.1110	Insignificant
SEED	7.49E+01	3.42E+01	2.19E+00	0.0329	significant
ORGN	-2.76E+03	2.83E+03	-9.75E-01	0.3340	Insignificant

IORG	-9.58E+02	8.81E+02	-1.09E+00	0.2820	Insignificant
INTG	-8.63E+01	1.02E+03	-8.48E-02	0.9330	Insignificant
EDU1	3.35E+02	8.43E+02	3.97E-01	0.6930	Insignificant
EDU2	-5.02E+02	7.62E+02	-6.58E-01	0.5130	Insignificant
EDU3	-7.23E+02	7.01E+02	-1.03E+00	0.3070	Insignificant
EDU4	-3.20E+00	1.29E+03	-2.47E-03	0.9980	Insignificant
EXP	1.04E+00	2.58E+01	4.01E-02	0.9680	Insignificant
EXT	-1.11E+03	1.72E+03	-6.44E-01	0.5220	Insignificant
CREDIT	3.11E+01	1.11E+03	2.81E-02	0.9780	Insignificant

There were a number of insignificant variables at the 5% level of significance. This indicated that only farm size and quantity of seed sown were significant in explaining the sorghum yield. The residual deviance of the model is 2.201E+9 with 54 degrees of freedom.

#### *Interpretation for the parameters of SIZE and SEED*

**In this interpretation, the 5% significance level was taken as criterion of significance.**

- The parameter of SIZE was significant, one hectare increase in farm SIZE with all other variables held fixed, would correspond to a 919 kg/ha decrease in the yield.

- The parameter of SEED was also significant. A unit increase in quantity of seed sown, with all other variables held fixed, would correspond to a 74.9kg/ha increase in the yield.

In order to check the adequacy of the model, collinearity of the independent variables, outliers and influential observations were considered. The correlation matrix of the numerical independent variables is given in Table 7.8.

**Table 7.8 :** Correlation matrix of numerical independent variables on the old data.

	SIZE	NPK	SEED	EXP
SIZE	1.0000	-0.3015	-0.3824	-0.0238
NPK	-0.3015	1.0000	0.1170	0.0281
SEED	-0.3824	0.1170	1.0000	0.0841
EXP	-0.0238	0.0281	0.0841	1.0000

From this correlation matrix, it can be seen that there were no large pair-wise correlations.

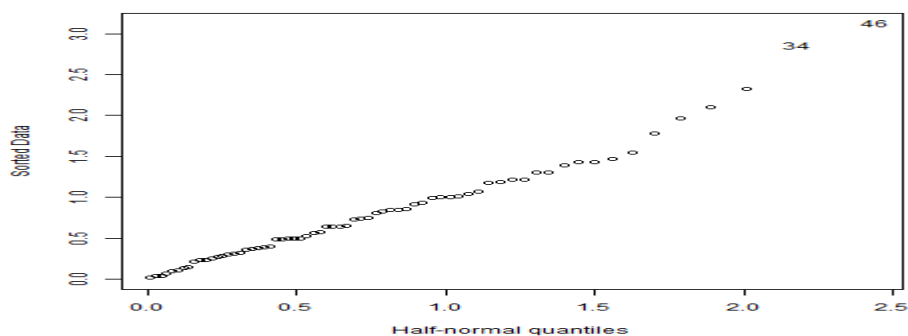
There is no need worry about correlations between two independent variables. The variance inflation factors for each numerical variable are given in Table 7.9.

**Table 7.9:** Variance inflation factors (VIF) of numerical independent variables on the old data.

Variables	VIF
SIZE	1.27102
NPK	1.100506
SEED	1.179045
EXP	1.007692

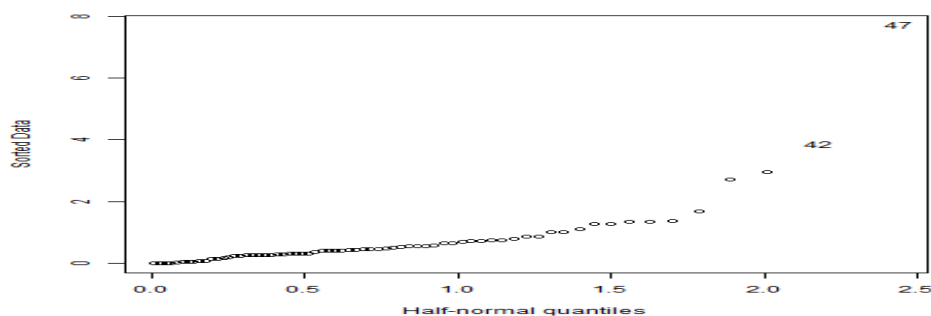
From Table 7.9, it can be seen there were no large variance inflation factors, which indicates that there is no serious problem with collinearity in the old data.

Outliers and influential observations in the model were now considered. The following plots are considered for the presence of outliers and influential observations: Half-Normal plots of the residuals, leverages and Cook's distance statistics. The Half-Normal plot of the residuals is given in Figure 7.5.



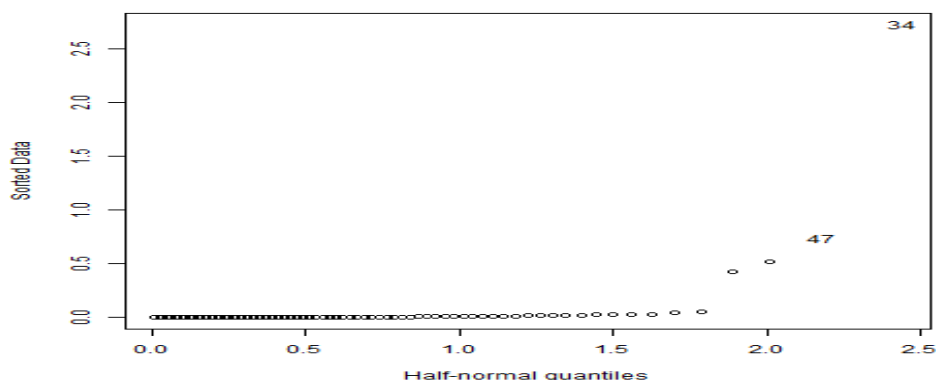
**Figure 7.5:** Half-Normal plot of residuals for the old data.

From Figure 7.5, there does not appear to be any sign of outliers. The Half-Normal plot of the leverages is given in Figure 7.6.



**Figure 7.6:** Half-Normal plot of leverages for the old data.

Figure 7.6 indicates that observations numbered 42 and 47 may have the potential to affect the fit of the model.



**Figure 7.7:** Half-Normal plot of the Cook's distance statistics for the old data

The Half-Normal plot of the Cook's distance statistics is given in Figure 7.7. This plot indicates that observations numbered 34 and 47 may be influential. A linear regression model was then fitted on the old data excluding observations numbered 34, 42, 46 and 47. The estimated parameters of this model were then compared to the model with no removed observations.

**Table 7.10:** Difference in the parameters when possible leverage and influential observations are removed from old data.

Variables	Influential observations include	Influential observations removed
(Intercept)	4.39E+03	4.54E+03
SIZE	-9.19E+02	-4.81E+02
NPK	4.00E+00	6.61E-01
SEED	7.49E+01	-2.16E+01
ORG	-2.76E+03	-8.17E+02
INOG	-9.58E+02	-2.33E+02
INTG3	-8.63E+01	-5.08E+02
EDU1	3.35E+02	-4.16E+02
EDU2	-5.02E+02	-6.35E+02
EDU3	-7.23E+02	-2.41E+02
EDU4	-3.20E+00	-6.33E+00
EXP	1.04E+00	-1.71E+02
EXT	-1.11E+03	2.05E+02
CREDIT	3.11E+01	4.54E+03

Table 7.10 shows the difference in the parameters when possible leverage and influential observations were removed. The first column gives the predictor variables with all observations in the old data and the second column gives the model parameters when the possible leverage and influential observations were removed. Looking at Table 7.10, the

differences in the estimated parameters between the models are minimal. Therefore, the removal of possible leverage and influential points would not significantly affect estimates from the old data.

### 7.5.2 Linear Regression Model on New Data

Two Fisher scoring iterations were needed for the parameters to converge. The estimated parameters are given in Table 7.11.

**Table 7.11:** Linear regression model fitted on the new data.

Predictors	Estimate	Std. Error	t value	Pr(> t )	Significant
(Intercept)	3.99E+03	1.04E+03	3.8428	< 0.001	Significant
SIZE	-7.60E+02	1.55E+02	-4.8935	0.001	Significant
NPK	-8.39E-01	1.46E+00	-0.5765	0.567	Insignificant
SEED	1.23E+01	1.95E+01	0.6333	0.530	Insignificant
ORG	-5.04E+02	9.64E+02	-0.5222	0.604	Insignificant
INOG	3.26E+02	5.03E+02	0.6472	0.521	Insignificant
INTG3	-7.73E+02	6.52E+02	-1.1843	0.242	Insignificant
EDU1	-1.48E+03	6.73E+02	-2.2064	0.032	Significant
EDU2	-7.46E+02	6.20E+02	-1.2028	0.235	Insignificant
EDU3	-5.94E+02	5.19E+02	-1.1443	0.258	Insignificant
EDU4	6.94E+02	9.56E+02	0.7258	0.472	Insignificant
EXP	-1.28E+00	1.37E+01	-0.0934	0.926	Insignificant
EXT	-1.43E+02	9.80E+02	-0.1464	0.884	Insignificant
CREDIT	1.78E+02	5.48E+02	0.3253	0.746	Insignificant

From Table 7.11, it can be seen that the same three variables that were significant. However, variables NPK and EDU4 had their coefficients change from negative in 'old' data to positive coefficients in the 'new' data, while CREDIT changed from positive to negative coefficient in the new data. The residual deviance of the model was 7.36E+06 with 48 degrees of freedom and AIC of 1073.1.

#### *Interpretation for the parameters of SIZE, NPK and SEED*

**In this interpretation, the 5% significance level as criterion for significance. .**

- The parameter of SIZE was significant. A unit hectare increase in farm SIZE , with all other variables held fixed, would lead to be a 492kg/ha decrease in the sorghum yield.
- The parameter of NPK is not significant. A unit increase in quantity NPK fertilizer, with all other variables held fixed, would lead to a 0.46 kg/hectare increase in the sorghum yield.

- The parameter of SEED significant A one -kg/ha increase in SEED alone , would correspond to a 51.8kg per hectare increase in the sorghum yield.

The parameter of IORG was not significant. A unit increase in IORG with all other variables held fixed would correspond to a 284kg/ha above the reference intercept value .

**Table 7.12:** Correlation matrix of the numerical independent variables on the new data

	<b>SIZE</b>	<b>NPK</b>	<b>SEED</b>	<b>EXP</b>
<b>SIZE</b>	1.0000	-0.3187	-0.4044	0.0195
<b>NPK</b>	-0.3187	1.0000	0.3839	-0.2290
<b>SEED</b>	-0.4044	0.3839	1.0000	0.0083
<b>EXP</b>	0.0195	-0.2290	0.0083	1.0000

From this correlation matrix, we see that there were no large pair-wise correlations. The largest correlation is - 0.4044 between SIZE and SEED. The other pair-wise correlations were all very small and insignificant.

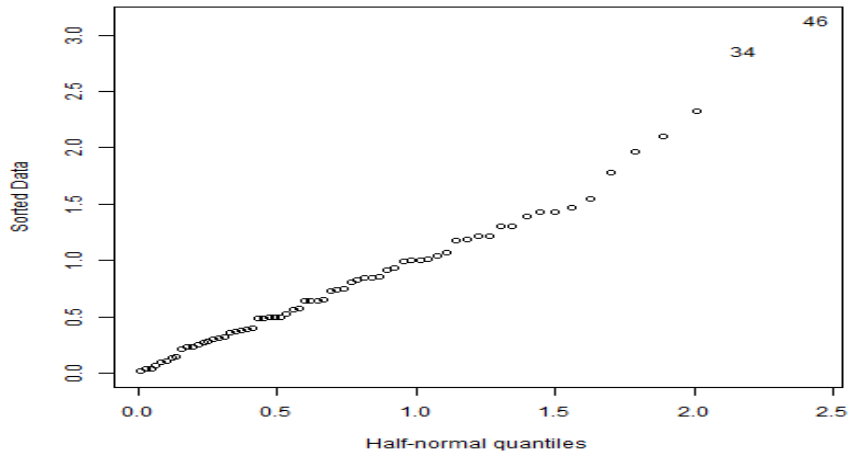
The variance inflation factors for each numerical variable are given in Table 7.13.

**Table 7.13:** Variance inflation factors (VIF) of numerical independent variables on the new data.

<b>Variable</b>	<b>VIF</b>
SIZE	1.243
NPK	1.301
SEED	1.320
EXP	1.068

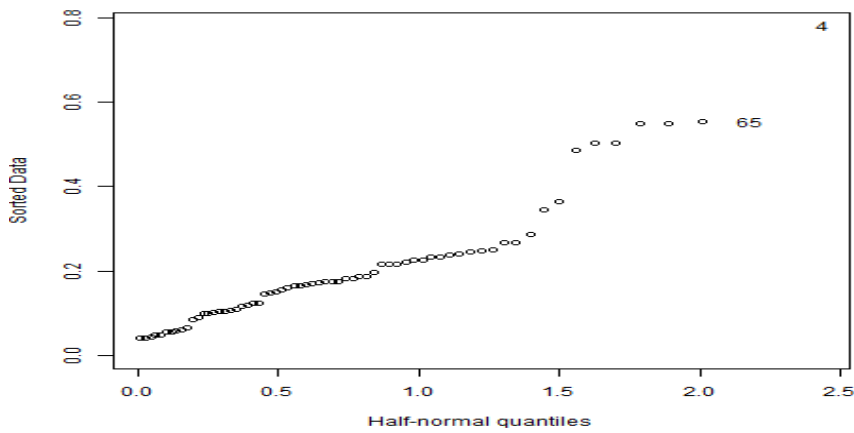
From Table 7.13, there were no large variance inflation factors. Therefore, there was no serious problem with collinearity.

Outliers and influential observations in the model are now considered. A Half-Normal plot of the residuals is given in Figure 7.17.



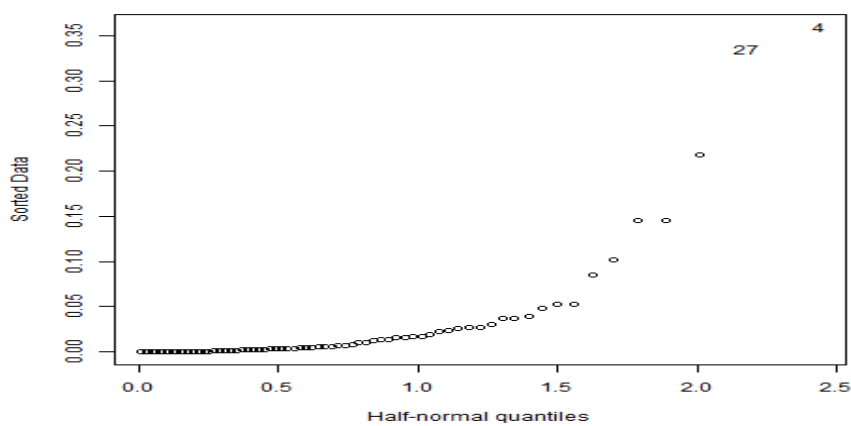
**Figure 7.8:** Half-Normal plot of residuals for the new data.

Figure 7.8 suggests outliers for observations 34 and 46. A Half-Normal plot of the leverages is given in Figure 7.9.



**Figure 7.9:** Half-Normal plot of leverages for the new data.

Figure 7.9 suggests leverage from observations 4 and 65. Finally a Half-Normal plot of the Cook's distance statistics is given in Figure 7.10.



**Figure 7.10:** Half-Normal plot of the Cook's distance statistics for the new data.

From Figure 7.10, there may have been some leverage from observations numbered 4 and 27. Thus, observations numbered 4, 27, 34, 46 and 65 may have been influential observations. In order to see whether these observations were influential they were deleted from the new data and the model refitted. The estimated parameters of this model were then compared to the parameters of the model with no observations deleted. Table 7.14 shows the estimated coefficients arising from the two fittings.

**Table 7.14** Difference in the parameters when possible leverage and influential observations are removed on new data.

<b>Variables</b>	<b>Influential observations include</b>	<b>Influential observations Excluding</b>
Intercept)	2.79E+03	3.99E+03
SIZE	-4.92E+02	-7.60E+02
NPK	0.46E+00	-8.39E-01
SEED	5.18E+01	1.23E+01
ORGN	-3.73E+02	-5.04E+02
IORG	2.84E+02	3.26E+02
INTG	-7.30E+02	-7.73E+02
EDU1	-7.07E+02	-1.48E+03
EDU2	-4.17E+02	-7.46E+02
EDU3	-6.65E+02	-5.94E+02
EDU4	1.31E+03	6.94E+02
EXP	-1.19E+01	-1.28E+00
EXT	-2.43E+02	-1.43E+02
CREDIT	-4.30E+02	1.78E+02

Table 7.14 shows the difference in the parameters when possible influential observations were removed. The first column gives the model parameters with all observations in the new data and the second column gives the model parameters when the possible influential observations were removed. It can be seen that there were significant differences in the estimated coefficients including sign changes in NPK and CREDIT, while SEED became not significant when influential observations were removed. Thus, the model without influential observations was retained for prediction.

A summary of the model fitted on the data with the influential observations omitted is given in Table 7.15.

**Table 7.15:** Linear regression model fitted on the new data.

Variables	Estimate	Std. Err.	t value	Pr(> t )	significant
Intercept)	3.99E+03	1.04E+03	3.8428	0.0004	significant
SIZE	-7.60E+02	1.55E+02	-4.8935	0.0000	significant
NPK	-8.39E-01	1.46E+00	-0.5765	0.5670	Insignificant
SEED	1.23E+01	1.95E+01	0.6333	0.5295	insignificant
ORGN	-5.04E+02	9.64E+02	-0.5222	0.6040	Insignificant
IORG	3.26E+02	5.03E+02	0.6472	0.5206	Insignificant
INTG	-7.73E+02	6.52E+02	-1.1843	0.2421	Insignificant
EDU1	-1.48E+03	6.73E+02	-2.2064	0.0322	significant
EDU2	-7.46E+02	6.20E+02	-1.2028	0.2350	Insignificant
EDU3	-5.94E+02	5.19E+02	-1.1443	0.2582	Insignificant
EDU4	6.94E+02	9.56E+02	0.7258	0.4715	Insignificant
EXP	-1.28E+00	1.37E+01	-0.0934	0.9260	Insignificant
EXT	-1.43E+02	9.80E+02	-0.1464	0.8842	Insignificant
CREDIT	1.78E+02	5.48E+02	0.3253	0.7463	Insignificant

The GLM model given in Table 7.15 was used for prediction, although only one of the 13 predictor variables was significant. The residual deviance of the model is 7.36E+07 with 48 degrees of freedom and AIC value of 1073.1.

### ***Interpretations:***

**In this interpretation, the 5% significance level was adopted as criterion of significance..**

From Table 7.15, it can be observed that SIZE and dummy EDU1 were significant by this criterion. .

- The parameter of SIZE was significant, one hectare increase in SIZE, keeping other variables fixed, would result to a 760kg/ha decreases in the sorghum yield.

The parameter of dummy EDU1 was significant A unit increase in EDU1 with all other variables held fixed, would correspond to 1480kg/ha less from the group reference intercept value in the sorghum yield.

### **7.5.3 Bayesian Linear Regression Model on new Data**

The MCMCregress function was applied from the MCMCpack in R. Two Bayesian linear regression models were fitted: one with an informative prior and the other non-informative prior. The MCMC algorithm used was a random walk Metropolis-Hastings algorithm (see Section 4.1.7). The Bayesian linear regression with informative priors used parameters from the linear regression on the old data as priors. The Bayesian linear regression model using a non-informative prior used a constant prior. The influential observations identified from the

linear regression model on the new data were removed. In order to obtain posterior estimates, a Markov chain with 10,000 samples was generated for both models. The first 5,000 samples were excluded (to allow enough time for the Markov chain to converge to its stationary distribution) which left a Markov chain of 10,000 samples. Therefore, the burn-in period was 5,000.

### ***Bayesian linear regression model with an informative prior***

Prior information was computed from the model fitted on the old data. The model fitted on the old data served as expert agronomic information characteristic of the current agricultural area. This expert knowledge on the linear regression parameters was then used as prior information for the model on the limited amount of new data in the new agricultural location/field. A multivariate normal prior was assumed for the parameters. The prior parameters were also assumed to be independent. The prior coefficients were the coefficients from the linear regression model on the old data. Each coefficient had corresponding information represented in a 14 x 14 diagonal matrix including the intercept.

The prior coefficients and corresponding element in the diagonal matrix are given in Table 7.16.

**Table 7.16:** Prior parameters for an informative Bayesian linear regression model

Variable	Coefficients	information
Intercept)	4.540E+03	1.629E+03
SIZE	-4.809E+02	1.597E+02
NPK	6.605E-01	1.541E+00
SEED	-2.158E+01	4.846E+01
ORGN	-8.170E+02	4.560E+02
IORG	-2.329E+02	5.535E+02
INTG	-5.077E+02	4.793E+02
EDU1	-4.160E+02	3.902E+02
EDU2	-6.347E+02	3.645E+02
EDU3	-2.406E+02	6.720E+02
EDU4	-6.334E+00	1.397E+01
EXP	-1.711E+02	8.934E+02
EXT	2.051E+02	5.668E+02

The posterior estimates of the Bayesian regression model with information prior is given in Table 7.14

**Table 7.17:** Summary of Bayesian Regression Estimates with informative prior (Expert)

<b>Variables</b>	<b>Mean</b>	<b>SD</b>	<b>2.50%</b>	<b>97.50%</b>
(Intercept)	4.37E+03	1.27E+02	4.12E+03	4.63E+03
SIZE	-8.81E+02	4.18E+01	-9.61E+02	-8.00E+02
NPK	2.20E+00	7.78E-01	7.12E-01	3.75E+00
SEED	6.67E+01	6.09E+00	5.50E+01	7.87E+01
ORG	-2.71E+03	1.68E+02	-3.04E+03	-2.38E+03
INOG	-7.23E+02	2.03E+02	-1.12E+03	-3.31E+02
INTG3	-5.15E+02	3.13E+02	-1.13E+03	1.22E+02
EDU1	4.67E+01	2.98E+02	-5.30E+02	6.48E+02
EDU2	-3.92E+02	2.63E+02	-9.09E+02	1.16E+02
EDU3	-7.26E+02	4.23E+02	-1.56E+03	1.20E+02
EDU4	-3.14E+00	5.80E+00	-1.45E+01	8.13E+00
EXP	-1.19E+00	1.25E+01	-2.55E+01	2.40E+01
EXT	-1.06E+03	3.87E+02	-1.84E+03	-3.20E+02
CREDIT	6.03E+01	1.26E+02	-1.85E+02	3.11E+02
sigma2	2.66E+06	5.48E+05	1.79E+06	3.90E+06

**SD= standard deviation**

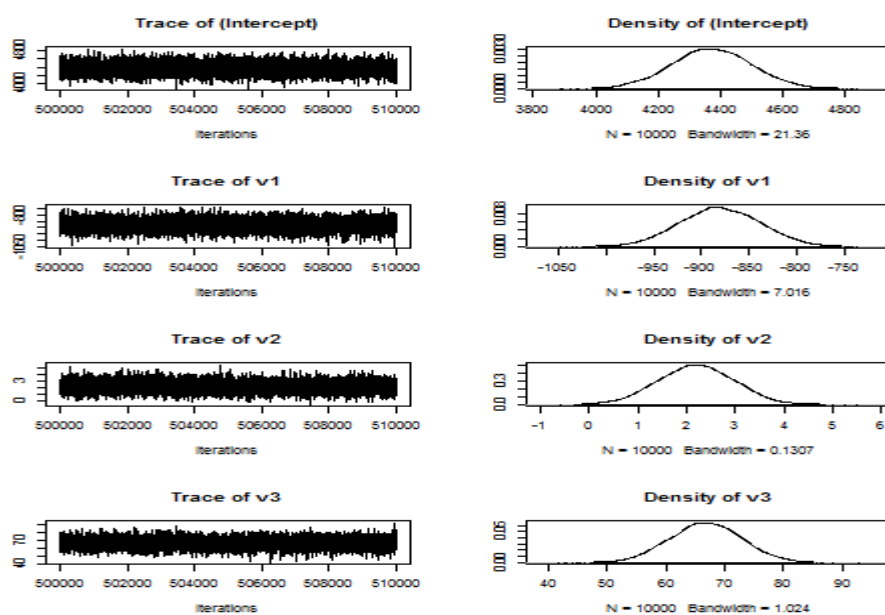
The posterior mean provided the estimate for the parameter. From Table 7.17, the quantiles for each variable can be used to determine which variables were significant at the 5% significance level. The values from the 2.5% to the 97.5% quantiles provided a 95% credible interval for each variable. Eight of the thirteen credible intervals did not include zero, showing that they were significant at 5% level of significance. This shows that many variables included in the model were significant in predicting sorghum crop yield. The parameter estimates had the same interpretation as already discussed for the parameters of NPK and SEED. Again, in the discussion below the absence of zero in the credible interval was taken as showing that the parameter was significant using the 5% significance level as criterion.

- The parameter of SIZE was significant; one hectare increase in FARM SIZE with all other variables held fixed, would produced a 881kg/ha predicted decrease in the sorghum yield.
- The parameter of NPK was significant. A unit increase in NPK with all other variables held fixed, would lead to a 2.20kg/ha predicted increase in the sorghum yield.
- The parameter of SEED was significant. A unit hectare increase in quantity of SEED sown would lead to a 66.7/ha increase in the yield.

Dummy variables EDU[1,2, 3] and INTG variable are not significant. The parameters EDU4, EXT and CREDIT were significant. Their interpretations are as follows:

- The parameter of dummy EDU4 is significant. The presence of EDU4 level, with all other variables held fixed, signifies 16.8kg/ha predicted decrease to the group reference intercept value in the sorghum yield.
- The parameter of dummy EXT was significant. This means that availability of EXT to farmers, with all other variables held fixed, would correspond to a 803kg/ha decrease from the group reference value on the output of sorghum yield.
- The parameter of dummy CREDIT was significant. The availability of CREDIT with all other variables held fixed, would lead to an addition of 6.03kg per hectare on the reference intercept in the sorghum yield.

In conclusion, the effect of informative prior caused a substantial and significant impact on the posterior estimates of the parameters. Trace plots of the Markov chain and density plots of the posterior distributions are given in Figure 7.11. Trace and density plots are only given for the first four parameters. The remaining trace and density plots can be found in the Appendix D1.



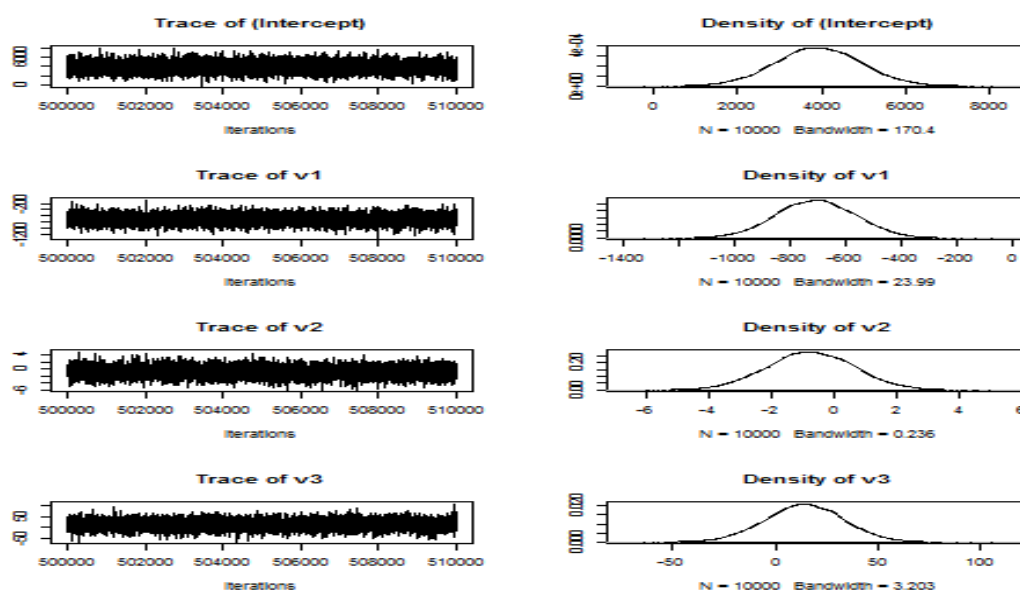
**Figure 7.11:** Trace and density plots of the posteriors for the first four variables using an informative prior

From Figure 7.11, the Markov chain can be seen to have been relatively stationary. This implies that the Markov chain had reached its stationary distribution. The Geweke diagnostic statistics were next calculated for each variable and are given in Table 7.18.

**Table 7.18:** Geweke diagnostic statistics for each variable of the Bayesian linear regression model with informative prior.

variables	z-score
(Intercept)	0.6237
SIZE	1.0410
NPK	-0.9620
SEED	-0.4102
ORG	-0.8095
INOG2	-0.3574
INTG3	1.3210
EDU1	-0.9248
EDU2	-0.0028
EDU3	0.2141
EDU4	0.6376
EXP	-0.3409
EXT	-0.5831
CREDIT	-1.1638
sigma2	-0.4040

Table 7.18 shows that the variables had  $|z| < 2$ . Therefore, all variables had converged according to the Geweke diagnostic. The asymptotic standard normal distribution displayed from the trace plots in Figure 7.12.



**Figure 7.12:** Trace and density plots of the posteriors for the first four variables using a non-informative prior

Trace plots of the Markov chain and density plots of the posterior distributions are given in Figure 7.12. Trace and density plots are only given for the first four parameters. The remaining trace and density plots can be found in the Appendix D1.

**Bayesian Regression model with Non-informative Prior on 'new' data**

**Table 7.19:** Bayesian Regression Estimates with Non-informative Prior on 'new' data

<b>Variables</b>	<b>Mean</b>	<b>SD</b>	<b>2.50%</b>	<b>97.50%</b>
(Intercept)	3.96E+03	1.04E+03	1.93E+03	6.03E+03
SIZE	-7.45E+02	1.48E+02	-1.03E+03	-4.60E+02
NPK	-9.02E-01	1.43E+00	-3.71E+00	1.90E+00
SEED	1.42E+01	1.93E+01	-2.38E+01	5.16E+01
ORG	-4.84E+02	9.67E+02	-2.41E+03	1.39E+03
INOG2	3.16E+02	4.84E+02	-6.35E+02	1.27E+03
INTG3	-6.92E+02	6.14E+02	-1.89E+03	5.29E+02
EDU1	-9.34E+02	6.72E+02	-2.23E+03	3.96E+02
EDU2	-5.81E+02	5.32E+02	-1.63E+03	4.67E+02
EDU3	-6.35E+02	5.08E+02	-1.64E+03	3.52E+02
EDU4	6.92E+02	9.57E+02	-1.17E+03	2.59E+03
EXP	-4.12E+00	1.31E+01	-2.97E+01	2.19E+01
EXT	-9.18E+01	9.81E+02	-2.04E+03	1.85E+03
CREDIT	2.27E+02	5.02E+02	-7.60E+02	1.23E+03
sigma2	1.53E+06	3.18E+05	1.02E+06	2.25E+06

The posterior mean provided the estimate for the parameter. From Table 7.19, looking at the quantiles for each variable it was possible to determine which variables were significant at the 5% significance level. The values from the 2.5% to the 97.5% quantiles provided a 95% credible interval for each variable. Three parameters out of 13 of the credible intervals did not include zero, these predictor variables are SIZE, SEED and CREDIT, and therefore they are significant at 5% level of significance. The interpretations are the same as given in Table 7.19.

The plots show that the Markov chain had reached its stationary distribution. The Geweke diagnostic statistics were next calculated for each variable and the results are given in Table 7.20.

**Table 7.20:** Geweke diagnostic statistics for each variable of the Bayesian linear regression model with non-informative prior.

<b>variables</b>	<b>z-score</b>
(Intercept)	-0.3381
SIZE	0.6067
NPK	-0.2841
SEED	0.5504
ORG	0.5192

INOG2	0.1148
INTG3	0.4448
EDU1	0.4170
EDU2	1.8895
EDU3	0.9256
EDU4	2.3858
EXP	0.9725
EXT	-0.4234
CREDIT	-1.5466
sigma2	2.6594

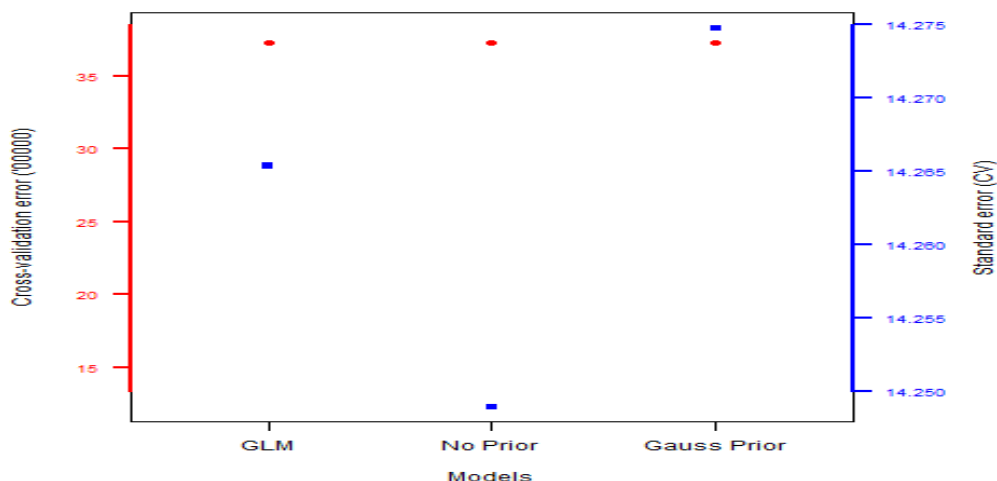
Table 7.20 shows that the variables had  $|z| < 2$ , except EDU1 and sigma2. Therefore, all other variables have converged, except these two variables.

#### 7.5.4 Leave One Out Cross Validation for Model Comparison for Sorghum

Leave-one-out cross validation (LOOCV) was implemented to compare the GLM (M1) and Bayesian linear model with non-informative (M2) and informative (M3) priors. To implement the LOOCV, one observation was excluded from the data and the remaining ( $D_{-i}$ ) data are fitted to the model. The response output was then predicted using the coefficients and compared with the test set (a single value). This process was repeated for all the observations in the data set and the prediction error was computed at each stage. The summary of average prediction error i.e. CV error and the corresponding standard error i.e. SE (CV) are presented on Table 7.21 for the three crop yield models.

**Table 7.21:** Models comparison using the Leave one out cross validation

	Models		
	GLM	No Prior	Gaussian Prior
CV Error	3729490	3728663	3728829
SE(CV)	1426533	1424893	1427473



**Figure 7.13:** Average Cross Validation error and the standard error of CV error for the three models of Sorghum yield

From Figure 4.13, it can be seen that GLM (Model 1) had the least average cross validation error and highest CV error, while Bayesian model with non- informative prior (Model 2) had highest average cross validation error and least predictive error. The Bayesian model with informative prior (Model 3) had a moderate CV and predictive error. From this analysis it seems that GLM was a worse model than the two Bayesian models.

## 7.6 Results on Linear Regression Analysis on Split Data for Millet

Table 7.22 shows the MLE estimates fitted on the new data excluding the influential and outlying observations.

**Table 7.22:** Linear regression model fitted on New data

Variables	Estimate	Std. Err.	t value	Pr(> t )	significant
Intercept)	1.32E+03	4.03E+02	3.27618	0.00266	significant
SIZE	-7.94E+01	1.11E+02	-0.71624	0.47938	insignificant
NPK	1.78E-01	7.45E-01	0.23937	0.81245	Insignificant
SEED	-5.01E+00	6.00E+00	-0.83488	0.41039	insignificant
ORGN	6.36E+02	3.54E+02	1.79828	0.08220	Insignificant
IORG	-4.05E+02	2.34E+02	-1.73203	0.09354	Insignificant
INTG	-2.54E+02	2.93E+02	-0.86634	0.39318	Insignificant
EDU1	-5.04E+02	2.55E+02	-1.97624	0.05739	Insignificant
EDU2	9.63E+01	2.06E+02	0.46799	0.64317	Insignificant
EDU3	2.75E+02	1.90E+02	1.44342	0.15927	Insignificant
EDU4	1.35E+01	8.82E+00	1.52750	0.13711	Insignificant
EXP	3.89E+02	4.35E+02	0.89268	0.37914	Insignificant
EXT	1.32E+03	4.03E+02	3.27618	0.00266	significant
CREDIT	-7.94E+01	1.11E+02	-0.71624	0.47938	Insignificant

The model given in Table 7.22 was used for prediction, although only one of the 13 predictor variables was significant. The residual deviance of the model was 4.53E+06 with 41 degrees of freedom and AIC value of 631.97.

***Interpretations:***

From Table 4.22, it can be observed that only dummy EXT was significant at 5% level of significance.

- The parameter of EXT was significantly different from its reference intercept at 5% ( $p < 0.05$ ), it indicated that a unit increase in EXT, keeping other variables fixed, would result to a 1320kg/ha addition on the reference value of millet yield.

### 7.6.1 Bayesian Regression Model on New Data

Table 7.23 shows the posterior estimates of the Bayesian model with informative prior fitted on the new data excluding the influential and outlying observations.

***Bayesian Regression model with informative Prior on new data***

**Table 7.23:** Summary of Bayesian Regression Estimates with Informative Prior (Expert) on new data

Predictors	Mean	SD	2.50%	97.50%
Intercept)	2.60E+03	8.36E+01	2.43E+03	2.76E+03
SIZE	-3.77E+02	3.10E+01	-4.37E+02	-3.16E+02
NPK	05.81E-01	6.15E-01	-6.24E-01	1.79E+00
SEED	-1.12E+01	3.09E+00	-1.73E+01	-5.15E+00
ORGN	1.79E+03	1.03E+02	1.59E+03	1.99E+03
IORG	-4.25E+02	1.54E+02	-7.29E+02	-1.25E+02
INTG	-6.04E+02	1.62E+02	-9.22E+02	-2.86E+02
EDU1	-1.31E+01	1.72E+02	-3.48E+02	3.27E+02
EDU2	-2.48E+02	1.55E+02	-5.51E+02	5.55E+01
EDU3	1.37E+02	1.91E+02	-2.44E+02	5.07E+02
EDU4	9.96E+01	2.90E+00	9.40E+01	1.05E+02
EXP	6.62E+00	7.88E+00	-8.77E+00	2.22E+01
EXT	-1.61E+02	1.73E+02	-5.00E+02	1.81E+02
CREDIT	-3.56E+02	8.43E+01	-5.22E+02	-1.91E+02
sigma2	4.40E+05	1.02E+05	2.83E+05	6.77E+05

SD= standard deviation

The posterior mean provided the estimates for the parameters. From Table 7.23, looking at the quantiles for each variable it was possible to determine which variables were significant at the 5% significance level. The values from the 2.5% to the 97.5% quantiles provided a 95% credible interval for each variable. Many of the credible intervals did not include zero

showing that they were significant in predicting millet crop yield. The parameter estimates suggested that the parameters of SIZE and SEED were significant at the 5% level of significance.

### *Interpretations*

**In this interpretation, the absence of zero in the credible interval was taken as showing that the parameter was significant using the 5% significance level as criterion.**

- The parameter of SIZE was significant. A 1 hectare increase in SIZE with all other variables held fixed, would correspond to a 377 kg/ha decrease in millet yield.
- The parameter of NPK was not significant. A unit increase in NPK on its own would correspond to a 0.581 kg/ha increase in the millet yield.
- The parameter of SEED was significant. A 1kg/ha increase in SEED on its own would correspond to 11.2kg/ha decrease in millet yield.
- The parameter of the dummy variables of SOIL (ORG, IORG and INTG) was significantly different from the group reference intercept, when no soil supplement is used. The parameter ORGN would contribute 1790kg/ha to millet yield. The parameters of IORG and INTG would cause respectively 425kg/ha increase and 604kg/ha decrease to the production of millet.
- The parameter of EDU4 was significant. The presence of .EDU4 level would decrease millet yield by 161kg/ha. Other parameters of EDU were not significant at 5%.
- The parameter of CREDIT was significant. The availability of credit on its own would decrease by 356kg/ha from group reference on millet yield.

**Table 7.24:** Geweke diagnostic statistics for each variable of the Bayesian linear regression model with informative prior (millet)

<b>variables</b>	<b>z-score</b>
(Intercept)	0.408181
SIZE	0.268472
NPK	-0.17353
SEED	-0.00624
ORG	-0.94699
INOG2	-0.65987
INTG3	-0.61933
EDU1	-2.3394
EDU2	-0.02558
EDU3	1.15249

EDU4	-0.02457
EXP	0.818889
EXT	-0.27162
CREDIT	-1.92081
sigma2	-1.15997

Table 7.24: shows that the variables all had  $|z| < 2$ .

**Bayesian Regression model with Non-informative Prior on new data**

**Table 7.25:** Bayesian Regression Estimates with Non-informative Prior on new data

Variables	Mean	SD	2.50%	97.50%
(Intercept)	1.44E+03	7.13E+02	3.02E+01	2.85E+03
SIZE	-1.53E+02	1.44E+02	-4.38E+02	1.32E+02
NPK	-3.61E-01	9.59E-01	-2.25E+00	1.53E+00
SEED	-6.05E+00	8.36E+00	-2.25E+01	1.04E+01
ORG	6.07E+02	4.97E+02	-3.71E+02	1.59E+03
INOG2	-1.79E+02	3.09E+02	-7.85E+02	4.28E+02
INTG3	-1.71E+02	3.90E+02	-9.38E+02	5.96E+02
EDU1	-1.60E+02	2.86E+02	-7.25E+02	4.05E+02
EDU2	-9.88E+00	2.85E+02	-5.75E+02	5.48E+02
EDU3	3.44E+02	2.59E+02	-1.67E+02	8.57E+02
EDU4	2.03E+03	5.81E+02	8.83E+02	3.18E+03
EXP	1.89E+00	1.18E+01	-2.13E+01	2.51E+01
EXT	2.99E+02	4.43E+02	-5.76E+02	1.17E+03
CREDIT	-2.51E+01	4.85E+02	-9.82E+02	9.32E+02
sigma2	3.02E+05	7.92E+04	1.84E+05	4.91E+05

Table 7.25 shows the model estimates with non-informative prior. Many of the variables include zero in the credible intervals of 2.5% and 97.5%. The result is similar to the GLM outputs. The significant parameters are intercepts, ORG, and EDU4

**Table 7.26:** Geweke diagnostic statistics for each variable of the Bayesian linear regression model with non-informative prior (millet)

variables	z-score
(Intercept)	0.50869
SIZE	-0.11359
NPK	-0.18651
SEED	-0.03420
ORG	-0.88970
INOG2	-0.84246
INTG3	-1.01854
EDU1	-1.84971
EDU2	-0.40990

EDU3	0.25283
EDU4	-0.32977
EXP	0.27468
EXT	-0.00334
CREDIT	-0.92988
sigma2	-0.86976

Table 7.26 shows that the variables have  $|z| < 2$ . Therefore, all variables had converged.

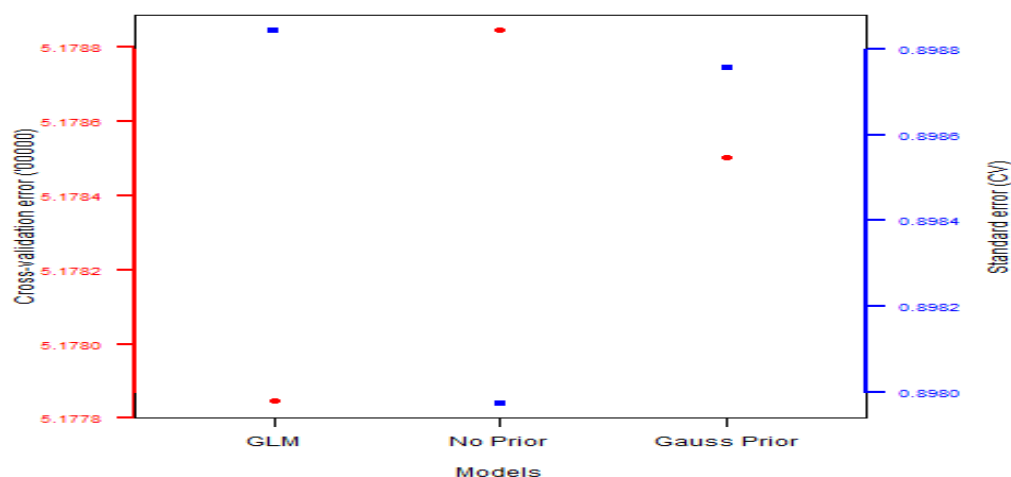
### 7.6.2 Leave One Out Cross Validation for Model Comparison

Leave-one-out cross validation (LOOCV) was implemented to compare the GLM (M1) and Bayesian linear model with non-informative (M2) and informative (M3) priors. The resulting average CV errors and SE(CV error) are given in Table 7.26 for millet yield.

**Table 7.27:** Models comparison using the Leave one out cross validation

	Models		
	GLM	No Prior	Gaussian Prior
CV Error	517784.6	517884.4	517850.2
SE(CV)	89884.43	89797.29	89875.78

Table 7.23 shows cross validation error and its standard error for the three models. The dot plot of the three models is presented in Figure 7.14



**Figure 7.14:** Average LOOCV error and the standard error CV for the three models

From Figure 7.14, it can be seen that GLM (Model 1) had the least average cross validation error and highest CV error, while Bayesian model with non-informative prior (Model 2) had highest average cross validation and least predictive error. Bayesian model with informative

prior (Model 3) had moderate CV and predictive errors. Again, it seems from this that GLM was a worse model than the two Bayesian models.

## 7.7 Assessing the Model Accuracy and Residual Plots

In this section, we investigate the model accuracy to predict the new data. We have plotted the observed yield versus predicted (*left panel*) and the residual plots versus the average of the observed and predicted (*right panel*). A plot of residuals  $\varepsilon_i$ 's against the fitted values  $\hat{y}_i$ 's is the residual plot, which is a simple and convenient tool for regression model diagnosis. The residuals evenly distributed on both sides of  $y = 0$  imply that the assumptions  $E(\varepsilon) = 0$  and constant variance. The purposes of the residual analysis are to detect model misspecification and to verify model assumptions. Among other uses of residual plot in regression analysis model, it is used to check overfitting, model underfitting, and for outlier detection. Overall, residual analysis is useful for assessing a regression model diagnostics.

### 7.7.1 Residual Plots on Sorghum

Figures 7.15 – 7.21 display the graphs of respective models: linear model fitted by the OLS methods (M1), Bayesian non-informative prior (M2) and Bayesian with informative prior (M3) on sorghum. Figure 7.15- 7.21 demonstrate how one can make sense about the accuracy of the model is related the observed value. The left panel graphs show that the scatterplots between the observed and fitted are weakly correlated. This pattern indicates that the yield can be predicted from the models, but with low accuracy. The residual plots showed that the points are clustered about line  $y=0$ . Models (M1 and M2) clustered closely, while M3 scattered about line  $y=0$ . M1 underpredicted the yield as the pattern looks like concave up, perhaps some higher order factor terms were missing. M1 underpredicted the yield (Figure 7.16) as many points are above the line zero. M2 predicted more accurately (Figure 7.18) and M3 overpredicted (Figure 7.20) as some points can be seen below line  $y=0$ .

### Model 1: Linear Model

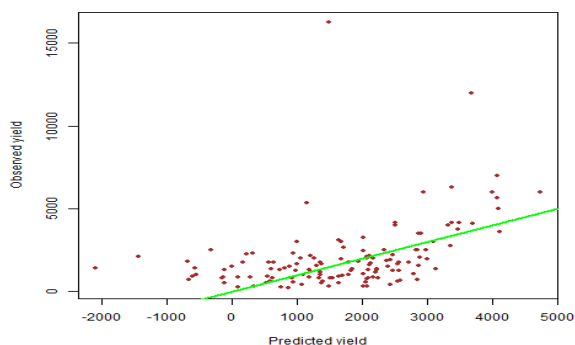


Figure 7.15: Plot of Predicted versus observed values(GML)

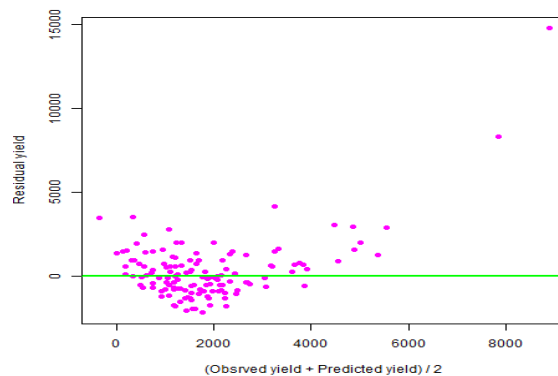


Figure 7.16: Half of Predicted and observed versus residuals plot

**Model 2: Non informative Prior**

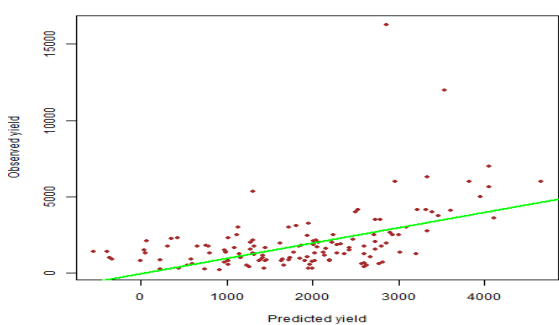


Figure 7.17: Plot of Predicted versus observed values

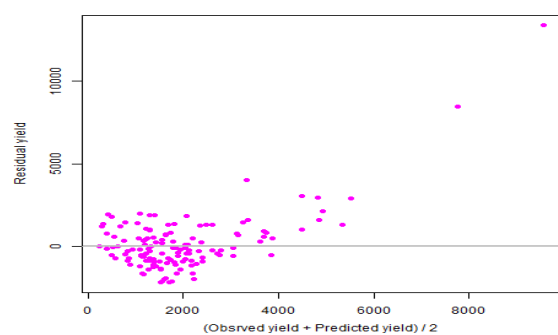


Figure 7.18: Plot of Predicted and observed versus residual plot

**Model 3: Informative Prior**

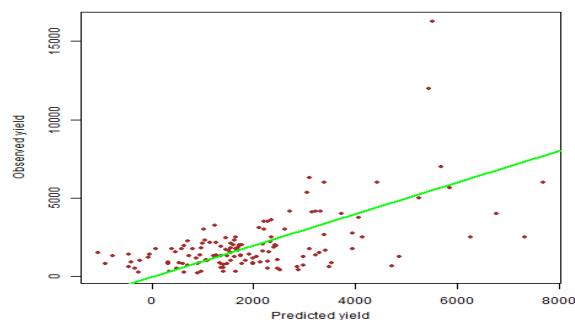


Figure 7.19: Plot of Predicted versus observed values

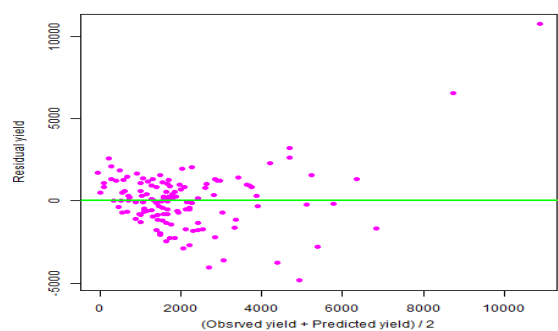


Figure 7.20: Plot Half of Predicted and observed versus residual plot

**Combined residual Plot**

Figure 7.18 presents the combined graph of the residual plot. M3 predicted more accurately than the other two models. M3 clustered closely to the line  $y=0$ , M3 would predict more accurately than M1 and M2.

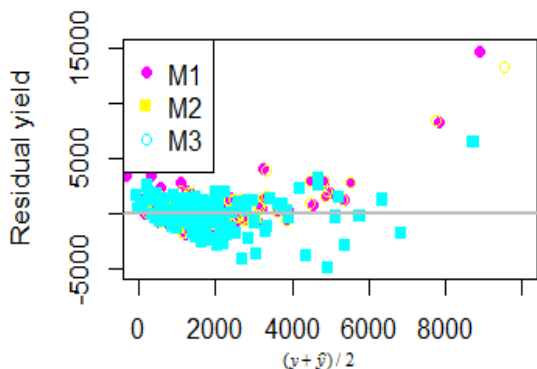


Figure 7.21: Combined Model: Half of Predicted and observed versus residual plot(Sorghum)

### 7.7.1 Residual Plots on Millet Yields

In the left panels, the plots between the predicted and observed for the three models did not show defined patterns, therefore there was a weak relationship between the observed and model predicted value.

The residual plots for the models were not evenly distributed around line  $y=0$ , thus these models may not accurately predict the millet yield. The linear model M1, the residual plot shows a downward curvature and overestimated the yield. This indicates that the variance is not homogenous and possible some higher order terms are missing in the model. M2 underfitted as most residual points can be seen above the line  $y=0$ .

#### Model 1: Linear model

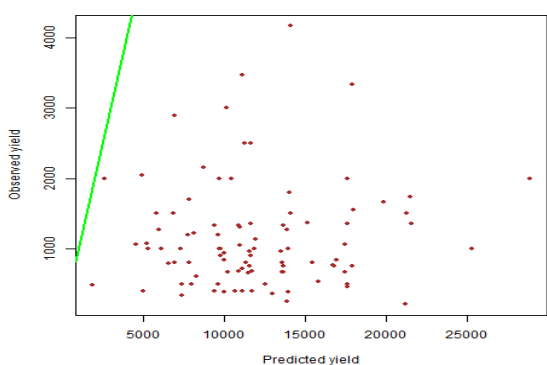


Figure 7.22: Plot of predicted versus observed value (GLM)

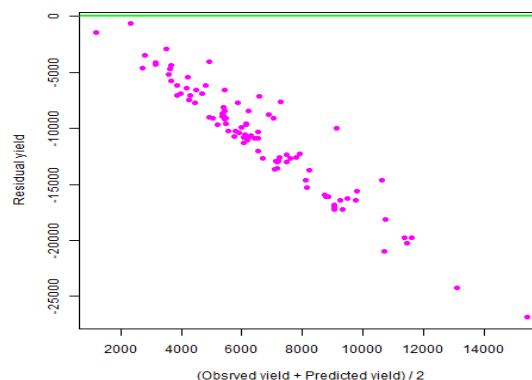


Figure 7.23: Half of predicted and observed versus residual plots (GLM)

**Model 2: Bayesian non informative Prior**

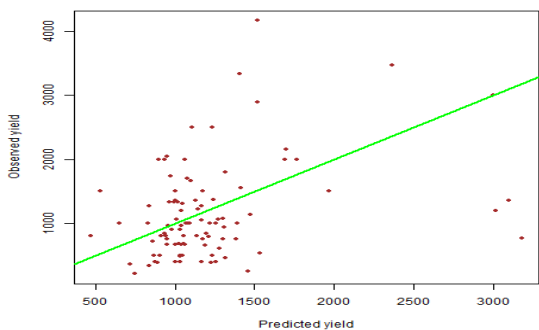


Figure 7.24: Plot of predicted versus observed value (Non informative)

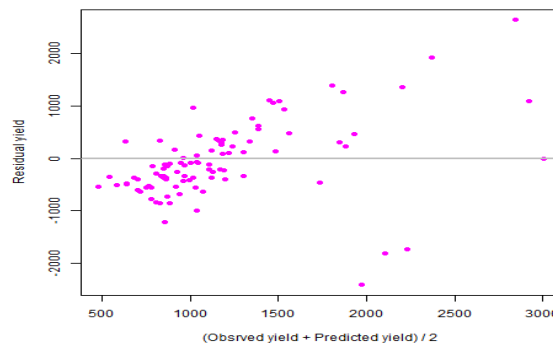


Figure 7.25: Half of predicted and observed versus residual plots (Non-informative)

**Model 3: Bayesian with informative Prior**

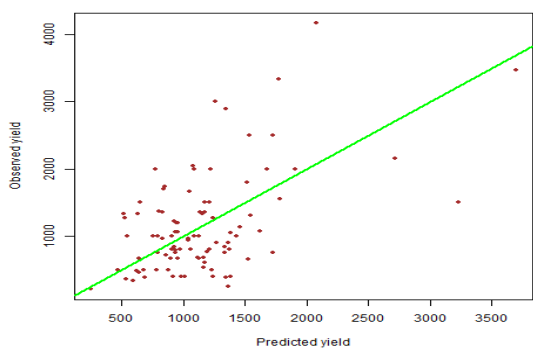


Figure 7.26: Plot of predicted versus observed value (Informative)

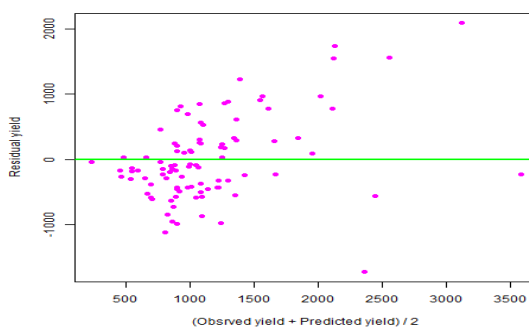


Figure 7.27: Half of predicted and observed versus residual plots (Informative)

**Combined Model Plots**

Figure 7.22 displays the graph of the half of the predicted and observed values versus the residuals. Model M3 predicted more accurately than the other two models. Model M1 had overpredicted as it can be seen the purple points fall below the zero line.

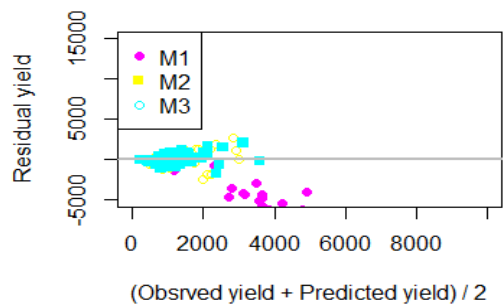


Figure 7.28: Combined Model Plot of predicted and observed versus residual plot (Millet)

## 7.8 Summary and Discussion

In the first part of this chapter, classical approaches were applied to fit a general linear model to the full data: maximum likelihood estimation and Bayesian framework... It was observed that the frequentist estimates results were similar to Bayesian analysis when non-informative prior was used on the crop yields.

Bayesian analysis has the advantage of incorporating external information such as historical data from previous research work (e.g. Meta analysis research) The expert beliefs or hyperparameters incorporated into the regression analysis yielded a narrower credible interval, indicating the importance of prior belief in reducing the variability and uncertainty in the estimates. The uncertainty in the estimates is reduced when the informative prior is used. The objective credible interval has better performance in this regard than any of the classically derived confidence intervals. Furthermore, it can be seen that the objective Bayesian intervals are, on average, smaller than the classically derived intervals, (See Mossman and Berger (2001) for more extensive computations. The analysis presented in this chapter has shown that expert knowledge can play a useful role in deciding important variables for predictions.

In contrast, vague prior are sometimes used in this analysis to represent a state of little or no prior knowledge about the process. This decision is in line with Gamerman (1997) and Lambert (2005). Uniform priors or Jeffrey's prior are assumed non-informative priors. A vague prior i.e. non-informative prior can influence a standard choice for parameters with large variance. The use of vague priors can be problematic due to small amount of data. Hence choosing a vague prior distribution is heavily dependent on the situation. Discussion about the effect of different priors is conducted in Gamerman (1997) and Lambert (2005).

## 8. Results - Analysis of Bayesian Model Selection and Model Average

This chapter presents the results of application of two concepts: Bayesian Model selection (BMS) and Model Averaging (BMA). The theoretical concepts discussed in chapter 4 were applied to analyse crop yield data and make relevant inferences. BMA package was implemented in R environment using the **bicreg** and **BMA** commands within the *LearnBayes* package. The R code for the analysis is presented in Appendices F and G

### 8.1 Bayesian Model Average (BMA) analysis

This section reports on the application of the BMA Package to fully analyse the data on the three cereal crops (Sorghum, Millet and Millet). The methods generated posterior inclusion probability (PIP), expected values and standard deviation for the predictors. BMA analysis was expected to generate  $M = 2^k = 2^{13} = 8192$  models, where  $k = 13$  is the number of predictor variables in the model. This could be quite computationally demanding and demand cumbersome interpretation.

**The BMA Package:** The BMA package (an acronym for Bayesian model averaging), performs BMA analysis assuming a uniform distribution on model priors and using a simple BIC (Bayesian Information Criterion) approximation to construct the prior probabilities on the regressions coefficients (Raftery, et al., 2010). This package is built upon Raftery (1995) algorithm. The main functions in the BMA package to implement a BMA regression analysis is the function **bicreg** in R environment.

#### 8.1.1 Bayesian Model Averaging (BMA) analysis for Sorghum

Table 8.1 presents the summary posterior estimates obtained from applying Bayesian model averaging and t-statistics for significance test of coefficients. The dependent variable was sorghum yield (kg/ha) of the sampled population.

The posterior estimates resulted from application of the BMA package and R code is presented in Appendix F.1 for sorghum. The number of explanatory variables in this case was 13 and considered the full enumeration of model space in term of posterior inclusion probability over the possible combinations.

From Table 8.1, the importance of each predictor can be judged by its posterior inclusion probability  $p(\beta \neq 0 | D)$ . Suppose one intends to decide which variables are to be included in the final model, to column (3) under BMA. It can be observed that farm size had the posterior inclusion probability of 1.00. It is obvious that land size is an important factor of production, especially farming operation. Also, the posterior distribution of the predictor quantity of seed sow (**kg/ha**) had posterior inclusion probability (PIP) of 0.49.

For the categorical variable, education level of farmers, the posterior distribution probability of variable (x9) was 0.30 with a negative coefficient on sorghum yield. This means that a farmer with that higher level of education would have a reduced contribution on the output relative to the non educated farmer (reference level). Moreover, variable x9 had a larger negative coefficient (-229) than variable x8 coefficient (-32.9). This means that  $x_9$  had a lesser contribution than variable x8 on their differential scale of output relative to the non educated farmer (reference level).

Farmer experience (continuous variable) had a positive effect on the yield, but this was not significant at 5% significance level.

On the same Table 8.1, the binary variable, ( extension service,x12) is dropped out, meaning that access to extension services (coded 1) had negligible contribution on response output relative to those who had no contact with extension officers.

The binary variable (credit facility, x13) has a positive coefficient with a posterior probability of 0.03. The effect is not significant different from those farmer who had no access to credit facility (reference level intercept).

**Table 8.1:** Summary estimates and inclusion Probability using BMA and p-value on sorghum

Predictors	BMA analysis			Frequentist
	Mean $\beta   D$	SD $\beta   D$	$p(\beta \neq 0   D)$	P-values Pr(> t )
Intercept	3846	90.9	1.00	0.00225 **
Farm Size ( $x_1$ )	-842.9	140.8	1.00	1.22e-06 ***
Fertilizer ( $x_2$ )	0.272	0.796	0.15	0.31499
Seed Quantity ( $x_3$ )	19.55	23.83	0.49	0.06481

SOIL1 ( $x_4$ )	-51.12	317.3	0.06	0.44880
SOIL2 ( $x_5$ )	DROPPED			0.75013
SOIL3 ( $x_6$ )	27.53	138	0.07	0.73349
EDU1 ( $x_7$ )	24.14	152	0.05	0.98254
EDU2 ( $x_8$ )	-32.9	169	0.07	0.36663
EDU3 ( $x_9$ )	-229	417	0.30	0.08619
EDU4 ( $x_{10}$ )	49.6	352	0.05	0.46806
Exp. ( $x_{11}$ )	0.06	1.53	0.15	0.96171
Ext. ( $x_{12}$ )	DROPPED			0.86824
Credit ( $x_{13}$ )	9.28	106	0.03	0.74691

\*\*\* Significant at 1%, \*\* significant at 5% significance level.

PIP=Probability of inclusion, SD=Standard deviation,

BMA was applied on sorghum yield (*kg/ha*) as dependent variable, 27 models were selected by the OCCAM window, the results for the five top models are written out below together with their individual model posterior probabilities. The cumulative posterior probability of these five top models was 0.563. These models are averaged over all possible combination of predictors. The posterior mean and standard deviation of individual variable is provided in the Table 8.1.

The five top models are fully written out below including the intercepts and BIC selection values. Detailed print out of the R code and analysis is provided in Appendix E.1.

- $M_1: y = 4432.2 - 903.4 x_1$   
 $R^2 = 28.7\%$  BIC =  $-40.75$ ,  $p(m_1 | y, X) = 0.167$  .
- $M_2: y = 3230.1 - 787.77 x_1 + 40.85 x_3$   
 $R^2 = 31.2\%$  BIC =  $-40.74$ ,  $p(m_2 | y, X) = 0.165$  .
- $M_3: y = 4593.1 - 918.2 x_1 + (\beta_0 - 783.9) * x_6$  ( $\beta_0$  is reference intercept)  
 $R^2 = 30.6\%$ , BIC =  $-39.54$ ,  $p(m_3 | y, X) = 0.091$  .
- $M_4: y = 4593.1 - 918.2 x_1 + 39.46 x_3 + (\beta_0 - 748.3) x_6$  ( $\beta_0$  is ref intercept)

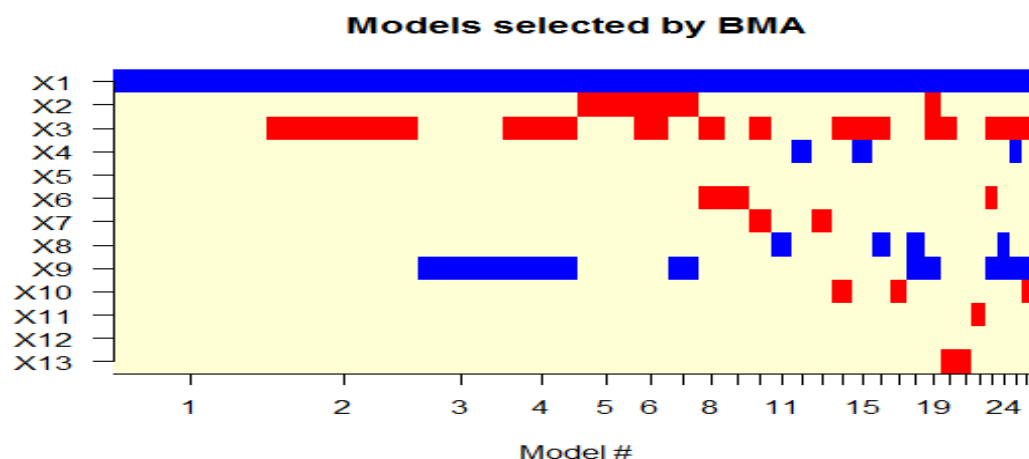
$$R^2 = 33\% \quad \text{BIC} = -39.31, \quad p(m_4 | y, X) = 0.081$$

- $M_{.5}: y = 4593.1 - 836.1x_1 + 1.985 * x_2$

$$R^2 = 30.2\% \quad \text{BIC} = -38.73, \quad p(m_5 | y, X) = 0.060$$

Using BMA, the selection criterion of highest posterior probability,  $\text{prob} = 0.167$ ,  $M_1$  would be selected best among these models, That choice coincides with lowest Bayesian Information criterion of -40.75 among the selected models. In a similar way, a frequentist analyst would select a model with highest  $R^2$ . Thus, the best model for sorghum is  $y = 4432.2 - 903.4x_1$  and a competing model  $m_2: y = 3230.1 - 787.77 * x_1 + 40.85x_3$  with  $\text{BIC} = -40.74$ , and posterior probability equal to 0.165 .

The output was displayed visually in Figure 8.1. The resulting models are listed along the horizontal axis in order of BIC approximation close to the minimum (R code is provided in Appendix F.1). The blocks indicate significant regression coefficients coloured red for positive and blue for negative effect on sorghum yield. The posterior mean of coefficients averaged over all these models are given in the output data frame. Those variables with more blocks will more often be selected in the result.

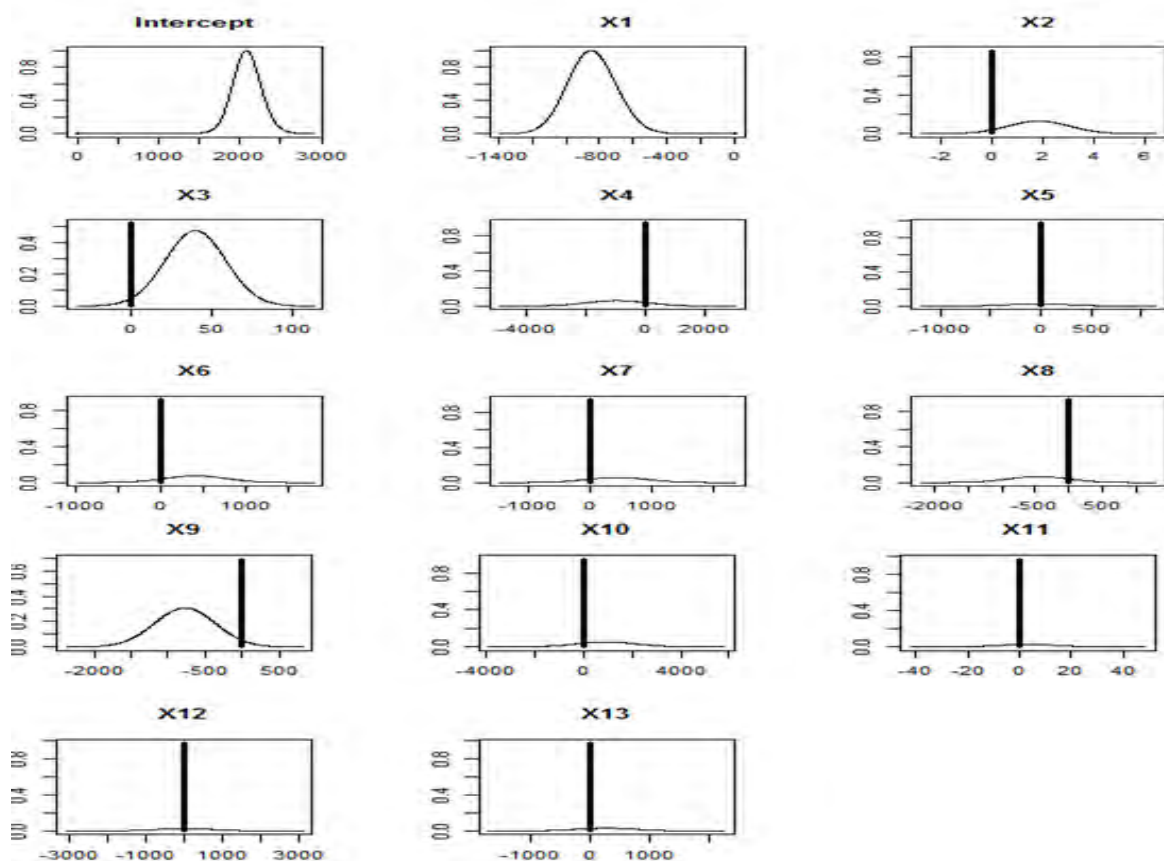


**Figure 8.1:** BMA image plot of selected models sorghum yield (kg/ha) and associated Predictors. See text above for explanation.

Figure 8.1 presents model combinations and the regression coefficients. A long block of bar for the predictors  $(x_1, x_2, x_3, x_9)$  indicates substantial influence on the outcome. The predictor variables  $(x_2, x_3, x_6, x_{10})$  had positive contribution to the sorghum yield. From Table 8.1, it can be seen that  $x_2$  had little significant effect on the response and  $x_3$  had a significant positive effect on sorghum yield.

- The predictor variable  $x_6$  was a categorical variable level of soil supplements; it can be interpreted as having positive contribution to sorghum yield above the reference level that reflected no additional soil supplement ( $\beta_0 = 3846$ ).

The individual plot of the posterior distribution of each variable for sorghum yield is displayed in Figure 8.2. The density plots are displayed in 5 rows and 3 columns. It can be observed that among the density plots, those variables of little influence on yield contain a spike at zero. e.g.  $x_{11}$ ,  $x_{12}$ ,  $x_{13}$ . Their coefficients were centered on and most likely were, zero. The marginal posterior probability percentages of individual variables excluding the intercept, are (89.9, 3.7, 27.3, 64.3 64.9, 40.7, 2.8, 10.1, 2.8 18.9 7.4 4.0 14.0). The density plots of the predictors are given in Figure 8.2.



**Figure 8.2:** Density plots of the predictor variables on Sorghum yields using BMA

From Figure 8.2, it can also be observed that the density plots of the predictors with low posterior probability were flattened. It would be included in the models. Variables  $x_3$  and  $x_9$  contributed moderately to yield,  $x_2$ ,  $x_4$  and  $x_6$  were less substantial, while the density plots of the remaining variables ( $x_5$ ,  $x_7$ ,  $x_{10}$ ,  $x_{11}$ ,  $x_{12}$ ,  $x_{13}$ ) are almost flat, indicating they did not contribute significantly to the response variable of interest. Hence, the risk of overfitting

could be minimized. Thus, BMA analysis excludes redundant variables from the model and addresses the problem of covariates uncertainty selection. The variables ( $x_1$ ,  $x_3$ , and  $x_9$ ) are identified as variables for prediction of sorghum production.

### 8.1.2 Model Selection Using AIC Weight of Evidence on Sorghum Yield

Here, BMA analysis was used as a trade off criterion for selection 10 top models with highest posterior probabilities. The AIC value of each model computed and presented in column 4. Here, the individual AIC value is not directly interpretable due to an unknown constant. AIC is only comparable, relative to other AIC values in the set of selected candidate models. Accordingly, the difference between the minimum AIC (selected as the best) and the other AIC values was computed, as is important and useful in information theory.

**Table 8.2:** AIC Weights Estimates for Model Comparison models on Sorghum

Model $m$	Predictors	$p(m y, X)$	AIC	$\Delta_i$ AIC	$l(m y)$	$w_i$
1	$x_1$	0.1484	2401.60	4.3751	0.1122	0.0310
2	$x_1, x_3$	0.1476	2398.71	1.4806	0.4770	0.1319
3	$x_1, x_9$	0.0809	2399.91	2.6825	0.2615	0.0723
4	$x_1, x_3, x_9$	0.0724	2397.23	0.0000	1.0000	0.2766
5	$x_1, x_2$	0.0540	2400.72	3.4896	0.1747	0.0483
6	$x_1, x_2, x_3$	0.0332	2398.71	1.4806	0.4770	0.1319
7	$x_1, x_2, x_9$	0.0290	2399.06	1.8302	0.4005	0.1108
8	$x_1, x_3, x_6$	0.0250	2399.36	2.1287	0.3450	0.0954
9	$x_1, x_6$	0.0243	2402.32	5.0883	0.0785	0.0217
10	$x_1, x_3, x_7$	0.0209	2399.71	2.4811	0.2892	0.0800

$p(m|y, X)$  is the posterior model probability from BMA.

AIC- Akaike information criterion

$\Delta_i$ AIC=AIC-AIC<sub>min</sub> is the information loss

$l(m|y)$  is the likelihood of the model given the data

$$w_i = \frac{\exp(-\frac{1}{2}\Delta AIC_i)}{\sum_{r=1}^R \exp(-\frac{1}{2}\Delta AIC_r)}$$

$w_i$  is the Akaike weight or relative probability for model  $i$ , which can be interpreted as relative posterior probability of the model given the data for this set of the candidate models.

From Table 8.3, the column (3) is posterior probability of the model as estimated by BMA among the possible predictor combination. Column 4 is the AIC value the model and column 5 is the difference between the individual AIC value and the least AIC, assume as the best model relative to the other candidate models. The column 6 is the likelihood of the model given the data, called the probability of correct selection of the model. The relative Akaike weight for model  $m_4$  was 0.2766 and it would be selected as the best model given the data. Hence, the best model for sorghum yield included small number of predictor variables: farm size, NPK fertilizer and degree of education.

Generally, column  $w_i$  can be interpreted as relative posterior probability of each model given the data and it can be termed literarily as measure of evidence. The ratio of the best model Akaike weight to other model is the relative importance or the chance of being selected among the candidate models. The competing models to the selected model are  $m_2$  and  $m_6$ .

For example  $\frac{w_3}{w_2} = \frac{0.2766}{0.1319} = 2.097$ . This value is termed as evidence ratio; it literally means

that given the set of candidate models, the best model will be selected 2 times before any of these two models is selected. It is clearly seen that model 2 is nested in model 6. This suggests that the rank of a model might change if the analyst takes another set of independent samples of identical size from the model combinations of  $2^{13}$  models. In other words, there would be a high degree of uncertainty regarding the best model. Akaike weights are also useful to give a measure of the relative importance for model comparison.

### 8.1.3 Bayesian Model Average analysis for Millet

Table 8.4 presents the posterior inclusion probabilities (PIP) of individual variables, expected value and standard deviation and t-test of significance for each variable in the predictors for millet yields. R code and detailed output of the analysis are presented in Appendix F.3.

The t-test indicated that the variable farm size had a significant negative effect on millet production at 1% probability; this corresponds to Bayesian estimates with a highest inclusion probability 0.89. The three soil supplement categories and access to credit were significant at 5% probability level although these variables (except organic soil supplement) had negative

effect on the millet yield. The combinations of these significant predictors constitute most of the top 10 models listed in Table 8.4. Hence BMA had been able to identify potential variables for prediction of millet yields.

**Table 8.3** Summary estimates and inclusion Probabilities using BMA and p-value on Millet

Predictors	BMA			p-value
	Mean ( $\beta D$ )	Standard deviation ( $\beta D$ )	PIP $p(\beta \neq 0 D)$	Pr(> t )
Int.	2033.2	394.1	1.00	<0.001***
Farm Size	-233.2	122.4	0.89	0.004***
Fertilizer	DROPPED			0.418
Seed QTY	-3.49	6.736	0.28	0.052
SOIL1	657	616	0.63	0.041**
SOIL2	-346.3	324.8	0.64	0.021**
SOIL3	-234.1	321.3	0.42	0.030**
EDU1	DROPPED			
EDU2	-17.3	84.5	0.07	0.689
EDU3	DROPPED			
EDU4	118	277	0.02	0.069
Exp.	0.44 5	2.4	0.05	0.099
Ext.	DROPPED			
Credit	-42	141	0.12	0.038**

\*\*\* Significant at 1%, \*\* significant at 5%,  $p(\beta \neq 0|D)$ = posterior inclusion probability,

Int.=Intercept.

The number of explanatory variables is 13; the model space was small enough to allow for full enumeration of model space.

By applying BMA methods on millet yield (*kg/ha*) as dependent variable, 44 models were selected by OCCAM window, and the five top models had cumulative posterior probability of 0.263. The 5 top models are written out below and include the intercepts with their BIC values. R code and detailed analysis is provided in Appendix F.3

- $M_1: y = 2242.4 - 228.0 * x_1 + (\beta_0 - 683.4) * x_5 + (\beta_0 - 602.0) * x_6$

$$R^2 = 20.3\%, \text{ BIC} = -8.145, \text{ post.prob} = 0.095$$

- $M_2: y = 1722.3 - 256.7 * x_1 + (\beta_0 + 1191.6) * x_4$

$$R^2 = 16.3\%, \text{ BIC} = -8.01, \text{ post.prob} = 0.089.$$

- $M_3: y = 1877.7 - 232.5 * x_1 + (\beta_0 + 985.5) * x_4 + (\beta_0 - 307.7.0) * x_5$

$$R^2 = 19.9\%, \text{ BIC} = -7.59, \text{ post.prob} = 0.072$$

- $M_4: y = 2215.5 - 319.2 * x_1 - 13.28 * x_3 + (\beta_0 + 1162.4) * x_4$

$$R^2 = 19.8\%, \text{ BIC} = -7.47, \text{ post.prob} = 0.068$$

- $M_5: y = 2110 - 368.9 x_1 + 742.2 x_4 - 552.7 x_5 - 471.1 x_6$

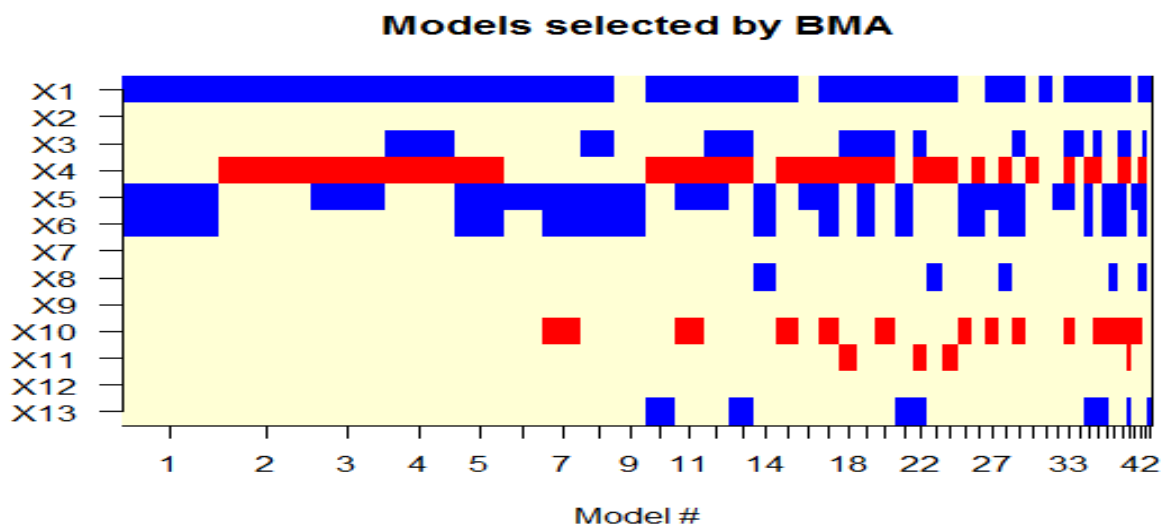
$$y = 2110 - 227.29x_1 + 2852.17x_4 + 1557.34x_5 + 1639.0x_6$$

$$R^2 = 22.9\%, \text{ BIC} = -6.76, \text{ post.prob} = 0.048$$

Using the criterion of highest posterior probability,  $\text{prob} = 0.095$ ,  $M_1$  was the best model for prediction of millet yield. Thus, the best model for millet yield is  $y = 2242.4 - 228.0 x_1 - 683.4 x_5 - 602.0 x_6$  and  $\text{BIC} = -8.145$ .

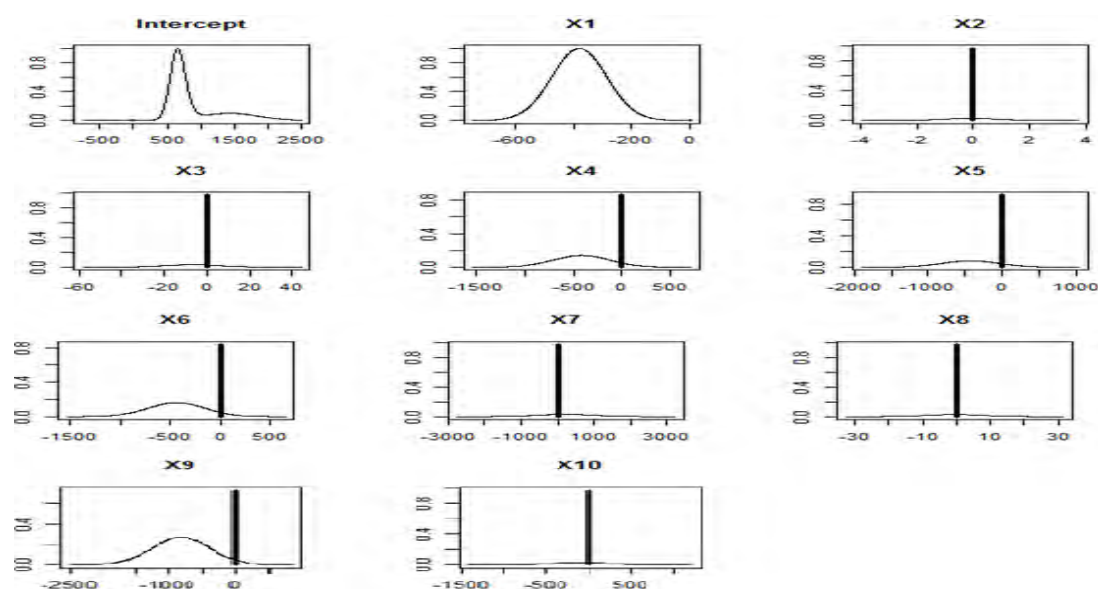
From Table 8.4, It can be observed that farm size had a posterior inclusion probability of 0.89. The organic soil supplement  $x_4$  had inclusion probability of 0.629 with a positive coefficient. Inorganic supplement  $x_5$  and integrated soil supplement  $x_6$  also showed inclusion probabilities of 0.639 and 0.415 respectively, but both had negative coefficients, thus one would prefer to select  $x_4$  for inclusion to be included in the model. It is worthwhile to note that soil management strategies were key determinants in the millet production.

The image plot of the all generated models and the associated predictors on millet yield ( $kg/ha$ ) is shown in Figure 8.3.



**Figure 8.3:** BMA image plot of selected models millet yield (kg/ha) and associated predictors

The five top models had a cumulative posterior probability of 0.3711. The corresponding plots of individual distribution for the predictor variable is displayed in Figure 8.4



**Figure 8.4:** Density plot of predictor variables on Millet yield using BMA model.

From Figure 8.4, it can be seen again that those variables which did not contribute significantly to the response variable are shown with spikes at zero in their density plots; indicating that their coefficients were most likely zero. It can be recognized that farm size,  $x_1$  was a substantial factor; so it would be included in most models.  $x_3$  and  $x_9$  contributed moderately to yield,  $x_2$ ,  $x_4$  and  $x_6$  were less substantial, while the density plots of the remaining variables ( $x_5$ ,  $x_7$ ,  $x_{10}$ ,  $x_{11}$ ,  $x_{12}$ ,  $x_{13}$ ) are almost flat showing they made no significant contribution to the response variable of interest. Thus,  $x_1$ ,  $x_3$ ,  $x_4$ ,  $x_6$  and  $x_9$  can be identified as variables for predicting millet production.

It can be seen that the density plots of the predictors with low posterior probability are almost flattened. The BMA can be used to address overfitting by excluding redundant variables from the model.

BMA identified the first 10 models that had higher posterior probabilities are presented in column 2 of Table 8.6. The individual AIC value computed for each model and the difference, Delta AIC ( $\Delta_i$ ). Akaike weight is relative probability of the model with respect to the set of candidate model given data. The model with highest Akaike weight is selected as the best model relative to candidate models given the data. This is empirical obtained given the selected candidate models. This means that there is an evidence of uncertainty in rank of these models in the model selection, however if another of set of candidate models has been selected, their position would have changed.

**Table 8.4:** AIC Weights Estimates for Model Comparison on Using on Millet

Model $m$	Predictors	$p(m y, X)$	AIC	$\Delta_i$	$l(m y)$	$w_i$
1	$x_1, x_5, x_6$	0.068	1512.23	1.2081	0.5466	0.1368
2	$x_1, x_4$	0.063	1515.39	4.3604	0.1130	0.0283
3	$x_1, x_4, x_5$	0.051	1512.81	1.7812	0.4104	0.1027
4	$x_1, x_3, x_4$	0.048	1513.65	2.6233	0.2694	0.0674
5	$x_1, x_4, x_5, x_6$	0.034	1511.03	0.0000	1.0000	0.2503
6	$x_1, x_5$	0.027	1516.71	5.6798	0.0584	0.0146
7	$x_1, x_5, x_6, x_{10}$	0.026	1511.36	0.3331	0.8466	0.2119
8	$x_1, x_3, x_5, x_6$	0.023	1512.15	1.1271	0.5692	0.1425
9	$x_5, x_6$	0.021	1516.81	5.7829	0.0555	0.0139
10	$x_1, x_4, x_{13}$	0.021	1515.18	4.1517	0.1254	0.0314

$p(m|y, X)$  is the posterior model probability from BMA

AIC- Akaike information criterion

Delta AIC =  $\Delta_i = \text{AIC} - \text{AIC}_{\min}$  is the information loss

$l(m|y) = \exp(-0.5 * \Delta)$  is the likelihood of the model given the data e.g.  $\exp(-$

$0.5 * 1.2082) = 0.5466$

$$w_i = \frac{\exp\left(-\frac{1}{2} \Delta AIC_i\right)}{\sum_{r=1}^R \exp\left(-\frac{1}{2} \Delta AIC_r\right)} \quad (8.1)$$

$w_i$  is the Akaike weight or relative probability for model  $i$ , which can be interpreted as a relative posterior probability of the model given the data.

The weight  $w_i$  is interpreted as the measure of evidence. Model  $m_5$  would be chosen as the best approximating model given the data and the set of candidate models. Alternatively, the Akaike weight  $w_i$  can be interpreted as the relative posterior probability for each model given the data. Model  $m_5$  had a highest Akaike weight of 0.2503, and thus,  $m_5$  would be selected as the best model given the data and set of candidate models. That is the model that entails farm size, organic manure and soil supplement regime. It should be noted that the likelihood value of each model is likelihood relative to the best model given the data.

The Akaike weights serve to focus the evidence for or against the various models. The importance of carefully defining a small set of candidate models, based on the objective and what is known about the problem, cannot be overemphasized. Initially, we had 13 explanatory variables; we are not expected to learn much from the data. The BMA approach demands a substantial supporting algorithm to help reducing the number of models to be considered. In this case, there would be  $2^{13}$  models (many more if transformations or interaction terms were allowed) and overfitting would surely be a risk.

## 8.2 Models selection by Zellner's g-prior

The distinction between model average and model selection is important for celerity of the concept. Wasserman (2000) gives a clear definition of the two analytical processes. In simple terms, model selection refers to the problem of using the data to select one model from the list of candidate models  $M_1, M_2, \dots, M_k$ . Model averaging on the other hand refers to the process of estimating some quantity under each model  $M_j$  and then averaging the estimates according to how likely each model is to be selected.

Thus, if one identifies the potential significant predictors for response output of interest then it would save computation time and reduce the complexity of estimating  $2^k$  model combinations arisen from BMA. The analysis here is based on Bayesian Model Selection using Zellner's  $g$ -prior discussed in section 4.2.2. One of the crucial points in its application in Bayesian variable selection is in the determination of prior density for the regression coefficients and constant  $c$ .

### 8.2.1 Bayesian Model Selection for predicting Sorghum yield

In this section, we used Zellner's  $g$ -prior with constant  $c=100$  for models selection. The predictors involved are  $X_\gamma = (x_1, x_3, x_6, x_{10}, x_{13})$ , which takes value  $\gamma = 1$ , if  $x_\gamma$  is included in the model and zero otherwise. The analysis generates  $2^6 = 64$  model combinations with their posterior probabilities and associated  $-2\log$  likelihood. The predictors considered are based on the coefficients of the predictor variables being far from zero.

Fifteen model combinations out of 64 models are presented in Table 8.5. R code that generated the results is provided in Appendix G.1 and submodels presented on Table 8.5.

**Table 8.5:** Summary Estimates of  $-2 \log$  likelihood and marginal posterior probability for Model Selection using Zellner's  $g$ -prior with constant  $c=100$  on Sorghum yields

Model $m$	x1	x2	x3	X6	X10	x13	Log $p(y X, m)$	Post.Prob $p(m y, X)$
1	0	0	0	0	0	0	-1225.49	0.0000
2	1	0	0	0	0	0	-1205.51	0.2475
4	1	1	0	0	0	0	-1206.43	0.0099
6	1	0	1	0	0	0	-1205.46	0.2622
8	1	1	1	0	0	0	-1206.85	0.0653
10	1	0	0	1	0	0	-1207.21	0.0452
12	1	1	0	1	0	0	-1208.44	0.0133
14	1	0	1	1	0	0	-1207.13	0.0493
16	1	1	1	1	0	0	-1208.77	0.0096
26	1	0	0	1	1	0	-1209.26	0.0059
30	1	0	1	1	1	0	-1208.93	0.0081
32	1	1	1	1	1	0	-1210.61	0.0015
34	1	0	0	0	0	1	-1207.73	0.0270
38	1	0	1	0	0	1	-1207.58	0.0314
56	1	1	1	0	1	1	-1210.97	0.0011
62	1	0	1	1	1	1	-1211.05	0.0010

To implement Zellner  $g$ -prior approach,  $\beta_0$  assumed zero and  $c = 100$  was chosen to be a large value reflecting the vague belief representing the uncertainty associated with the model selection. The prior on  $(\beta, \beta_0)$  for this full model is  $N(\beta_0 + \hat{\beta}X, c\sigma^2(X'X)^{-1}) \times 1/\sigma^2$ .

Then for any submodel defined by reduced design matrix  $X_\gamma$ , we take the prior on  $(\beta, \beta_0)$  to be  $N(\beta_0 + \hat{\beta}X, c\sigma^2(X'_\gamma X_\gamma)^{-1}) \times 1/\sigma^2$ .

The model considered best is the one that maximizes the log likelihood and has highest posterior probability. From Table 8.5, it can be seen that the model having the highest posterior probability is  $X_\gamma = (1, 0, 1, 0, 0, 0)$ . This is the optimal model for sorghum yield ( $kg/ha$ ) as a function of the farm size and the quantity of seed sown. The model had posterior probability of 0.2622 and log-likelihood of -1205.46 and so was considered best for predicting the sorghum yield production.

### 8.2.2 Bayesian Model Selection for predicting Millet yield

Table 8.6 was first used to identify predictors with higher inclusion probability to be considered to build a single optimal model. Predictor variables ( $x_1, x_3, x_4, x_5, x_6$ ) were considered based on their computed inclusion probabilities. Then 32 models could be generated from the combinations of these 5 predictors. Then posterior model probability and -log likelihood under Zellner's' g-prior with constant  $c = 100$  for the model were computed.

Table 8.6 presents the result of model design. The model consists of the five predictors ( $x_1, x_3, x_4, x_5, x_6$ ). The R code and outputs are presented in Appendix G.

The procedure utilized constant zero beta coefficient,  $c=100$  and predictors  $X_\gamma = (x_1, x_3, x_4, x_5, x_6)$  takes value  $\gamma = 1$ , if  $x_\gamma$  is included in the model and zero otherwise.

The process generated  $2^5 = 32$  combinations of predictors with their posterior probabilities and corresponding -2log likelihood.

**Table 8.6:** Summary Estimates of -2 log likelihood and marginal posterior probability for Model Selection using Zellner's' g-prior with constant  $c=100$  on Millet

Model (m)	x1	x3	x4	x5	x6	Log $p(y X, m)$	Post.Prob $p(m y, X)$
1	0	0	0	0	0	-767.37	0.0037
2	0	0	0	0	1	-765.69	0.0199
6	0	0	1	0	1	-763.80	0.1315
7	0	0	1	1	0	-767.46	0.0034
8	0	0	1	1	1	-764.34	0.0768
9	0	1	0	0	0	-765.28	0.0300
10	0	1	0	0	1	-764.44	0.0697
11	0	1	0	0	0	-767.58	0.0030
12	0	1	0	0	1	-765.90	0.0162
13	0	1	1	1	0	-764.87	0.0450
22	1	0	1	0	1	-766.13	0.0128
25	1	1	0	0	0	-764.49	0.0663
26	1	1	0	0	1	-763.66	0.1517

27	1	1	0	1	0	-766.77	0.0068
28	1	1	0	1	1	-765.00	0.0397
29	1	1	1	0	0	-765.36	0.0277
30	1	1	1	0	1	-764.46	0.0682
31	1	1	1	1	0	-767.63	0.0029
32	1	1	1	1	1	-765.72	0.0194

From Table 8.6, it can be observed that the optimal model is  $X_\gamma = (1, 1, 0, 0, 1)$ . The linear combination of farm size, quantity of seed sown, and integrated soil supplements are included in the model. That is, the model included variables  $x_1$ ,  $x_3$  and  $x_6$ . This model had highest posterior probability of 0.1517 and log likelihood of -763.66 and so was considered best for predicting millet production.

It can be seen that the selected model “highlighted green” is optimal. That model consists of 3 terms for millet production: farm size, seed quantity and integrated soil supplements. It seems then that, farm size, quantity of seed sow and integrated soil supplement are important predictions for the production of millet yield.

The integrated soil supplement is a farmer strategy that involves ploughing with animal manure at the land preparation stage and later applying NPK fertilizer during crop growth development to enhance the crop yield. Referencing different group intercept for no soil supplement  $\beta_0 = 2033.2$ , then  $\beta_0 + \beta_{4(3)}X_6 = 2033.2 + (-234.1)(1) = 1799.1$  (refer to Table 8.5 for the coefficient contribution). Thus, the optimal model for millet yield can be written as  $y = 2033 - 233.2x_1 + 1799.1x_3 + 1999.1x_6$ . This means that a farmer, who adopts an integrated soil supplement, would harvest 1999.1 kg/ha more than a fellow farmer without additional soil supplements on his millet field.

### 8.3 Summary and Discussion

In this chapter, we have applied BMA to address the problem of model uncertainty by selecting the best candidate models among the combinations of predictors. BMA analysis allows the posterior probability of the true model to be computed. When faced with several model combinations, one may either choose a single model or averages over all the models. Bayesian model average methods provide a set of tools for these problems. Bayesian methods also give us a numerical measure of the relative evidence in favour of competing theories.

The main point in this chapter has been the Bayesian model selection and model averaging in a conceptually simple and unified approach. For models that are moderate in terms of their

posterior probability, AIC weight provides a useful approximation to the log Bayes factor as established in chapter 4. Choosing one model may be risky for prediction with regard to fitting a regression function. It is possible to average the predictions from several models and decide the final optimal model.

Reviewing the analysis carried out in this chapter, this approach is well suited to address the uncertainty in model structure and covariate selection. Under a traditional approach to data analysis, researchers would consider fitting different model structures, examining the model fits criterion such as R-square, AIC, BIC and drawing inferences from the final models, thus ignoring alternative estimates and uncertainty arising from the model selection process and other models. Bayesian variable selection can still be meaningful in a decision-theoretic sense, where the task is to select the model with the best predictive performance. In many situations, the model selection process can be improved using a modified Bayes factor that is similar to a cross-validation criterion. Furthermore, approximation of AIC weight for model comparison can be found Burnham and Anderson (2002), even for non nested competing models.

In this chapter, we also applied Zellner g-prior to variable selection. The result yielded a single optimal model. Bayesian variable selection addresses model covariates uncertainty given a particular model structure; which was a linear combination of predictors in this case. The work investigated uncertainty in analytical process in term of estimation of the model coefficients and determining the model probabilities.

Model uncertainty arising from covariates selection was addressed by comparing two strategies for implementing model averaging where the approach averaged over all possible model combinations. The method computed Bayesian information criterion as a benchmark for model selection as an AIC approximation to the log of Bayes factor. There are other approaches to Bayesian model selection that encourage the use of prior distributions, which consequently allows the elements of regression coefficients to be forced to zero. The work reported by George and McCulloch (1993) used Gibbs sampling to do model selection and predictive least quasi-deviance (Quine et al. 1996).

## 9. Conclusions and Suggestions

This thesis investigated influences of different uncertainty model structures while building models in two conceptual settings through standard probability theory and uncertainty statistics. The chapter gives a concluding remark on uncertainty modelling of agricultural data.

### 9.1 Conclusions

The goal of this thesis was to construct a predictive model for crop yield prediction in presence of uncertainty in the model building process. Both a model based approach (classical model) and knowledge based approach (uncertainty theory) were explored.

With the development of modern data collection approaches, researchers may collect hundreds to millions of variables, yet may not need to utilize all explanatory variables available in predictive models. Hence, choosing models that entails a subset of potential variables often becomes a crucial step. In a linear regression, variable selection will not only reduce to a parsimonious model, but also prevents over-fitting. From a Bayesian perspective, prior specification for model parameters plays an important role in model selection as well as parameter estimation, and often prevents over-fitting by removal of redundant variable and model averaging. The use of expert knowledge as an additional source of information has been successfully utilized in many scientific inquiring, particular where there are insufficient data. This analysis of split data would be very useful, when a farmer is adopting a new farming technology or cultivating at new farming location, but there are no sufficient information.

### 9.2 Suggestion for Further Research

Bayesian Model Average (BMA) discussed in this thesis can be extended to involve generalized linear models and survival functions. The process of determining the prior in regression estimation and selection should be investigated further especially where data contains sparse matrix, and correlation among the regression predictor. Stochastic variable selection and other subset selection can be employed to analyse high dimension like quantitative trait loci mapping and gene identification in microarray data (e.g. cancer problem Lee et al. Marina Vannucci, Bioinformatics, 2003). Further research into Bayesian variable

selection might offer ground breaking results in areas where the number of predictors is larger than the sample size.

There is a need for further investigation on influence of the prior sensitivity in generalized linear models. Especially, there is a need to carry out more research on the specification and elicitation of prior belief in distributions with multiple parameter setting. For example, different distributional forms of Weibull density are good examples. The idea of vague or non information prior employed in this work can be extended to Generalized Linear Models (GLMs) or through a non-conjugate prior from expert and model averaging.

We have also suggested that other distributions such as Gamma, log-normal, Weibull for crop yields as a remedy for negative prediction in this study. This would tackle the skewness nature of crop yield density as reported in many empirical studies.

For GLMs, there is a need to further investigate the mixtures of Zellner's g-priors under different scenarios while still maintaining good distributional properties such as tractable (approximate) marginal and asymptotic consistency in model selection. In the literature, a potential problem of g- prior identified is that it inherits in the instability of ordinary least squares (OLS) estimates when predictors are highly correlated. Marin and Robert (2007) suggested that the model could be built with a hierarchical prior based on scale mixtures of independent normal, which incorporates invariance under rotations within models like ridge regression and the g-prior, but has heavy tails like the Zellner-Siow Cauchy prior.

Uncertainty theory is applicable when belief degrees are available. In order to rationally deal with belief degrees, an uncertainty theory was developed by Liu (2007). When no sample is available, we need to invite some domain experts so as to evaluate the belief degree function about indeterminate quantities. It has been reported that the belief degree function has much larger variance than the long run cumulative frequency and hence the probability theory may no longer be applicable. Further research need to be conducted in this area of multivariate aspect and applicability of the uncertainty statistics. This work is not exhaustive but it has hopefully opened new area for further researches. The criteria for goodness of fit for constructing uncertainty distributions need attention. The distance measure between two uncertain variables is used to compare the difference (divergence) between a model and its data realization. More research can be opened in the aspect of the uncertainty distance measure similar to Kullback-Leibler divergence.

A variety of methods for specifying opinions have been developed for unidimensional prior distributions in the literature. However, practical statistical models invariably contain many unknown parameters, and it is usually important to obtain the expert's information about several parameters.

Until now, there are few methods for specifying a multivariate prior distribution, most of these methods are only applicable to specific classes of problems, or rely on restrictive conditions such as independence of variables, or else require the elicitation of variances and covariance. Kadane and Wolfson, (1998) noted that most experts interrogated are not generally able to specify second order moments reliably. This is a problem that needs to be addressed in the future and a more flexible method of elicitation for multivariate prior distributions should be developed that is applicable to a wide class of practical problems.

# Bibliography

- [1] Adeoti, J. O., & Olubamiwa, O. (2009). Towards an innovation system in the traditional sector: the case of the Nigerian cocoa industry. *Science and Public Policy*, 36(1), 15-31.
- [2] Akaike, H. (1973). Information Theory and an Extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *2nd International Symposium on Information Theory* .(pp. 267-281).Akad. Kiado, Budapest.
- [3] Albert, J. (2008). *Bayesian computation with R* (Vol. 747389981). New York: Springer
- [4] Bellow, M. E. (2007). *Comparison of Methods for Estimating Crop Yield at the County Level*.
- [5] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag, 2<sup>nd</sup> Edition.
- [6] Berger J. (2006). The Case for Objective Bayesian Analysis, *Bayesian Analysis* vol. 1, Number 3, pp. 385-402.
- [7] Bishop C.M. (2005). *Neural Networks for Pattern Recognition*. Oxford University Press ISBN 0-19-853864-2.
- [8] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- [9] Bozdogan, H. (2000). Akaike's Information Criterion and recent developments in information complexity. *Journal of mathematical psychology*, 44(1), 62-91.
- [10] Bremen, H., Groot, J. & van Keulen, H. (2001). Resource limitations in Sahelian Agriculture. *Global Environmental Change* 11: 59-68.
- [11] Brooks, S. P., & Roberts, G. O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, 8(4), 319-335.
- [12] Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- [13] Carlin, B. P. and Louiss. T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC.
- [14] CBN (2006) Statistical Bulletin of the Central Bank of Nigeria, Abuja.
- [15] CBN (2007) Statistical Bulletin of the Central Bank of Nigeria, Abuja.
- [16] Chandrahas, A. and Rai, T. (2001). Pilot study for developing Bayesian probability forecast model based on farmers' appraisal data on wheat crop. Project Report, IASRI, New Delhi.

- [17] Chipman, H., George, E. I., McCulloch, R. E., Clyde, M., Foster, D. P., & Stine, R. A. (2001). The practical implementation of Bayesian model selection. *Lecture Notes-Monograph Series*, 65-134.
- [18] Clyde, M., & George, E. I. (2004). Model uncertainty. *Statistical science*, 81-94.
- [19] Cooke R.M. (1991). Experts in Uncertainty-Opinion And Subjective Probability in Science
- [20] Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 883-904.
- [21] Day R.H. (1965). Probability Distributions of the Field crop Yields. *Jornal of Farm Economics*. Vol. 47, No.3:713 – 741.
- [22] Draper N.R. and H. Smith (1998). Applied Regression Analysis in Probability and Statistics, Wiley Series.
- [23] Efron, B. & Tibshirani, R. (1997). Improvements on cross-validation: The bootstrap method. *Journal of the American Statistical Association*, 92(438), 548 - 560.
- [24] Eugene, N., Lee, C. and Famoye F. (2002). Beta-normal distribution and its applications, *Commun. Stat.-Theory Methods* 31 pp. 497–512.
- [25] Fahrmeir, L., & Kneib, T. (2009). Propriety of posteriors in structured additive regression models: Theory and empirical evidence. *Journal of Statistical Planning and Inference*, 139(3), 843-859.
- [26] Feldkircher, M. and S. Zeugner (2011) Bayesian Model Averaging with BMS," R-package vs. 0.3.0, URL: <http://cran.r-project.org/web/packages/BMS/>
- [27] Fernandez, C., Ley, E., & Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100(2), 381-427.
- [28] Fridley, B. L. (2009). Bayesian Variable and model selection methods for genetic association studies. *Genetic Epidemiology*, 33:27–37.
- [29] Gamerman, D. (1997). Markov chain Monte Carlo: stochastic simulation for Bayesian inference.
- [30] Gao X. (2009). Some properties of continuous uncertain measure, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Vol.17, No.3, 419-426.
- [31] Gelman, A., J.B. Carlin, H.S. Stern, H. and D.B Rubin (2004), Bayesian Data Analysis, New York: Chapman & Hall/CRC.
- [32] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3), 515-534.
- [33] George, E.I. and R. E.McCulloch. (1993). Variable Selection via Gibbs Sampling. *Journal of American Statistical Association*, 88 (423).

- [34] George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica sinica*, 7(2), 339-373.
- [35] Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Vol. 196). *Federal Reserve Bank of Minneapolis*, Research Department.
- [36] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Vol. 196). In J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith (eds.), *Bayesian Statistics 4*, 169–193, Oxford University Press, Oxford, U.K..
- [37] Guo, R., Cui, Y.H. and Guo, D. (2012). Uncertainty Statistics. *Journal of Uncertain Systems*. Vol.6, No.3: 163-185. <http://www.jus.org.uk>.
- [38] Guo, R., D. Guo, and Thiart, C. (2010). Liu's uncertainty Normal Distribution. *Proceedings of the First International Conference on Uncertainty Theory*. Urumchi and Kashi, China, 191-207.
- [39] Guo, R., Cui, Y.H. and Guo, D. (2010). Uncertainty Decision Theory. *Proceedings of the First International Conference on Uncertainty Theory*, Urumchi, China, 5-14.
- [40] Harrell, F. E. (2001). *Regression modelling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag, Inc.
- [41] Hocking, R.R.. (1972). Criteria for selection of a subset regression: which one should be used, *Technometrics*. 14, 967-970.
- [42] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401.
- [43] Jeffreys, H. (1961). *Theory of Probability*, 3rd Ed., London.: Oxford University Press.
- [44] Johnson, N., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions Volume 2*, 2nd ed. Wiley: New York.
- [45] Jones, M.C. (2004). Families of distributions arising from distributions of order Statistics. *Test*. 13, 1–43.
- [46] Jones, P. G., & Thornton, P. K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global environmental change*, 13(1), 51-59.
- [47] Just, R. E., & Weninger, Q. (1999). Are crop yields normally distributed?. *American Journal of Agricultural Economics*, 81(2), 287-304.
- [48] Kadane, J., & Wolfson, L. J. (1998). Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 3-19.
- [49] Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.

- [50] Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Statist. Assoc.*, 90, 928-934.
- [51] Kass, R. E., and Wasserman L. (1996). The Selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* vol.91, 1343-1370.
- [52] Kolmogorov, A.N. (1950). *Foundations of the Theory of Probability*, Chelsea, New York.
- [53] Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, 128(4), 224-233.
- [54] Kuhnert, P. M., Martin, T. G., Mengersen, K., & Possingham, H. P. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics*, 16(7), 717-747.
- [55] Kuha, J. (2004). AIC and BIC comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2), 188-229.
- [56] Lambert, P.C. and Sutton, A.J. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distribution in MCMC using WinBUGS. *Statistics in Medicine*, 24 2401-2428.
- [57] Lee, P. M. 1997. *Bayesian statistics: An introduction*, 2nd ed. Arnold, London
- [58] Liang, F., Paulo, R., Molina, G., Clyde, M., and Berger, J. (2008), "Mixtures of g-priors for Bayesian variable selection," *Journal of the American Statistical Association*, 103, 410- 423.
- [59] Liu, H., Weiss, R. E., Jennrich, R. I., & Wenger, N. S. (1999). PRESS model selection in repeated measures data. *Computational statistics & data analysis*, 30(2), 169-184.
- [60] Liu, B.D.(2007). *Uncertainty Theory*, 2nd Edition, Springer-Verlag, Heidelberg, Berlin.
- [61] Liu, B.D. (2011). *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*, Springer-Verlag, Berlin.
- [62] Liu, B.D. (2009). Some research problems in uncertainty theory. *Journal of Uncertain Systems*. vol.3, no.1: 3-10.
- [63] Liu, B.D. (2010). Uncertain risk analysis and Uncertain reliability analysis. *Journal of Uncertain Systems*. vol.4, No.3: pp.163-170.
- [64] Liu, B.D. (2011). Uncertain logic for modelling human language, *Journal of Uncertain Systems*. vol.5, no.1, 3-20.
- [65] Liu, Y.H. and Ha, M. H. (2010). Expected value of function of Uncertain variables, *Journal of Uncertain Systems*, Vol.4, No.3, 181-186.
- [66] Liu, W., & Xu, J. (2010). Some properties on expected value operator for uncertain variables. *Information: An International Interdisciplinary Journal*, 13(5), 1693-1699. Available at [www. http://orsc.edu.cn/online/090611.pdf](http://www.orsc.edu.cn/online/090611.pdf)

- [67] Li X. and Liu, B. (2009). Chance measure for hybrid events with fuzziness and randomness, *Soft Computing*, Vol.13, No.2, 105-115.
- [68] Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. London: Sage Publications.
- [69] Ludwig, B. G. (1994). Internationalizing Extension: An exploration of the characteristics evident in a state university Extension system that achieves internationalization. Unpublished Ph.D dissertation, The Ohio State University, Columbus.
- [70] Madigan, D. and Raftery, A.E. (1994), Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Statist. Assoc*, 89 pp. 1535–1546.
- [71] Marin, J. and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer-Verlag, New York.
- [72] Martin, A. and Quinn, K. (2006). *Applied Bayesian Inference in R Using MCMCpack*. Available at :<http://mcmcpack.wustl.edu>.
- [73] Martin T.G., Kuhnert, P.M., Mengersen, K. and Possingham, H.P. (2005). The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. *Ecological Applications* **15**: 266-280.
- [74] Masterson, T. & Rao, J.M. (1999). *Agricultural Productivity in Paraguay: Analysis and Proposals For economic Growth Policies*. Technical report, Ministerio de Hacienda de la Republica del Paraguay and United Nations Development Programme.
- [75] McDaniels, T.L. (1995) Using Judgement in Resource Management: A Multiple Objective Analysis of a Fisheries Management Decision. *Operations Research* 43 415-426.
- [76] Montgomery, D.C. and Peck, E.A. (1982). *Introduction to linear regression analysis*. Wiley, New York.
- [77] Nmadu J. N., Eze, G. P., and Jirgi, A. J. (2012). Determinants of Risk Status of Small Scale Farmers in Niger State, Nigeria, *British Journal of Economics, Management & Trade*. 2(2): 92-108.
- [78] Ogallo, L., Bessemoulin, P., Ceron, J. P., Mason, S., & Connor, S. J. (2008). Adapting to climate variability and change: the Climate Outlook Forum process. *Bulletin of the World Meteorological Organization*, 57(2), 93-102.
- [79] O'Hagan, A., C.E. Buck, A. Daneshkhah, R. Eiser, P. Garthwaite P, D.J. Jenkinson, J. Oakley J and T. Rakow ( 2006). *Uncertain Judgements Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd, Chichester.
- [80] Ozaki, V. A., Goodwin, B. K., & Shiota, R. (2008). Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40(9), 1151-1164.
- [81] Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681-686.

- [82] Peng, Z.X., and Iwamura, K. () A sufficient and necessary condition of uncertainty distribution, *Journal of Interdisciplinary Mathematics*. (to appear).
- [83] Persson, T., Ortiz, B. V., Bransby, D. I., Wu, W., & Hoogenboom, G. (2011). Determining the impact of climate and soil variability on switchgrass (*Panicum virgatum* L.) production in the south-eastern USA; a simulation study. *Biofuels, Bioproducts and Biorefining*, 5(5), 505-518.
- [84] Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437), 179-191.
- [85] Rao, J.N.K. (2003). Small Area Estimation. Wiley Series in Survey Methodology: ISBN- 0-471-41374
- [86] Rao, J. N. K. (2004). Small area estimation with applications to Agriculture. *Agricultural survey methods*, 9, 139-147.
- [87] R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [88] Ricci V. (2005). Fitting Distributions with R , Release 0.4-21 February 2005.
- [89] Roosen, J. and Hennessy D.A. (2001) Capturing Experts' Uncertainty in Welfare Analysis: An application to Organophosphate Use Regulation in U.S. Apple Production. American
- [90] Sanchez, P. E. (2002). Soil Fertility and Hunger in Africa, *Science*, 295:2019–2020.
- [91] Sanchez, P. A. & Jama, B.A. (2002). Soil Fertility replenishment takes off in the East Africa. In: Vanlauwe, B. , Diels, J., Sanginga , N., Merckx, R.(Eds). *Integrated Plant Nutrient Management in Sub Saharan African*. CABI Publishing , Oxon, UK. Pp. 23-45.
- [91] Seber, G. A. F. (1977) *Linear Regression Analysis*. Wiley, New York, 1977.
- [92] Stewart. G. W. (1977) *Introduction to Matrix Computations*. Academic Press, New York, 1973.
- [93] Shannon R. B., Mohammad, T. K., Sittichai K. and Byung, R.Ch. (2009). A logistic approximation to the cumulative normal distribution. *Journal of Industrial Engineering and Management*, 2(1): 114-127, ISSN: 2013-0953 available at doi:10.3926/jiem.2009.v2n1.p114-127.
- [94] Smith, J., & Mandac, A. M. (1995). Subjective versus objective yield distributions as measures of production risk. *American Journal of Agricultural Economics*, 77(1), 152-161.
- [95] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Linde, A.V. (2003). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4), 583-616.

- [96] Stasny, E., Goel, P., Cooley, C. and Bohn, L. (1995). Modeling County-Level Crop Yield with Spatial Correlations Among Neighboring Counties”, Technical Report No. 570, Department of Statistics, The Ohio State University.
- [97] USDA. (1998). *Crop Progress/Crop Weather: Terms and Definitions.* National Agricultural Statistics Service {NASS}, Washington
- [98] Valeeva, N. I., Meuwissen, M. P., Bergevoet, R. H., Lansink, A. G. O., & Huirne, R. B. (2005). Improving food safety at the dairy farm level: farmers' and experts' perceptions. *Applied Economic Perspectives and Policy*, 27(4), 574-592.
- [99] Walter, G., & Augustin, T. (2010). Bayesian Linear Regression-Different Conjugate Models and Their (In) Sensitivity to Prior-Data Conflict. In *Statistical Modelling and Regression Structures* (pp. 59-78). Physica-Verlag HD.
- [100] Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, 11(1), 192-196.
- [101] Wang, X.S. (2010). Moment estimation of uncertainty distribution, Proceedings of the Eighth Annual Conference on Uncertainty, 168-172.
- [102] Wang X.S, Z.C. Gao, and Guo H. Y.(2012). Delphi method for estimating uncertainty distributions, *Information: An International Interdisciplinary Journal*, Vol.15, No.2.
- [103] Xue, F., Tang, W.S. and Zhao R.Q. (2008), The expected value of a function of a fuzzy variable with a continuons membership function, *Computers and Mathematics with Applications*, vol.55, pp.1215–1224.
- [104] Yan, X. and Su, X.G. (2009). *Linear regression analysis: Theory and Computing*, World Scientific Publishing Co. Pte. Ltd. ISBN-10: 981-283-410-9.
- [105] Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, vol.1, 3-28.
- [106] Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distribution regression using Bayesian variable selection. In *Bayesian inference and decision techniques: Essays in Honour of Bruno De Finetti*, pages 233-243. North-Holland / Elsevier.

# Appendix A: Mathematical Derivations

## A.1 Jeffreys' Prior

Suppose we cannot easily find the natural scale on which the likelihood is in data-translated format, Jeffreys (1961) proposed a general prior in such cases, based on the Fisher information  $I$  of the likelihood. Recall that

$$I(\theta | x) = -E_x \left( \partial^2 \frac{\ln \ell(\theta | x)}{\partial \theta^2} \right) \quad (\text{a.1})$$

Jeffreys' stated that the prior takes

$$p(\theta) \propto \sqrt{I(\theta | x)} \quad (\text{a.2})$$

A full discussion, with derivation, can be found in Lee (1997, Section (3.3))

## A.2 Derivation of Posterior Distribution for a normal prior

To introduce the basic ideas of Bayesian analysis, consider the case where data are drawn from a normal distribution, so that the likelihood function for the  $i^{\text{th}}$  observation,  $x_i$  is

$$\ell(\mu, \sigma^2 | x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right) \quad (\text{a.3})$$

The resulting likelihood for all  $n$  data points is

$$\begin{aligned} \ell(\mu, \sigma^2 | \mathbf{x}) &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left( -\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 - 2\mu\bar{x} + n\mu^2 \right) \end{aligned} \quad (\text{a.4})$$

### Known Variance and Unknown Mean

Assume the variance ( $\sigma^2$ ) is known, while the mean  $\mu$  is unknown. For a Bayesian analysis, one need to specify the prior for  $\mu$ , say  $p(\mu)$ . Suppose we assume a Gaussian prior,  $\mu \sim N(\mu_0, \sigma_0^2)$  so that

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left( -\frac{(x_i - \mu_0)^2}{2\sigma_0^2} \right) \quad (\text{a.5})$$

The mean and variance of the prior,  $\mu_0$  and  $\sigma_0^2$  are referred to as **hyperparameters**.

In calculating the posterior distribution, we ignore terms that are constants with respect to the unknown parameters. Suppose  $\mathbf{x}$  denotes the data and  $\Theta_1$  is a vector of a known model parameters, while  $\Theta_2$  is a vector of unknown parameters. Then, the posterior is written as

$$p(\Theta_2 | \mathbf{x}, \Theta_1) = f(\mathbf{x}, \Theta_1) \cdot g(\mathbf{x}, \Theta_1, \Theta_2) \quad (\text{a.6})$$

Then

$$p(\Theta_2 | \mathbf{x}, \Theta_1) \propto g(\mathbf{x}, \Theta_1, \Theta_2) \quad (\text{a.7})$$

Using the prior given in equation (a.5), we write the resulting posterior as

$$p(\mu | \mathbf{x}) \propto \ell(\mu | \mathbf{x}) \cdot p(\mu) \quad (\text{a.9})$$

$$\propto \exp\left[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2} - \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 - 2\mu\bar{x} + n\mu^2\right)\right] \quad (\text{a.10})$$

We can factor out additional terms not involving  $\mu$  to give

$$p(\mu | \mathbf{x}) \propto \exp\left[-\frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} + \frac{\mu n\bar{x}}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right] \quad (\text{a.11})$$

Factorizing in terms of  $\mu$ , the exponential becomes

$$\frac{-\mu^2}{2} \left(\frac{1}{2\sigma_0^2} - \frac{n}{2\sigma^2}\right) + \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) = -\frac{\mu^2}{\sigma_*^2} + \frac{2\mu\mu_*}{2\sigma_*^2} \quad (\text{a.12})$$

Where

$$\sigma_*^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{2\sigma^2}\right)^{-1} \quad \text{and} \quad \mu_* = \sigma_*^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) \quad (\text{a.12})$$

Finally, by completing the square, we have

$$p(\mu | \mathbf{x}) \propto \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2} + f(\mathbf{x}, \mu_0, \sigma^2, \sigma_0^2)\right) \quad (\text{a.13})$$

Thus, the posterior density function for  $\mu$  becomes

$$p(\mu | \mathbf{x}) \propto \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2}\right) \quad (\text{a.14})$$

Now, that the density function for  $z \sim N(\alpha, \beta)$  is

$$p(\mu | \mathbf{x}) \propto \exp\left(-\frac{(\mu - \alpha)^2}{2\beta^2}\right) \quad (\text{a.15})$$

This shows that the posterior density function for  $\mu$  is a normal with mean  $\mu_*$  and variance  $\sigma_*^2$  e.g.

$$\mu \sim N(\mu_*, \sigma_*^2) \quad (\text{a.16})$$

This is the posterior density is in the same form as the prior. This is known as the prior conjugated with the likelihood function.

# Appendix B: Expert Survey Questionnaire

## B.1: Instructions

1. This expert Survey Questionnaires differ from design Questionnaires used in Social Sciences and Agricultural Economics and Extension Services.
2. Respondent here is referred to as ``Expert". Experts must be a staff of Ministry of Agriculture, Agricultural Development Program (ADP), Research Institute or member of staff in School of Agriculture.
3. The information is sought regarding the farmers yield production of cereal crops in the last growing session (November 2012) in Niger State, Nigeria and the associated degree of belief, denoted by  $(x_i, \alpha_i)$  .
4. Each data point  $x_i$  represents an uncertain variable, which can be regarded as imprecise quantity such, ``About 50 kg", ``Approximately 30 tonne", ``Moderately good", `` high", ``low", `` better than previous year", and any appropriate description of the respondent view (using unrestricted language).
5. Assume that the real value is not exactly known to us. Thus we think the production level is an imprecise quantity.
6. The  $\alpha_i$  is the degree of belief or intuition level attached to  $i^{th}$  data point provided, which will be based on your knowledge or past experience. It can be stated in Percentage(1%, 20%, 50%, 90%, even 100% assured level) or top 10%, down 10%, perfectly assured (100%).
7. The respondent should provide a consistent in assessment, the percentage of belief must be increasing order, where first score has the least percentage attached and the last value reported must attract highest percentage of belief and maintain a monotonic function.

**PLEASE TURN TO NEXT PAGE**

## B.2: Expert Eliciting Information on Uncertain Variable

University of Cape Town, South Africa

### Department of Statistical Sciences

The information is sought from the small-scale farmers on level of production/productivity of some cereal crops in the last cropping session in Niger State, Nigeria. The required information will be used for research purpose in pursuance of a Master degree in Statistics in the above named institution. The information provided will be treated confidential and purely for the research purpose. We want to establish a model to get crop yields based expert knowledge and experience. Assume that the crop yield value is not exactly known to us. Thus we think the crop yield is an imprecise quantity. The consultation process of a questionnaire survey for estimating such process is of the following format.

#### Sample Copy Questionnaire

\* Please, provide information about demographic information of the expert reporting.

(a.) Name: \_\_\_\_\_ (b.) Sex: \_\_\_\_\_ (c.) Age: \_\_\_\_\_ .

(d.) Highest Education \_\_\_\_\_ (e.) Occupation \_\_\_\_\_

(f.) Department: \_\_\_\_\_ (f.) Organization: \_\_\_\_\_

A. Among the crops listed below, tick or circle which crop(s) you would like to provide the information about the annual production in your locality.

(a.) Maize (b.) Millet (c.) Sorghum (d.) Rice (e.) Millet . (f.) Benniseed  
(Sesame) (g.) Wheat (h.) Soybean (i.) Groundnut (j.)

Others(specify) \_\_\_\_\_

B. Provide information about farmers' production output in last cropping season in Niger state.

1. May I ask you how can you describe the farmers' crop output last cropping season?

Q1: What do you think is the minimum production level (tonne /ha) for the crop(s) you have selected (2)above?.

2 How confidence is you about this value? {Report your response in percentage in enclosed bracket}

A1: E.g. 2.5 / ha , [20% \_\_\_\_ ]

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].  
 (3) Crop \_\_\_\_\_ [.....]      (4) Crop \_\_\_\_\_ [.....].  
 (5) Crop \_\_\_\_\_ [.....]      (6) Crop \_\_\_\_\_ [.....].

E.g.(An experts experimental data (2.5, 0.20) is acquired)

Q2: What do you think is the maximum crop yield? And how sure are you?

A2: E.g. 12.5 ton / ha [90 % ], better than last year, poor harvest or bumper yield this year

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].  
 (3) Crop \_\_\_\_\_ [.....]      (4) Crop \_\_\_\_\_ [.....].  
 (5) Crop \_\_\_\_\_ [.....]      (6) Crop \_\_\_\_\_ [.....].

Q3: What is the belief degree that the real production level is less than this value given in Q2?

A3: E.g. 60 %.

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].  
 (3) Crop \_\_\_\_\_ [.....]      (4) Crop \_\_\_\_\_ [.....].  
 (5) Crop \_\_\_\_\_ [.....]      (6) Crop \_\_\_\_\_ [.....].

Q4: Is there another yield value for this crop may be?

A4: E.g. 7.5 ton / ha.

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].  
 (3) Crop \_\_\_\_\_ [.....]      (4) Crop \_\_\_\_\_ [.....].  
 (5) Crop \_\_\_\_\_ [.....]      (6) Crop \_\_\_\_\_ [.....].

Q5: What is the belief degree that the real production level last year is less than this value?

A5: E.g. 85%

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].  
 (3) Crop \_\_\_\_\_ [.....]      (4) Crop \_\_\_\_\_ [.....].  
 (5) Crop \_\_\_\_\_ [.....]      (6) Crop \_\_\_\_\_ [.....].

Q6: Is there another value this production output may be? A6: E.g. 6.0 ton / ha.

- (1) Crop \_\_\_\_\_ [.....]      (2) Crop \_\_\_\_\_ [.....].

(3) Crop\_\_\_\_\_ [.....]      (4) Crop\_\_\_\_\_ [.....].

(5) Crop\_\_\_\_\_ [.....]      (6) Crop\_\_\_\_\_ [.....].

Q7: What is the belief degree that the real crop yield is less than this value in Q6?

A7: E.g. 50 % as average, or good (65%)

((1) Crop\_\_\_\_\_ [.....]      (2) Crop\_\_\_\_\_ [.....].

(3) Crop\_\_\_\_\_ [.....]      (4) Crop\_\_\_\_\_ [.....].

(5) Crop\_\_\_\_\_ [.....]      (6) Crop\_\_\_\_\_ [.....].

Q8: Do you think of another yield value? A8: E.g. No idea.

(1) Crop\_\_\_\_\_ [.....]      (2) Crop\_\_\_\_\_ [.....].

(3) Crop\_\_\_\_\_ [.....]      (4) Crop\_\_\_\_\_ [.....].

(5) Crop\_\_\_\_\_ [.....]      (6) Crop\_\_\_\_\_ [.....].

### THANK YOU FOR YOUR COOPERATION

### B.3: R Code for Estimation of Uncertainty moments

This section implements the R code for computations in Uncertainty statistics. The results of this section are presented in Chapter 6. R programmes require that all comments and other documentations are preceded with symbol # and the commands are just statements. To facilitate easy understanding, the codes are written within the R program environment, so that these codes can be used by unchanged. This appendix was structured into segments, and the first line (italic sentence as comment command) of each segment is not part of R code, it is meant to guide the reader.

```
#=====
```

```
# Estimation Expectation of Empirical Distributions
```

```
#=====
```

```
# Uncertainty Empirical Distribution
```

```
emp.dist <- function(x, alpha){
```

```
  n <- length(x)
```

```
  part1 <- mean(alpha[1:2]) * x[1]
```

```
  part2 <- 0
```

```
  for(i in 2:(n-1)){
```

```
    part0 <- 0.5 * (alpha[i+1] - alpha[i-1]) * x[i]
```

```

    part2 <- part2 + part0
  }
  part3 <- (1 - mean(alpha[(n-1):n])) * x[n]

  result <- sum(part1, part2, part3)
  return(result)
}
# Maize
x <- c(1.8, 2, 2.5, 2.7)
alpha <- c(0.1, 0.5, 0.75, 1)
emp.dist(x, alpha)
#=====

# Rice
x <- c(0.8, 1, 1.2, 1.5, 2)
alpha <- c(0,0.4, 0.7, 0.9,1)
emp.dist(x, alpha)
#=====

# Sorghum
x <- c(0.8, 1, 1.2, 1.5)
alpha <- c(0.2, 0.5, 0.9, 1)
emp.dist(x, alpha)
#=====

# Soybean
x <- c(1, 1.5, 1.8, 2.0)
alpha <- c(0.2, 0.5, 0.75, 1)
emp.dist(x, alpha, 1)

#=====

# Millet
x <- c(0.8, 1.0, 1.2, 1.5, 1.8)
alpha <- c(0, 0.4,0.8, 0.9, 1)
emp.dist(x, alpha)
#=====

```

```
# Estimation First and second moment Uncertainty Linear Distributions
```

```
#=====
```

```
# Estimation of Uncertainty Distribution
```

```
# Uncertainty Moment estimation
```

```
mom.dist <- function(x, alpha, k){
  n <- length(x)
  apa1 <- alpha[1] * x[1]^k
  apa2 <- 0
  for(i in 1:(n-1)){
    for(j in 0:k){
      apa0 <- (alpha[i+1] - alpha[i]) * (x[i]^j) * (x[i+1]^(k-j))
      apa2 <- apa2 + ((1/(k+1)) * apa0)
    }
  }
  apa3 <- (1 - alpha[n]) * (x[n]^k)
  answer <- sum(apa1, apa2, apa3)
  return(answer)
}
```

```
mom.dist(x, alpha, 1)
```

```
mom.dist(x, alpha, 2)
```

```
#=====
```

```
# Maize
```

```
x <- c(1.8, 2, 2.5, 2.7)
```

```
alpha <- c(0.1, 0.5, 0.75, 1)
```

```
mom.dist(x, alpha, 1)
```

```
[1] 2.1525
```

```
mom.dist(x, alpha, 2)
```

```
[1] 4.731
```

```
#=====
```

```
# Rice
```

```
x <- c(0.8, 1, 1.2, 1.5,2)
```

```
alpha <- c(0,0.4, 0.7, 0.9,1)
```

```
mom.dist(x, alpha, 1)
```

```

[1] 1.135
mom.dist(x, alpha, 2)
[1] 1.363667
=====

# Sorghum
x <- c(0.8, 1, 1.2, 1.5)
alpha <- c(0.2, 0.5, 0.9, 1)
mom.dist(x, alpha, 1)
[1] 1.005
> mom.dist(x, alpha, 2)
[1] 1.040333
=====

# Soybean
x <- c(1, 1.5, 1.8, 2.0)
alpha <- c(0.2, 0.5, 0.75, 1)
mom.dist(x, alpha, 1)
[1] 1.4625
mom.dist(x, alpha, 2)
[1] 2.260833
=====

# Millet
x <- c(0.8, 1.0, 1.2, 1.5, 1.8)
alpha <- c(0, 0.4, 0.8, 0.9, 1)
mom.dist(x, alpha, 1)
[1] 1.1
mom.dist(x, alpha, 2)
[1] 1.266667
#=====

# Aggregated sesame data
x <-c(230, 240, 250, 260, 270, 275, 280, 290,300)
alpha <- c(0.15, 0.40, 0.45, 0.50, 0.60, 0.64, 0.76, 0.90, 1)
emp.dist(x, alpha)
[1] 258.35

```

```

# =====
# Estimation of Parameters of Uncertainty Normal Distributions
# =====
# Computation of Parameters for Uncertainty normal Distribution
guass.dist <- function(x, alpha){
  n <- length(x)
  mex <- mean(x)
  nu1 <- 0
  nu2 <- 0
  nu12 <-0
  for(i in 1:n){
    nu1 <- nu1 + log (alpha[i]/(1 - alpha[i]) )
    nu2 <- nu2 + log (alpha[i]/(1 - alpha[i]) ) * x[i]
    nu12 <- nu12 + log (alpha[i]/(1 - alpha[i]) ) * log
(alpha[i]/(1 - alpha[i]) )
  }
  sem<-(pi * n )
  ses1 <- (mex*nu1)- nu2
  ses2 <- (nu1*nu1)-(n*nu12)
  ses3 <- ses1/ses2
  sigma<-(sem*ses3*sqrt(3))/3
  mes1<- nu1/sem
  mes2<-sigma*sqrt(3)
  mes3 <-mes1*mes2
  mean <- mex-mes3
  mean1 <- mex + mes3
  answer<-list(n,mex,mean, mean1,sigma)
  return(answer)
}
=====

# MaizeNr
x <- c(1.8, 2, 2.5, 2.7)
alpha <- c(0.1, 0.5, 0.75, 0.99)

```

```
guass.dist(x, alpha)
```

```
=====
```

```
# RiceNr
```

```
x <- c(0.8, 1, 1.2, 1.5,2)
```

```
alpha <- c(0.001,0.4, 0.7, 0.9,0.99)
```

```
guass.dist(x, alpha)
```

```
=====
```

```
# SorghumNr
```

```
x <- c(0.8, 1, 1.2, 1.5)
```

```
alpha <- c(0.2, 0.5, 0.9, 0.99)
```

```
guass.dist(x, alpha)
```

```
=====
```

```
# SoybeanNr
```

```
x <- c(1, 1.5, 1.8, 2.0)
```

```
alpha <- c(0.2, 0.5, 0.75, 0.99)
```

```
guass.dist(x, alpha)
```

```
=====
```

```
# MilletNr
```

```
x <- c(0.8, 1.0, 1.2, 1.5, 1.8)
```

```
alpha <- c(0.01, 0.4,0.8, 0.9, 0.99)
```

```
guass.dist(x, alpha)
```

```
#=====
```

```
# Estimated Uncertainty Linear and Normal Distributions for Sesame
```

```
#=====
```

```
# Sesame first and second moments estimation
```

```
x <-c(250, 260, 270, 275, 280, 290, 295, 300)
```

```
alpha <- c(0.22, 0.43, 0.66, 0.81, 0.89, 0.96, 0.975, 1)
```

```
mom.dist(x, alpha, 1)
```

```
[1] 264.375
```

```
mom.dist(x, alpha, 2)
```

```
[1] 70057.42
```

```
=====
```

```
# Sesame uncertainty normal estimation
```

```

> x <-c(250, 260, 270, 275, 280, 290, 295, 300)
> alpha <- c(0.22, 0.43, 0.66, 0.81, 0.89, 0.96, 0.975, 0.99)
> guass.dist(x, alpha)
[[5]]
[1] 15.555
#####

# R Code for Graphs
#=====
# Plot for Estimated Uncertainty Linear Distributions
#=====
rm(list=ls())
# Linear uncertainty function
linear.sbean <- function(a, b, x, n=1000){
x <- seq(x[1], x[2], length=n)
phi.x <- numeric(1)
theta1 <- 1/(b-a)
theta2 <- -1 * a * theta1
for(i in 1:n){
if(x[i] < a){phi.x[i] <- 0} else if(x[i] >= b){phi.x[i] <- 1} else {
phi.x[i] <- theta1*x[i] + theta2
}
}
plot(x, phi.x, xlab="x", ylab=expression(phi(x)),
type="l", col="red", lwd=1)
}
# Linear estimated millet
linear.sbean(a=0.2008, b=1.8412, x=c(0,3))
# Linear estimated sorghum
linear.sbean(a=0.7035, b=1.3065, x=c(0,3))
# Linear estimated soybean
linear.sbean(a=0.8577, b=2.0673, x=c(0.5,3))
# linear estimate maize
linear.sbean(a=1.611, b=2.694, x=c(1.2,3))
# linear estimate rice

```

```

linear.sbean(a=0.6582, b=1.6118, x=c(0.5,3))
#=====
# Plot for Estimated Uncertainty Normal Distributions
#=====
normal.sbean <- function(mu, sigma, alpha=c(), x=T, n=1000){
  if(x[1]){
    low <- max(0, mu-(4.5*sigma))
    up <- mu + (4.5*sigma)
    x <- seq(low, up, length=n)
  }else{ x <- seq(x[1], x[2], length=n) }
  #end of default unspecified normal x
  phi.x <- numeric(1)
  for(j in 1:n){
    temp <- exp( (pi*(mu-x[j])) / (sigma*sqrt(3)))
    phi.x[j] <- (1 + temp)^(-1)
  }
  plot(x, phi.x, xlab="Estimated value x", ylab=expression(phi(x)),
  type="l", col="red", lwd=2)
  predicted <- c()
  inv.func <- function(alpha){mu +
  (sigma*sqrt(3)/pi) * log(alpha/(1-alpha))}

  If (length(alpha) > 0){
    predicted <- inv.func(alpha)
    for(k in 1:length(predicted)) {
      arrows(x0=0, y0=alpha[k], x1=predicted[k], y1=alpha[k],
      length=0, col="green")
      arrows(x0=predicted[k], y0=0, x1=predicted[k],
      y1=alpha[k], length=0, col="green")
    }
  }
  answer <- cbind(alpha, predicted)
  return(answer)

```

```

}
# Maize at predicted level of Alpha
normal.sbean(mu=2.128, sigma=0.246, alpha=c(0.4,0.5))
# rice predicted at level of Alpha
normal.sbean(mu=1.294, sigma=0.174, alpha=c(0.5))
# sorghum predicted at level of Alpha
normal.sbean(mu=0.9713, sigma=0.20632, alpha=c(0.5))
# soybean predicted at level of Alpha
normal.sbean(mu=1.4104, sigma=0.27729, alpha=c(0.5))
# millet predicted level of Alpha
normal.sbean(mu=1.1906, sigma=0.1980, alpha=c(0.5))
# Sesame predicted at level of Alpha
normal.sbean(mu=277.50, sigma=15.56, alpha=c(0.5))
normal.sbean(mu=2.128, sigma=0.246) # Normal maize
normal.sbean(mu=1.294, sigma=0.174)# Normal Rice
normal.sbean(mu=0.9713, sigma=0.20632)# Normal Sorghum
normal.sbean(mu=1.4104, sigma=0.27729)# Normal Soybean
normal.sbean(mu=1.1906, sigma=0.1980)# Normal Millet
list(x, phi.x)
#=====
# Plot for Empirical distribution of Crop yield raw data from expert
#=====
# Maize plot
x <- c(1.5,2,2.5,2.7)
y <- c(0.1,0.5,0.75,1)
plot(x,y,type="l", xlim=c(1.0,2.8),xlab="(tonne/ha)",ylab=".",lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)
text(x,y,c("(1.5, 0.1)","(2, 0.5)", "(2.5, 0.75)", "(2.7, 1.0)"),
adj=1.5, cex=1.0)
# Rice plot
x <- c(0.8, 1, 1.2, 1.5,2)
y <- c(0,0.4, 0.7, 0.9,1)
plot(x,y,type="l", xlim=c(0.5,2.8),xlab="(tonne/ha)",ylab=".",lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)

```

```

text(x,y,c("(0.8, 0)", "(1.0, 0.4)", "(1.2, 0.7)", "(1.5, 0.9)", "(2.0, 1.0)"),
adj=1.2, cex=1.0)
# Surghum plot
x <- c(0.8, 1, 1.2, 1.5)
y <- c(0.2, 0.5, 0.9, 1)
plot(x,y,type="l", xlim=c(0.5,2.0),xlab="(tonne/ha)",ylab=".",lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)
text(x,y,c("(0.8, 0.2)", "(1.0, 0.5)", "(1.2, 0.9)", "(1.5, 1.0)"),
adj=1.5, cex=1.0)
# Soyabean plot
x <- c(1, 1.5, 1.8, 2.0)
y <- c(0.2, 0.5, 0.75, 1)
plot(x,y,type="l", xlim=c(0.5,2.5),xlab="(tonne /ha)",ylab=".",lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)
text(x,y,c("(1.0, 0.2)", "(1.5, 0.5)", "(1.8, 0.75)", "(2.0, 1.0)"),
adj=1.2, cex=1.0)
# Millet plot
x <- c(0.8, 1.0, 1.2, 1.5, 1.8)
y <- c(0, 0.4, 0.8,0.9, 1.0)
plot(x,y,type="l", xlim=c(0.5, 2.0),xlab="(tonne/ha)",ylab=".",lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)
text(x,y,c("(0.8, 0.0)", "(1.0, 0.4)", "(1.2, 0.8)", "(1.5, 0.9)", "(1.8, 1.0)"),
adj=1.2, cex=1.0)
#=====
# Lebesque function
#=====
# Graph of Lebesque function for expected value
x <- seq(-1, 1, length=1000)
y <- function(x) (x^3)
plot(y(x), x, type="l", xaxt="n", yaxt="n", xlab="",ylab="y")
abline(v=0)
polygon(c(0,0,-1,y[x[1:500]],y[x[501:1000]],1,0,0), c(0,-1,-1,x[1:500],
x[501:1000],1,1,0), col="grey")
#=====
# Plot of Sesame Raw data

```

```

#=====
x <-c(250, 260, 270, 275, 280, 290, 295, 300)
y <- c(0.22, 0.43, 0.66, 0.81, 0.89, 0.96, 0.975, 0.99)
plot(x,y,type="l", xlim=c(225, 350), xlab="(kg /ha)",ylab=".",col="blue", lwd=2)
mtext(expression(alpha), side=2, padj=-5.7)
text(x,y,c("(250, 0.22)", "(260, 0.43)", "(270, 0.66)", "(275, 0.81)", "(280, 0.89)",
"(290, 0.96)", "(295, 0.975)", "(300, 0.99)"),adj=1.2, cex=1.0)
#=====
# Plot of Standard Gaussian Distribution and Uncertainty Normal Distribution
#=====
rm(list=ls())
class.uncert <- function(x, mu=0, sigma=1){
b <- (sigma * sqrt(3)) / pi
power1 <- (x - mu) / b
uncertain <- 1 / (1 + exp(-power1))
  gauss <- pnorm(x, mu, sigma)
  titre <- expression(paste(Psi(x), " ", Phi(x)))
plot(x, gauss, type="l", lwd=2, col=2, ylab=titre)
lines(x, uncertain, col=3, lwd=2)
legend("topleft", lwd=c(2,2), col=c(2:3),
legend=c(expression(Psi(x)),expression(Phi(x))))
}
x <- seq(-3, 3, by=0.1)
class.uncert(x)
#*****CODE ENDS HERE*****#

```

## Appendix C: R Code for Regression Model

Regression Analysis of “FULL” data

This appendix is subdivided into two parts, namely: Classical analysis (OLS and Bayesian estimation). We have implemented *MCMCpack* and *R2WinBUGS* in R environment. To facilitate easy implementation, our codes include comments # within the R program.

```
# Set working directory
```

```
rm(list=ls())
```

```
setwd('C:/Users/Adeyemi/Desktop/dddd')
```

```
# Read data
```

```
sorgh <- read.csv('SORGHUMS.csv')
```

```
str(sorgh) # structure of your data (no. of obs, variables,...)
```

```
head(sorgh) # view part of your data
```

```
tail(sorgh)
```

```
length(sorgh$y)
```

```
summary(sorgh)
```

```
## Summary Statistics
```

```
summary(sorgh)
```

```
*****
```

### C.1: R2WinBUGS Code for Bayesian model with non- informative priors

```
*****
```

```
# Loading packages
```

```
library(MCMCpack)
```

```
library(lattice)
```

```
library(MASS)
```

```
library(lattice)
```

```
library(coda)
```

```
library(boot)
```

```
library(R2WinBUGS)
```

```
*****
```

```
n <- nrow(sorgh)
```

```
p <- 13
```

```

# load packages
# WinBUGS code for regression
#*****

#Multiple ridge regression starts here
sink("multibayes.txt")
cat("
model
{
# Likelihood Model
# d <- 4;           # degrees of freedom for t
  for (i in 1 : n) {
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta[1] + (beta[2]*v1[i]) + (beta[3]*v2[i]) + (beta[4]*v3[i]) +
    (beta[5]*s1[i]) + (beta[6]*s2[i]) + (beta[7]*s3[i]) + (beta[8]*e1[i])
    + (beta[9]*e2[i]) + (beta[10]*e3[i]) + (beta[11]*e4[i]) + (beta[12]*v6[i])
    + (beta[13]*v7[i]) + (beta[14]*v8[i])
    #mu[i] <- beta0 + beta[1] * v1[i] + beta[2] * v2[i] + beta[3] * v3[i] +
    #beta[4] * v4[i] + beta[5] * v5[i] + beta[6] * v6[i] + beta[7] * v7[i]
    # + beta[8] * v8[i]
  }
# standard deviation of error distribution
sigma~ dunif(0,10)
tau <- 1 / (sigma*sigma)           # normal errors
# Priors
#beta0 ~ dnorm(0, 0.0001)
for (j in 1 : 14) {
  beta[j] ~ dnorm(0, 0.001) # coeffs independent
}
}
# model
",fill=TRUE)
sink()
#*****

```

```

# WinBUGS data and factoring predictors
#*****
data <- list(y= sorgh$y, v1= sorgh[,2], v2= sorgh[,3],v3= sorgh[,4],
  s1= sorgh[,5],s2= sorgh[,6],s3= sorgh[,7],e1= sorgh[,8],
  e2= sorgh[,9],e3= sorgh[,10],e4= sorgh[,11],
  v6= sorgh[,12],v7= sorgh[,13], v8= sorgh[,14],n=length(sorgh$y))
#data <- list(y,v1,v2,v3,sk,sm,sn,v5,v6,v7,v8)
#data <- list(y= sorgh$y, v1= sorgh[,2], v2= sorgh[,3],v3= sorgh[,4], v4= sorgh[,5],
#  v5= sorgh[,6], v6= sorgh[,7], v7= sorgh[,8],v8= sorgh[,9],p=8,n=length(sorgh$y))
#Initializing the parameters
#inits <- function(){list(beta0 =1500, beta=c(0,0,0,0,0,0,0,0,0), sigma = 0.001)}
inits <- function(){list(beta= rnorm(14,0,1), sigma=runif(1,0,10))}
#inits <- function(){list(beta= rnorm(15,0,1), sigma=0.01)}

#parameters monitored
pars <- c('beta', 'sigma')

out <- bugs(data=data,inits=inits, parameters=pars,
model.file="multibayes.txt",n.chains=2, n.iter=50000, n.burnin=3000, n.thin=3,
debug=TRUE, working.directory=getwd())
print(out,digit=2)
#print(out)

```

node	mean	sd	MC error	2.50%	median	97.50%	start
beta[1]	3192	6.206	0.03363	3180	3192	3204	1001
beta[2]	-776	0.8347	0.004934	-777.6	-776	-774.4	1001
beta[3]	1.113	0.007566	4.13E-05	1.098	1.113	1.128	1001
beta[4]	37.64	0.1106	5.78E-04	37.42	37.64	37.86	1001
beta[5]	-567.2	6.148	0.03794	-579.2	-567.2	-555.3	1001
beta[6]	175.2	2.298	0.01307	170.7	175.2	179.7	1001
beta[7]	449.2	2.784	0.01565	443.8	449.2	454.7	1001
beta[8]	113.9	3.059	0.01731	107.9	113.9	119.9	1001
beta[9]	-407.1	2.647	0.01456	-412.3	-407.1	-401.9	1001
beta[10]	-755.4	2.522	0.01425	-760.4	-755.4	-750.4	1001
beta[11]	811.7	7.21	0.04217	797.5	811.7	825.6	1001
beta[12]	-0.3939	0.07002	4.10E-04	-0.531	-0.3944	-0.2572	1001
beta[13]	-81.01	4.686	0.02676	-90.15	-81.02	-71.92	1001

beta[14]	180.4	3.083	0.01727	174.4	180.4	186.5	1001
deviance	3.63E+06	44.54	0.2433	3.63E+06	3.63E+06	3.63E+06	1001
sigma	10	0	1.65E-08	10	10	10	1001
	Dbar	Dhat	pD	DIC			
y	3633280	3633260	13.755	3633290			
total	3633280	3633260	13.755	3633290			

```
#Extracting parameters from sim outputs
```

```
names(out$sims.list)
```

```
beta<-out$sims.list$beta
```

```
beta[,1]<-out$sims.list$beta[,1]
```

```
beta[,2]<-out$sims.list$beta[,2]
```

```
beta[,3]<-out$sims.list$beta[,3]
```

```
beta[,4]<-out$sims.list$beta[,4]
```

```
beta[,5]<-out$sims.list$beta[,5]
```

```
beta[,6]<-out$sims.list$beta[,6]
```

```
beta[,7]<-out$sims.list$beta[,7]
```

```
beta[,8]<-out$sims.list$beta[,8]
```

```
beta[,9]<-out$sims.list$beta[,9]
```

```
beta[,10]<-out$sims.list$beta[,10]
```

```
beta[,11]<-out$sims.list$beta[,11]
```

```
beta[,12]<-out$sims.list$beta[,12]
```

```
beta[,13]<-out$sims.list$beta[,13]
```

```
beta[,14]<-out$sims.list$beta[,14]
```

```
sigma<-out$sims.list$sigma
```

```
#Autocorrelation Plots fro convergence
```

```
dev.off()
```

```
par(mfrow=c(4,4))
```

```
autocorr.plot(beta[,1],main=expression("beta0"),col="blue")
```

```
autocorr.plot(beta[,2],main="beta1",col="blue")
```

```
autocorr.plot(beta[,3],main="beta2",col="blue")
```

```
autocorr.plot(beta[,4],main="beta3",col="blue")
```

```
autocorr.plot(beta[,5],main="beta4",col="blue")
```

```
autocorr.plot(beta[,6],main="beta5",col="blue")
```

```

autocorr.plot(beta[,7],main="beta6",col="blue")
autocorr.plot(beta[,8],main="beta7",col="blue")
autocorr.plot(beta[,9],main="beta8",col="blue")
autocorr.plot(beta[,10],main="beta9",col="blue")
autocorr.plot(beta[,11],main="beta10",col="blue")
autocorr.plot(beta[,12],main="beta11",col="blue")
autocorr.plot(beta[,13],main="beta12",col="blue")
autocorr.plot(beta[,14],main="beta13",col="blue")
dev.off()
plot(beta[,2],beta[,1])
pairs(out)
A=data.frame(out$sims.list)
# MILLET plotting parameters
par(mar=c(5,7,1,1))
par(mfrow=c(4,3))
hist(beta[,1],prob=T,breaks=12,xlab=".", main="beta0")
lines(density(beta[,1]))
abline(v=3203, col=3, lwd=2)
abline(v=3449, col="red",lwd=2)
###=====
hist(beta[,2],prob=T,breaks=12,xlab=".", main="beta1")
lines(density(beta[,2]))
abline(v=-442, col=3, lwd=2)
abline(v=-445, col="red",lwd=2)
#par(mfrow=c(4,2))
hist(beta[,3],prob=T,xlab=".", main="beta2")
lines(density(beta[,3]))
abline(v=0.28, col=3, lwd=2)
abline(v=0.47, col="red",lwd=2)
###=====
hist(beta[,4],prob=T,breaks=12,xlab=".", main="beta3")
lines(density(beta[,4],col="blue"))
abline(v=-12, col=3, lwd=1)

```

```

abline(v=-12, col="red",lwd=1)
##=====
hist(beta[,5],prob=T,breaks=12, xlab=".", main="beta4")
lines(density(beta[,5],col="blue"))
  abline(v=466, col=3, lwd=2)
  abline(v=372, col="red",lwd=2))
##=====
hist(beta[,6], prob=T,breaks=12, xlab=".",main="beta5")
lines(density(beta[,6]))
abline(v=33, col=3, lwd=2)
abline(v=-80, col="red",lwd=2)
##=====
hist(beta[,7], prob=T,breaks=12,xlab=".", main="beta6")
lines(density(beta[,7]))
abline(v=-121, col=3, lwd=2)
abline(v=-241, col="red",lwd=2)
##=====
hist(beta[,8],prob=T,breaks=12, xlab=".", main="beta7")
lines(density(beta[,8]))
abline(v=-594, col=3, lwd=2)
abline(v=-677, col="red",lwd=2)
#####
hist(beta[,9], prob=T,breaks=12, xlab=".",main="beta8")
lines(density(beta[,9]))
abline(v=-803, col=3, lwd=2)
abline(v=-865, col="red",lwd=2)
dev.off()
#####

```

## **C.2: R code OLS and Direct MCMC Estimation**

```

#####
# Set working directory
rm(list=ls())
setwd('C:/Users/Adeyemi/Desktop/dddd')

```

```

# Read data
sorgh <- read.csv('MILLETS.csv')
str(sorgh) # structure
head(sorgh) # view part of your data
#*****

# OLS and Direct MCMC estimation using MCMCpack
#*****

library(MCMCpack)
posterior<-MCMCregress(y~ v1+ v2+v3+ factor(v4)+ factor(v5) +v6 +v7+v8, data= sorgh)
summary(posterior)
# Ordinary Least square estimate of Linear regression
cm<-lm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8, data= sorgh)
res<-resid(cm)
lm(formula=y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 +v8 , data= sorgh)
summary(cm)
confint(cm)
par(mar = c(4, 6,1,1))
plot(posterior,trace=FALSE)
dev.off()
#autocorrelation
autocorr.plot(posterior,col="red")
#*****

# Multiple linear regression OLS
#*****

cm<-lm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8, data= sorgh)
res<-resid(cm)
lm(formula=y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 +v8 , data= sorgh)
summary(cm)
confint(cm)
par(mfrow=c(2,2))
plot(cm)
hist(res,breaks=20, prob=T, col="blue")
lines(density(res))

```

```

dev.off()
# SORGHUM plotting parameters
par(mar=c(5,7,1,1))
par(mfrow=c(4,3))
hist(beta[,1],prob=T,breaks=12, main="beta0")
lines(density(beta[,1]), col="blue")
abline(v=3192, col=3, lwd=2)
abline(v=3470, col="red",lwd=2)
#*****

```

### **C.3: Bayesian Predictive model for set of covariates**

```

#*****
rm(list=ls())
setwd('C:/Users/Adeyemi/Desktop/dddd')
# Read data
sorgh <- read.csv('SORGHUMS.csv')
str(sorgh) # structure of your data (no. of obs, variables,...)
head(sorgh) # view part of your data
tail(sorgh)
library(MCMCpack)
cm<-lm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8, data= sorgh)
res<-resid(cm)
lm(formula=y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 +v8 , data= sorgh)
library(LearnBayes)
length(sorgh$y)
fit=lm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8, data= sorgh,x=TRUE,y=TRUE)
summary(fit)
# Simulated sample of 1000 draws from the joint posterior ditribution
theta.sample=blinreg(fit$y,fit$x,1000)
mean(theta.sample$beta[,2])
mean(theta.sample$beta[,3])

```

```

mean(theta.sample$beta[,4])
mean(theta.sample$beta[,5])
mean(theta.sample$beta[,6])
mean(theta.sample$beta[,7])
mean(theta.sample$beta[,8])
mean(theta.sample$beta[,9])
mean(theta.sample$beta[,10])
mean(theta.sample$beta[,11])
mean(theta.sample$beta[,12])
mean(theta.sample$beta[,13])
sd(theta.sample$beta[,2])
sd(theta.sample$beta[,3])
sd(theta.sample$beta[,4])
sd(theta.sample$beta[,5])
sd(theta.sample$beta[,6])
sd(theta.sample$beta[,7])
sd(theta.sample$beta[,8])
sd(theta.sample$beta[,9])
sd(theta.sample$beta[,10])
sd(theta.sample$beta[,11])
sd(theta.sample$beta[,12])
sd(theta.sample$beta[,13])

```

```
# A set of Covariates # commented in R
```

```

#cov1=c(1, 1.2, 83.33, 23.33, 0, 1, 0, 0, 0, 0, 0, 38, 1, 0)
#cov2=c(1, 2.5, 200, 18.4, 0, 1, 0, 0, 1, 0, 0, 37, 1, 1)
#cov3=c(1, 3.5, 42.86, 19.43, 0, 1, 0, 0, 1, 0, 0, 20, 1, 1)
#cov4=c(1, 2.5, 200, 18.4, 0, 1, 0, 0, 1, 0, 0, 37, 1, 1)

```

```
# Another set of Covariates
```

```
cov5=c(1, 2.7, 74.07, 20, 0, 1, 0, 0, 0, 0, 0, 30, 1, 0)
```

```

cov6=c(1, 1.7, 147.06, 23.53, 0, 1, 0, 0, 0, 0, 38, 1, 1)
cov7=c(1, 3, 133.33, 20.33, 0,1, 0, 0, 0, 1, 0, 30, 0, 0)
cov8=c(1, 2.3, 86.96, 20, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1)

```

```
# X1=rbind(cov1,cov2,cov3,cov4) # Combining the set of Covariates
```

```
X2=rbind(cov5,cov6,cov7,cov8) # Combining another set of covariates
```

```
mean.draws=blinregexpected(X2, theta.sample)
```

```
# 90% interval estimates for each groups
```

```
quantile(mean.draws[,1],c(0.05,0.95))
```

```
quantile(mean.draws[,2],c(0.05,0.95))
```

```
quantile(mean.draws[,3],c(0.05,0.95))
```

```
quantile(mean.draws[,4],c(0.05,0.95))
```

```
# 90% prediction intervals for each groups
```

```
pred.draws=blinregpred(X2,theta.sample)
```

```
quantile(pred.draws[,1],c(0.05,0.95))
```

```
quantile(pred.draws[,2],c(0.05,0.95))
```

```
quantile(pred.draws[,3],c(0.05,0.95))
```

```
quantile(pred.draws[,4],c(0.05,0.95))
```

```
mean(pred.draws[,1])
```

```
mean(pred.draws[,2])
```

```
mean(pred.draws[,3])
```

```
mean(pred.draws[,4])
```

```
mean(mean.draws[,1])
```

```
mean(mean.draws[,2])
```

```
mean(mean.draws[,3])
```

```
mean(mean.draws[,4])
```

```
#Histogram of Predicted values for set of covariates
```

```
par(mfrow=c(2,2))
```

```
hist(pred.draws[,1],prob=TRUE,main="Covariate set A",xlab="kg/ha")
```

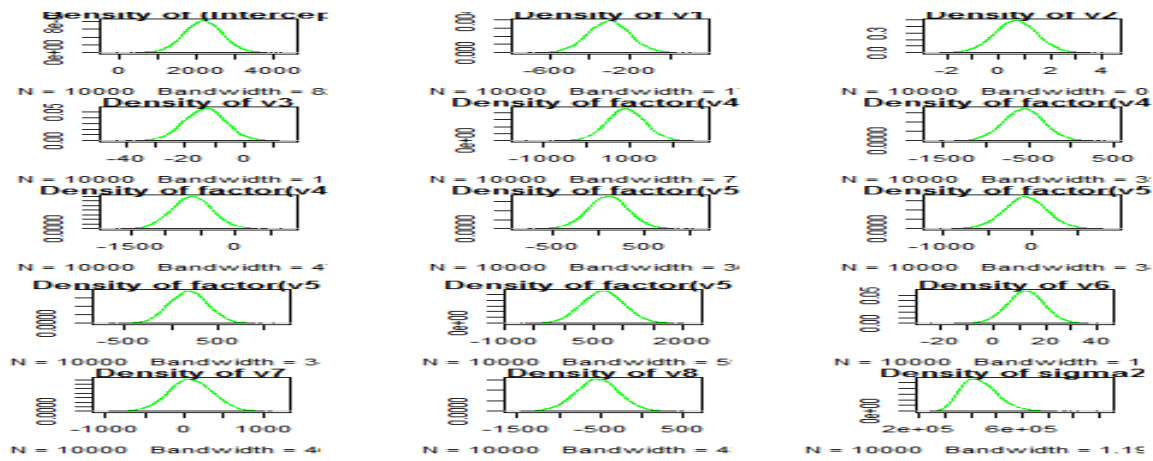
```
hist(pred.draws[,2],prob=TRUE,main="Covariate set B",xlab="kg/ha")
```

```
hist(pred.draws[,3],prob=TRUE,main="Covariate set C",xlab="kg/ha")
```

```

hist(pred.draws[,4],prob=TRUE,main="Covariate set D",xlab="kg/ha")
dev.off()
pred.draws=blinregpred(X2,theta.sample)
quantile(pred.draws[,1],c(0.05,0.95))
quantile(pred.draws[,2],c(0.05,0.95))
quantile(pred.draws[,3],c(0.05,0.95))
quantile(pred.draws[,4],c(0.05,0.95))
mean(pred.draws[,1])
mean(pred.draws[,2])
mean(pred.draws[,3])
mean(pred.draws[,4])
mean(mean.draws[,1])
mean(mean.draws[,2])
mean(mean.draws[,3])
mean(mean.draws[,4])
apply(mean(pred.draws[,1]),2,quantile(pred.draws[,1],c(0.05,0.95)))
apply(mean(pred.draws[,1]),2,(quantile(pred.draws[,1],c(0.05,0.95))))
hist(pred.draws[,1],main="Covariate set A",xlab="kg/ha")
dev.off()
par(mfrow=c(2,2))
hist(pred.draws[,1],main="Covariate set A",xlab="kg/ha")
hist(pred.draws[,2],main="Covariate set B",xlab="kg/ha")
hist(pred.draws[,3],main="Covariate set C",xlab="kg/ha")
hist(pred.draws[,4],main="Covariate set D",xlab="kg/ha")
dev.off()
par(mfrow=c(2,2))
hist(pred.draws[,1],prob=TRUE,main="Covariate set A",xlab="kg/ha")
hist(pred.draws[,2],prob=TRUE,main="Covariate set B",xlab="kg/ha")
hist(pred.draws[,3],prob=TRUE,main="Covariate set C",xlab="kg/ha")
hist(pred.draws[,4],prob=TRUE,main="Covariate set D",xlab="kg/ha")

```



Density Plots for Millet direct MCMC

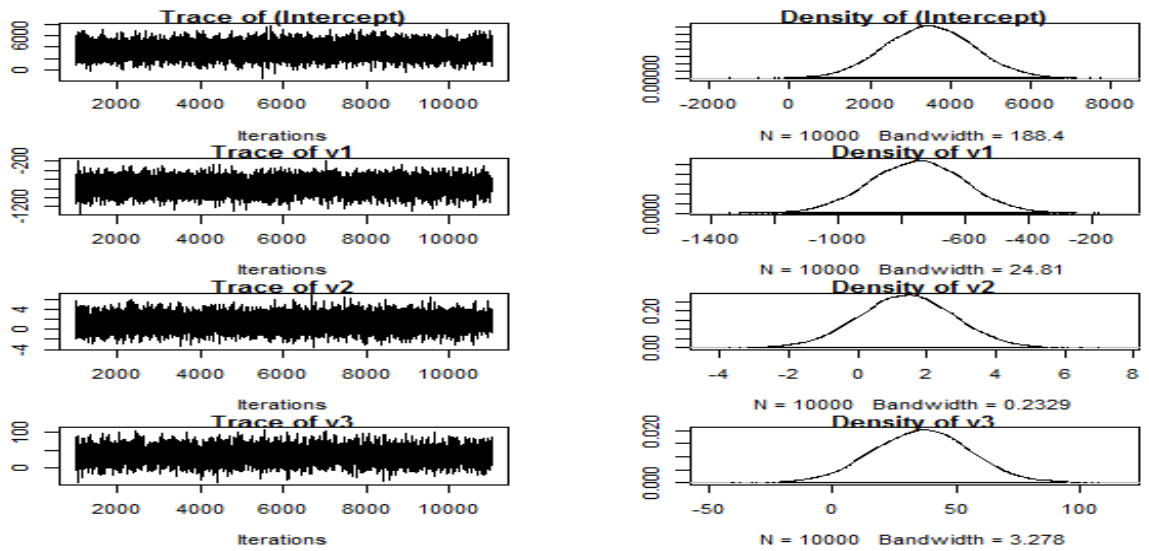


Figure 7.1: Density Plots and Trace plot of posterior estimates for sorghum *kg/ha* .

# Appendix D: Regression analysis on SPLIT data

```
#####
```

## D.1: R Code for Linear Regression analysis on Split data

```
#####
```

```
# Set working directory
rm(list=ls())
setwd("C:/Users/Adeyemi/Desktop/millet")
# Read data
sorgh <- read.csv('MILLETS.csv', header=TRUE)
n<-nrow(sorgh)
n2 <- round(n/2,0)
old_data <-sorgh[1:n2,]
new_data <- sorgh[(n2+1):n,]
# val_data <- sorgh[36:85,]
# test_data <- sorgh[86:135,]

#install.packages("faraway")
library(faraway)
old_data=na.omit(old_data) #removing missing values in categoricalvariables
new_data=na.omit(new_data)
# val_data=na.omit(val_data)
# test_data=na.omit(test_data)
#####Linear Model on old and newdata #####
mod_old=glm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8,
family=gaussian, old_data)
mod_new=glm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 + v7 + v8,
family=gaussian,new_data)
#####
#checking assumptions on old data
#collinearity
ss01 = summary(mod_old)
save01 = summary(mod_new)
write.csv(ss01$coef, "milletOldFile.csv")
write.csv(save01$coef, "milletNewFile.csv")
x1=model.matrix(mod_old)[-1]
```

```

x1=x1[,c(-4:-10,-12,-13)] #remove categorical variables
cor.matrix = cor(x1) #give correlation matrix of numerical independent variables
write.csv(cor.matrix, "milletOldFileCorrelation.csv")
vifVect = vif(x1) #give variance inflation factors of numerical independent
write.csv(vifVect, "milletVifOldFile.csv")
#variabels
#outliers and influential observations
png("oldHalfnormal%03d.png")
  halfnorm(rstudent(mod_old)) #Half-Normal plot of residuals
  ga=influence(mod_old)
  halfnorm(ga$hat) #half normal plot of influence
  halfnorm(cooks.distance(mod_old)) #half normal plot of Cooks statistics
dev.off()
# *****
# Linear model on old split data excluding the influential observations
# *****
mod_old1=glm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6
  + v7 + v8, old_data, subset=-c(48,23,35,36,3),family=gaussian)
result01 = summary(mod_old1)
write.csv(result01$coef, "milletOldmod_1File.csv")
# -----
compare00 <- cbind(coef(mod_old),coef(mod_old1)) #comparing parameters
write.csv(compare00, "milletOldComparedCoefs.csv")

#assumptions on new data
x2=model.matrix(mod_new)[,-1]
x2=x2[,c(-4:-10,-12,-13)]

write.csv(cor(x2), "milletNewCorrelation.csv")
write.csv(vif(x2), "milletNewVif.csv")

png("newHalfnormal%03d.png")
  halfnorm(rstudent(mod_new)) #residuals
  ga=influence(mod_new)
  halfnorm(ga$hat)
  halfnorm(cooks.distance(mod_new))
dev.off()

```

```

# *****

# Linear model on new split data excluding the influential and outlier observations
# *****

mod_new1=glm(y~ v1+ v2+v3+ as.factor(v4)+ as.factor(v5)+v6 +
            v7 + v8,family=gaussian, new_data, subset=-c(10,13,14,35,38))
# *****

summary(mod_new1)

compNEW<-cbind(coef(mod_new),coef(mod_new1)) #comparing parameters
write.csv(compNEW, "milletNewcomp3538.csv")
write.csv(summary(mod_new1)$coef, "milletNewL3538.csv")

D.2: R Code for Bayesian Models on SPLIT data
#*****
###BAYESIAN REGRESSION ANALYSIS ON SPLIT DATA STARTS HERE###
# *****

library(MCMCpack)
information=solve(vcov(mod_old1)) #information matrix for logit model on old data
information2=diag(diag(information),14,14) #diagonal information matrix
write.csv(summary(mod_new1)$coef, "milletNewLinearModel.csv")
write.csv(information, "milletVCOV.csv") # Variance Covariance matrix for Bayes prior
write.csv(information2, "milletDIAGprior.csv") # diagonal matrix for beta prior
# *****

#bayesian model on new data with informative prior:
bayes_mod=MCMCregress(y~ v1+ v2+v3+ factor(v4)+ factor(v5) +v6 +v7+v8,
                    data=new_data,subset=c(-13,-38,-14),burnin=5000,
                    mcmc=100000,b0=coef(mod_old),B0=information2)
# *****

sumb=summary(bayes_mod)
sb=sumb$statistics
sb_coefs=sb[,1] #coefficients of Bayesian linear model
geweke.diag(bayes_mod) # check geweke diagnostics
write.csv(sb, "milletBayesmod22.csv")
write.csv(geweke.diag(bayes_mod)[[1]], "mllgewekeBayesmod.csv")
png("millettrace Bayes.png")
plot(bayes_mod[,1:4])
dev.off()
# *****

#Bayesian model on new data with non-informative prior:

```



```

# --- Bayesian model on new data with informative prior --- #
# ----- #
bayesGaussPrior <- function(datum){
  olsModel <- lm(y ~ ., data=datum)
  newNames <- c("y", names(olsModel$coef)[-1])
  datum <- datum[, newNames]
  priorInformation <- solve( vcov(olsModel) )
  parameterSize <- nrow(summary(olsModel)$coef)
  priorInformation <- diag(diag(priorInformation), parameterSize, parameterSize)
  infoBayesModel <- MCMCregress(y ~ ., data=datum, burnin=500000, mcmc=10000,
                               b0=coef(olsModel), B0=priorInformation)
  model2est <- summary(infoBayesModel)$statistics
  model2est <- model2est[-nrow(model2est),]
  model2est
}
# --- Bayesian model on new data with non-informative prior --- #
# ----- #
bayesZeroPrior <- function(datum){
  zeroInfoBayesModel <- MCMCregress(y ~ ., data=datum, burnin=10000, mcmc=10000)
  model3est <- summary(zeroInfoBayesModel)$statistics
  model3est <- model3est[-nrow(model3est),]
  model3est
}
# *****
# --- Leave-one-out (LOO) cross-validation --- #
# *****
looCV <- function(datum, regressModel){
  dataObs <- nrow(datum)
  eror <- numeric()

  for(obs in 1:dataObs){
    newData <- datum[-obs,]

    bts <- switch(regressModel,
                  GLM = glmFunction(newData),

```

```

    ZEROprior = bayesZeroPrior(newData),
    GAUSSprior = bayesGaussPrior(newData))
bts <- as.matrix(bts[,1])

x.t <- datum[obs,c(2:ncol(datum))]
x.t <- as.numeric(c(1, x.t))
y.hat <- x.t %*% bts
y.t <- datum[obs,"y"]
error[obs] <- (y.t - y.hat)^2

print(sprintf("Completed run %.0f of %.0f", obs, dataObs))
}
t.error <- mean(error)
v.error <- var(error) / length(error)
std.error <- sqrt(v.error)

result <- rbind(t.error, std.error)
rownames(result) <- c("CV Error", "SE(CV)")
return(result)
}

D.4: Cross Validation output
# *****
# --- LOO cross-validation implementation for millet --- #
# *****

millet <- read.csv('MILLETS.csv', header=TRUE)
crossValidData <- millet[,1:14]
crossValidData <- crossValidData[-c(8,13),]
milletGLM <- looCV(crossValidData, "GLM")
milletZERO <- looCV(crossValidData, "ZEROprior")
milletGAUSS <- looCV(crossValidData, "GAUSSprior")
combinedAns <- cbind(milletGLM, milletZERO, milletGAUSS)
colnames(combinedAns) <- c("GLM", "No Prior", "Gauss Prior")
combinedAns

# *****LOOCV OUTPUT FO MILLET*****
alterMilletGLM <- milletGLM/100000

```

```

alterMilletZERO <- milletZERO/100000
alterMilletGAUSS <- milletGAUSS/100000
lowBound <- c(alterMilletZERO[1]-alterMilletZERO[2],
              alterMilletGLM[1]-alterMilletGLM[2],
              alterMilletGAUSS[1]-alterMilletGAUSS[2])
uppBound <- c(alterMilletZERO[1]+ alterMilletZERO[2],
              alterMilletGLM[1]+alterMilletGLM[2],
              alterMilletGAUSS[1]+ alterMilletGAUSS[2])
# *****
# --- LOO cross-validation implementation for sorghum --- #
# *****
sorghum <- read.csv('SORGHUMS.csv', header=TRUE)
sorghumData <- sorghum[,1:14]
sorghumData <- sorghumData[-c(2,25,63,32,44),]
sorghumGLM <- looCV(sorghumData, "GLM")
sorghumZERO <- looCV(sorghumData, "ZEROprior")
sorghumGAUSS <- looCV(sorghumData, "GAUSSprior")
combinedSorghumAns <- cbind(sorghumGLM, sorghumZERO, sorghumGAUSS)
colnames(combinedSorghumAns) <- c("GLM", "No Prior", "Gauss Prior")
combinedSorghumAns
# *****
alterSorghumGLM <- sorghumGLM/100000
alterSorghumZERO <- sorghumZERO/100000
alterSorghumGAUSS <- sorghumGAUSS/100000

lowBoundS <- c(alterSorghumZERO[1]-alterSorghumZERO[2],
              alterSorghumGLM[1]-alterSorghumGLM[2],
              alterSorghumGAUSS[1]-alterSorghumGAUSS[2])
uppBoundS <- c(alterSorghumZERO[1]+ alterSorghumZERO[2],
              alterSorghumGLM[1]+alterSorghumGLM[2],
              alterSorghumGAUSS[1]+ alterSorghumGAUSS[2])
# *****
# --- LOO cross-validation implementation for Millet --- #
# *****
Millet <- read.csv('MILLETS.csv', header=TRUE)

```

```

MilletData <- Millet[,1:14]
MilletmData <- MilletData[-c(5,20,33,50,74,75,31,68),]
MilletGLM <- looCV(MilletData, "GLM")
MilletZERO <- looCV(MilletData, "ZEROprior")
MilletGAUSS <- looCV(MilletData, "GAUSSprior")
combinedMilletAns <- cbind(MilletGLM, MilletZERO, MilletGAUSS)
colnames(combinedMilletAns) <- c("GLM", "No Prior", "Gauss Prior")
combinedMilletAns
alterMilletGLM <- MilletGLM/100000
alterMilletZERO <- MilletZERO/100000
alterMilletGAUSS <- MilletGauss/100000
lowBoundS <- c(alterMilletZERO[1]-alterMilletZERO[2],
               alterMilletGLM[1]-alterMilletGLM[2],
               alterMilletGAUSS[1]-alterMilletGAUSS[2])
uppBoundS <- c(alterMilletZERO[1]+ alterMilletZERO[2],
               alterMilletGLM[1]+alterMilletGLM[2],
               alterMilletGAUSS[1]+ alterMilletGAUSS[2])
# *****
# LOO cross-validation combined results summary --- #
# *****LOOCV GRAPHS*****
png("dotsorghumPlots1.png")
par(mar=c(5,5,2,5))
plot(combinedSorghumAns[1,c(1,2,3)]/100000, pch=16, col=2, xaxt="n", xlab="Models",
     xlim=c(0.5,3.5), ylab="Cross-validation error ('00000)",
     ylim=c(12.29,38.295), yaxt="n")
axis(2, col=2, lwd=2, cex=0.75, col.axis=2, las=1, cex.axis=0.75)
par(new=T)
plot(combinedSorghumAns[2,c(1,2,3)]/100000, pch=15, col=4, xaxt="n", xlab="",
     ylab="", yaxt="n", xlim=c(0.5,3.5), col.axis=4)
axis(1, 1:3,c("GLM", "No Prior", "Gauss Prior"))
axis(2, labels=F, col=2, lwd.ticks=0, lwd=2)
axis(4, col=4, lwd=2, cex=0.75, col.axis=4, las=1, cex.axis=0.75)
mtext("Standard error (CV)", 4, padj=5.2)
dev.off()
#***** Code ends here***** #

```

## Appendix E: Raw Data

### E.1: Raw data for Sorghum (15 Observations)

N	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	2166.67	3.00	0.00	20.00	0	0	0	0	0	1	0	10	1	0
2	1300.00	5.00	40.00	20.40	0	1	0	0	1	0	0	13	1	0
3	1166.67	3.00	66.67	18.00	0	1	0	0	1	0	0	15	1	0
4	1500.00	3.00	100.00	18.00	0	0	1	0	0	0	0	25	1	0
5	833.33	3.00	133.33	20.00	0	0	1	0	0	0	0	4	1	0
6	900.00	4.00	125.00	17.50	0	0	1	0	1	0	0	6	1	0
7	1800.00	2.50	20.00	19.20	0	1	0	0	0	1	0	9	1	0
8	1041.67	2.40	20.83	19.17	0	1	0	0	0	1	0	7	1	0
9	2187.50	1.60	62.50	21.25	0	1	0	0	1	0	0	11	1	0
10	2454.55	2.20	159.09	18.18	0	1	0	0	0	1	0	20	1	0
11	1333.33	3.00	83.33	18.67	0	1	0	0	0	0	0	7	1	0
12	4166.67	1.20	208.33	20.00	0	1	0	0	0	0	0	15	1	0
13	2000.00	2.50	0.00	20.00	0	0	0	0	0	0	0	18	1	0
14	3000.00	2.00	25.00	19.00	0	1	0	0	0	0	0	7	1	1
15	1700.00	3.00	66.67	20.67	0	1	0	0	0	0	0	14	1	0

### E.2: Raw data for Millet (15 observations)

SN	y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	1000.00	3.00	83.33	18.67	0	1	0	0	0	0	0	60	1	0
2	1333.33	3.00	66.67	21.33	0	1	0	0	1	0	0	30	1	0
3	1500.00	4.00	75.00	26.25	0	1	0	0	0	0	0	35	0	0
4	750.00	2.00	100.00	16.50	0	1	0	0	0	0	0	30	1	0
5	750.00	2.00	0.00	20.00	0	0	0	0	0	0	0	40	1	0
6	1000.00	3.00	0.00	41.67	0	0	0	0	1	0	0	30	1	0
7	400.00	2.50	100.00	44.00	0	1	0	0	0	1	0	22	1	0
8	1066.67	3.00	166.67	20.00	0	1	0	0	0	1	0	6	1	0
9	387.10	3.10	112.90	32.26	0	1	0	0	0	0	0	21	1	0
10	666.67	3.00	83.33	20.67	0	1	0	0	0	0	0	30	1	0
11	333.33	3.00	50.00	20.00	0	0	1	1	0	0	0	15	1	0
12	400.00	1.00	100.00	32.00	0	1	0	1	0	0	0	9	1	0
13	1550.00	2.00	0.00	15.00	0	0	0	0	0	0	0	40	1	0
14	384.62	2.60	0.00	28.46	0	0	0	0	0	0	0	30	1	0
15	362.50	4.00	62.50	17.50	0	1	0	1	0	0	0	30	1	0

# Appendix F: R code Bayesian Model Average

```
# Bayesian Model Averaging Analysis
library(LearnBayes)
#install.packages("BMA")
#loading packages Bayesian Model Avarage
library(BMA)
library(LearnBayes)
n <- nrow(sorgh)
p <- 13
# load packages
library(MCMCpack)
library(lattice)
library(MASS)
y=sorgh$y
x=cbind(sorgh$v1,sorgh$v2,sorgh$v3,sorgh$s1,sorgh$s2,sorgh$s3,sorgh$e1,sorgh$e2,
sorgh$e3,sorgh$e4,sorgh$v6,sorgh$v7,sorgh$v8)
fit1=bicreg(x,y)
summary(fit1)
imageplot .bma(fit1) # image plot of estimated coefficients and models
# plotting Posterior density of Coefficient from BMA
par(mar = c(4, 6,1,1))
par(mfrow=c(3,5))
plot(fit1) # plotting the posterior density of predictors contributions
dev.off()
```

## F.1: Sorghum (BMA analysis)

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	3845.54170	790.927	4432.164	3230.067	4593.132	3424.594	4055.794
X1	100.0	-842.93571	140.829	-903.405	-787.767	-918.213	-805.835	-836.016
X2	14.8	0.27199	0.796	.	.	.	1.985	.
X3	48.9	19.54975	23.827	.	40.846	.	39.457	.
X4	5.5	-51.11749	317.306	.	.	.	.	.

X5	0.0	0.00000	0.000	.	.	.	.	.	.
X6	6.8	27.53170	137.997	.	.	.	.	.	.
X7	5.4	24.14035	152.177	.	.	.	.	.	.
X8	7.3	-32.93968	168.853	.	.	.	.	.	.
X9	29.7	-229.24959	417.020	.	.	.	-783.896	-748.308	.
X10	5.0	49.63519	351.629	.	.	.	.	.	.
X11	1.5	0.06121	1.526	.	.	.	.	.	.
X12	0.0	0.00000	0.000	.	.	.	.	.	.
X13	3.3	9.28027	106.017	.	.	.	.	.	.
nVar			1	2	2	3	2		
r2		0.287	0.312	0.306	0.330	0.302			
BIC		-40.748	-40.738	-39.536	-39.312	-38.729			
post prob		0.166	0.165	0.091	0.081	0.060			
nVar			1	2	2	2	2		
r2		0.091	0.111	0.104	0.103	0.097			
BIC		-10.159	-8.652	-7.339	-7.158	-6.094			
post prob		0.345	0.162	0.084	0.077	0.045			
<b>F.2: Millet (BMA analysis)</b>									
	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5	
Intercept	100.0	2033.1896	394.125	2242.363	1722.257	1877.667	2215.540	2110.004	
X1	89.4	-233.1796	122.437	-227.951	-256.666	-232.521	-319.251	-227.296	
X2	0.0	0.0000	0.000	.	.	.	.	.	
X3	27.8	-3.4938	6.736	.	.	.	-13.278	.	
X4	62.9	657.2941	615.972	.	1191.595	985.481	1162.227	742.170	
X5	63.7	-346.3257	324.750	-683.442	.	-307.734	.	-552.662	
X6	41.5	-234.1125	321.364	-602.007	.	.	.	-471.099	
X7	0.0	0.0000	0.000	.	.	.	.	.	
X8	6.8	-17.2710	84.482	.	.	.	.	.	
X9	0.0	0.0000	0.000	.	.	.	.	.	
X10	22.0	118.4181	276.911	.	.	.	.	.	
X11	5.3	0.4444	2.360	.	.	.	.	.	
X12	0.0	0.0000	0.000	.	.	.	.	.	

X13	12.1	-41.9622	140.755	.	.	.	.	.	.
nVar			3	2	3	3	4		
r2			0.203	0.163	0.199	0.198	0.229		
BIC			-8.145	-8.008	-7.585	-7.466	-6.756		
post prob			0.095	0.089	0.072	0.068	0.048		

## Appendix G: R code Bayesian Model Selection Using Zellner's g-prior

```
# Set working directory
rm(list=ls())
setwd('C:/Users/Adeyemi/Desktop/dddd')
# Read data
sorgh <- read.csv('SORGHUMS.csv')
str(sorgh) # structure of your data (no. of obs, variables,...)
head(sorgh) # view part of your data
tail(sorgh)
# Loading packages
library(MCMCpack)
library(coda)
library(lattice)
library(LearnBayes)
```

### G .1: Bayesian Model Selection for Sorghum

```
# *****
x=cbind(sorgh$v2-mean(sorgh$v2),sorgh$v3-mean(sorgh$v3),
+ sorgh$s3,sorgh$s4, sorgh$v6-mean(sorgh$v6),sorgh$v8) # Sorghum2
> y=sorgh$y; c=100
> bayes.model.selection(y,x,c,constant=FALSE)
      log.m  Prob   NA   NA   NA   NA   NA   NA
1  FALSE FALSE FALSE FALSE FALSE FALSE -1225.49 0.00000
2  TRUE  FALSE FALSE FALSE FALSE FALSE -1205.51 0.24747
3  FALSE  TRUE  FALSE FALSE FALSE FALSE -1222.31 0.00000
4  TRUE  TRUE  FALSE FALSE FALSE FALSE -1206.43 0.09923
5  FALSE FALSE  TRUE  FALSE FALSE FALSE -1218.74 0.00000
6  TRUE  FALSE  TRUE  FALSE FALSE FALSE -1205.46 0.26215
7  FALSE  TRUE  TRUE  FALSE FALSE FALSE -1218.10 0.00000
8  TRUE  TRUE  TRUE  FALSE FALSE FALSE -1206.85 0.06531
9  FALSE FALSE FALSE  TRUE  FALSE FALSE -1227.67 0.00000
10 TRUE  FALSE FALSE  TRUE  FALSE FALSE -1207.21 0.04521
11 FALSE  TRUE  FALSE  TRUE  FALSE FALSE -1224.62 0.00000
12 TRUE  TRUE  FALSE  TRUE  FALSE FALSE -1208.44 0.01327
13 FALSE FALSE  TRUE  TRUE  FALSE FALSE -1220.82 0.00000
14 TRUE  FALSE  TRUE  TRUE  FALSE FALSE -1207.13 0.04925
```

```

19  FALSE TRUE FALSE FALSE TRUE FALSE -1223.83 0.00000
20  TRUE TRUE FALSE FALSE TRUE FALSE -1208.50 0.01246
26  TRUE FALSE FALSE TRUE TRUE FALSE -1209.26 0.00587
27  FALSE TRUE FALSE TRUE TRUE FALSE -1226.21 0.00000
28  TRUE TRUE FALSE TRUE TRUE FALSE -1210.37 0.00192
29  FALSE FALSE TRUE TRUE TRUE FALSE -1221.98 0.00000
30  TRUE FALSE TRUE TRUE TRUE FALSE -1208.93 0.00810

```

## G .2: Bayesian Model Selection for Millet

```

# *****
# loading data
attach(sorgh)
#x=cbind(sorgh$v1-mean(sorgh$v1),sorgh$v2-mean(sorgh$v2),sorgh$v3,sorgh$v6)
x=cbind(sorgh$v1-mean(sorgh$v1),sorgh$v3-mean(sorgh$v3),sorgh$s1,sorgh$s2, sorgh$s3)
y=sorgh$y; c=25
bayes.model.selection(y,x,c,constant=FALSE)
y=sorgh$y; c=50
bayes.model.selection(y,x,c,constant=FALSE)
y=sorgh$y; c=100
bayes.model.selection(y,x,c,constant=FALSE)

$converge
> y=sorgh$y; c=100
> bayes.model.selection(y,x,c,constant=FALSE)
$mod.prob
  log.m Prob  NA  NA  NA  NA  NA
1  FALSE FALSE FALSE FALSE FALSE -767.37 0.00371
2  TRUE  FALSE FALSE FALSE FALSE -765.69 0.01989
3  FALSE  TRUE  FALSE FALSE FALSE -769.40 0.00049
4  TRUE  TRUE  FALSE FALSE FALSE -766.17 0.01231
5  FALSE FALSE  TRUE  FALSE FALSE -765.41 0.02643
6  TRUE  FALSE  TRUE  FALSE FALSE -763.80 0.13149
7  FALSE  TRUE  TRUE  FALSE FALSE -767.46 0.00341
8  TRUE  TRUE  TRUE  FALSE FALSE -764.34 0.07681
9  FALSE FALSE FALSE  TRUE  FALSE -765.28 0.03002
10 TRUE  FALSE FALSE  TRUE  FALSE -764.44 0.06969
16 TRUE  TRUE  TRUE  TRUE  FALSE -765.25 0.03085
26 TRUE  FALSE FALSE  TRUE  TRUE  -763.66 0.15167
27 FALSE  TRUE  FALSE  TRUE  TRUE  -766.77 0.00679
28 TRUE  TRUE  FALSE  TRUE  TRUE  -765.00 0.03970
29 FALSE FALSE  TRUE  TRUE  TRUE  -765.36 0.02771
30 TRUE  FALSE  TRUE  TRUE  TRUE  -764.46 0.06816

```

```
31 FALSE TRUE TRUE TRUE TRUE -767.63 0.00285
32 TRUE TRUE TRUE TRUE TRUE -765.72 0.01941
# *****CODE END*****
```