

**OPTIMIZING MULTIPLE REGRESSION MODEL
FOR RICE PRODUCTION FORECASTING IN
NEPAL**



**A THESIS SUBMITTED TO THE
CENTRAL DEPARTMENT OF STATISTICS
INSTITUTE OF SCIENCE AND TECHNOLOGY
TRIBHUVAN UNIVERSITY
NEPAL**

**FOR THE AWARD OF
DOCTOR OF PHILOSOPHY
IN STATISTICS**

**BY
CHUDA PRASAD DHAKAL
JUNE 2015**

DECLARATON

This thesis entitled “**Optimizing multiple regression model for rice production forecasting in Nepal**” which is being submitted to the Central Department of Statistics, Institute of Science and Technology (IOST), Tribhuvan University, Nepal for the award of the degree of Doctor of Philosophy (Ph.D.), is a research work carried out by me under the supervision of Prof. Dr. Azaya Bikram Sthapit, Central Department of Statistics, Tribhuvan University and co-supervised by Prof. Dr. Naba Raj Devkota.

This research is original and has not been submitted earlier in part or full in this or any other form to any university or institute, here or elsewhere, for the award of any degree.

Chuda Prasad Dhakal

RECOMMENDATION

This is to recommend that **Chuda Prasad Dhakal** has carried out research entitled “**Optimizing multiple regression model for rice production forecasting in Nepal**” for the award of Doctor of Philosophy (Ph.D.) in **Statistics** under my/our supervision. To my /our knowledge, this work has not been submitted for any other degree.

He has fulfilled all the requirements laid down by the Institute of Science and Technology (IOST), Tribhuvan University, Kirtipur for the submission of the thesis for the award of Ph.D. degree.

Prof. Dr. Azaya Bikram Sthapit
Supervisor
Professor
Central Department of Statistics
Tribhuvan University
Kirtipur, Kathmandu,
Nepal

Prof. Dr. Naba Raj Devkota
Co-Supervisor
Director
Directorate of Research and Extension
Agriculture and Forestry University
Rampur, Chitwan,
Nepal

[June 2015]



TRIBHUVAN UNIVERSITY
CENTRAL DEPARTMENT OF STATISTICS
OFFICE OF THE HEAD OF DEPARTMENT

Kirtipur, Kathmandu
NEPAL

Ref. No.

Date:

Subject:

LETTER OF APPROVAL

18/06/2015

On the recommendation of Prof. Dr Azaya Bikram Sthapit / Naba Raj Devkota , this Ph. D. thesis submitted by **Chuda Prasad Dhakal**, entitled “**Optimizing Multiple Regression Model for Rice Production Forecasting in Nepal**” is forwarded by Central Department Research Committee (CDRC) to the Dean, IOST, T.U..

.....
Shankar Prasad Khanal
Professor,
Head,
Central Department of Statistics,
Tribhuvan University
Kirtipur, Kathmandu
Nepal

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my Principal Supervisor, Prof. Dr. Azaya Bikram Sthapit, outgoing Head of the Central Department of Statistics, Tribhuvan University, Kirtipur College for his invaluable advice and guidance during the five years of my study. I also appreciate the support of my Associate Supervisor: Prof. Dr Naba Raj Devkota, Agriculture and Forestry University Chitwan, for his tireless guidance for the completion of this study. And, I express my gratitude and I am very much grateful to Prof. Dr. Ramchandra Singh, Head of the Central Department of Statistics for his every official support and the high valued encouragement throughout the study.

I would like to thank Institute of Agriculture and Animal Sciences (IAAS) Rampur, Chitwan for granting me study leave to participate in the doctoral research programme. My research work included analysis of the time series data to investigate a multiple regression model for rice production forecasting in Nepal. I wish to thank International Rice Research Institute (IRRI), Food and Agriculture Organization Statistics (FAOSTAT), Ministry of Agriculture (MOA) and Department of Hydrology and Metrology (DHM) for I have used the data from these institutions in this study.

My sincere thank goes to former Department heads, Prof. Dr. Mrigendra Lal Singh and Prof. Dr. Devendra Bahadur Chhetry who provided academic assistance to the several stages of my study. I wish to thank all faculties (particularly Prof. Dr. Tika Ram Aryal, Prof. Dr. Shankar Prasad Khanal, Prof. Dr. Srijan Lal Shrestha, and Dr. Ram Prasad Khatiwada) and to all staff of the Central Department of Statistics Kirtipur for their academic and

administrative assistance. To add on this list my sincere thanks goes also to Mr. Bholanath Singh Basnet, rice expert Nepal.

I thank my friends Mr. Alan McWilliam (UK), Dr Patrick Bucker (Germany) and Philips Hoyle (Hungary), who helped me in providing some of the Book and journal articles which remained of very much importance for the study. I would like to thank Mr Bholanath Dhakal, director of Aadharshree College Nayabazar, Kathmandu for his enthusiasm and worthy cooperation, and to all of the colleagues there for providing me a friendly environment during my thesis write up period there at the Aadharshree Education Foundation (AEF) Nayabazar, Kathmandu.

Last but not the least, my sincere thanks to goes my parents, Late Basundhara Dhakal and Khageswor Dhakal , my wife Purna Dhungana, and sons Abbal Dhakal and Aviral Dhakal, for their continuing support and patience during this period. Thank you for encouraging and believing in me, which helped me to pursue my research study towards the end.

(Chuda Prasad Dhakal)

June, 2015

ABSTRACT

This research, testing the possibility of use of probable predictors, has optimized multiple regression model to be used for rice production forecasting in Nepal. Fifty years (1961-2010) time series data were divided into training sample (a sample which is used to build the model) (n=35), and test sample size of 15 through which the built model was cross validated for its reliability in forecasting.

This research has explored and used all the underlying principles of linear regression model building and its application in forecasting the production, mainly crop production such as rice. The model sustained with the three principle predictors: *harvested area*, *rural population* and *price at harvest* whereas these variables could explain 93% variation in *production*; the forecast variable. The model as such was parsimonious and as well the good fit with minimal (5%) mean absolute percentage error in its forecast.

It therefore, for this fit, was concluded that multiple regression model could be scientifically used in forecasting, and the concerned stakeholders could thus be benefited from the this model especially for the enhanced ease, and efficiency for rice production forecasting to be used for planning purpose at national level.

Future work might consider to increase the precision of the model in any aspects like making it more parsimonious and reliable than which have been purposed in this study.

LIST OF ACRONYMS AND ABBREVIATIONS

ME	Mean Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
IRRI	International Rice Research Institute
FAO	Food and Agriculture Organization
DHM	Department of Hydrology and Meteorology
TS	Tracking Signal
PCR	Principle Component Analysis
PLS	Partial Least Square
GLM	Generalized Regression Model
GAM	Generalized Additive Model
NASS	National Agriculture Statistics Service
ANN	Artificial Neural Network
NN	Neural Network
AIC	Akaikis Information Criteria
AICc	Corrected Akiaike's Information Criteria
BIC	Basian Information Criteria
DW	Durbin - Watson
OLS	Ordinary Least Square

sMAPE	'symmetric' Mean Absolute Percentage Error
PE	Percentage Error
MSE	Mean Squared Error
MAD	Mean Absolute Deviation
PRESS	Prediction Sums of Square
VIF	Variance Inflation Factor

LIST OF TABLES

	Page No.
Table 1: List of potential predictors	66
Table 2: SPSS forward selection	67
Table 3: Minitab forward selection	67
Table 4: SPSS backward selection	68
Table 5: Minitab backward selection	68
Table 6: SPSS stepwise selection	69
Table 7: Minitab stepwise selection	69
Table 8: List of the potential predictors	69
Table 9: Best subset regression	71
Table 10: List of appropriate predictors	72
Table 11: Correlation matrix	76
Table 12: Lack of fit test	79
Table 13: Breush-Pagon test	79
Table 14: Durbin-Watson test	81
Table 15: Shaprio-Wilk test	82
Table 16: VIF for the predictors	83
Table 17: Model after rural population was removed	84

Table 18: Description of unusual observations	84
Table 19: Detecting influential observations	85
Table 20: Model with and without the outliers	86
Table 21: Forecast accuracy measures	87
Table 22: Comparison of forecast model	89
Table 23: Goodness of fit statistics	91
Table 24: Zero- order partial and part correlations	93
Table 25: Regression outputs (Analysis of variance and the coefficients)	95
Table26: Possible and the appropriate list of the predictors	102

LIST OF FIGURES

	Page No.
Figure 1: Scatter plot of production versus harvested area	73
Figure 2: Scatter plot of production versus rural population	74
Figure 3: Scatter plot production versus price at harvest	75
Figure 4: Plot of residual versus fitted values	78
Figure 5: Plot of standardized residual versus observation order	80
Figure 6: Histogram of standardized residuals	82

TABLE OF CONTENTS

	Page No.
	i
Declaration	ii
Recommendation	iii
Certificate of Approval	iv
Acknowledgements	v
Abstract	vii
List of Abbreviations	viii
List of Tables	x
List of Figures	xii
CHAPTER 1: INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 OBJECTIVE	3
1.3 RATIONALE	4
1.4 RESEARCH QUESTIONS	5
1.5 SCOPE LIMITATIONS AND ASSUMPTION	6
1.6 METHODOLOGY	7
1.7 ORGANIZATION OF THE STUDY	9
CHAPTER 2: LITERATURE REVIEW AND THEORITICAL FRAMEWORK	10
2.1 LITERATURE REVIEW	10
2.1.1 Statistics and regression	11
2.1.2 Regression methods	13
2.1.3 Regression methods and forecasting	17

2.2 THEORITICAL FRAMEWORK	20
2.2.1 Linear regression theory	21
2.2.2 Linear regression as statistical modeling	23
2.2.3 Developing linear regression model for forecasting	25
CHAPTER 3: MATERIALS AND METHODS	57
3.1 MATERIALS	58
3.2 METHODS	59
CHAPTER 4: RESULTS AND DISCUSSION	64
4.1 VARIABLE SELECTION	66
4.2 OPTIMIZING OF THE MODEL	72
4.3 OUT OF SAMPLE CROSS VALIDATION OF THE MODEL	86
4.4 COMPARISON OF THE MODEL WITH A BENCHMARK	89
4.5 TEST OF GOODNESS OF FIT OF THE MODEL	90
4.6 INTERPRETATION OF THE MODEL INVESTIGATED	94
4.7 LIMITATION OF THE STUDY	99
CHAPTER 5: SUMMARY AND CONCLUSION	101
5.1 SUMMARY	101
5.2 CONCLUSIONS	103
5.3 RECOMMENDATIONS FOR FUTURE WORK	105
REFERENCES	106
APPENDIX	117