

**EFFICIENCY AND PRODUCTIVITY ESTIMATION
WITH A HIERARCHICAL DATA STRUCTURE:
AN AGGREGATION PERSPECTIVE**

Dissertation

to attain the degree Doctor of Philosophy (Ph.D.)
of the Faculty of Agricultural Sciences
Georg-August-Universität Göttingen

Submitted by

Yashree Mehta

born on 19th December 1987 in Ahmedabad, India

Göttingen, April 2020

First evaluator: Prof. Dr. Bernhard Bruemmer

Second evaluator: Prof. Dr. Stephan von Cramon-Taubadel

Additional examination board member: Prof. Dr. Helmut Herwartz

Date of oral examination: May 19, 2020

0.1 Joint summary

Repeated measurements on a subject of interest are statistically preferable than cross-sectional data due to confoundment of effect estimates in the latter. The confoundment is brought about due to between-subject heterogeneity which arises from the possibility that repeated measurements on the same subject tend to be more similar (as indicated by high correlation) as compared to measurements across different subjects. The method of hierarchical modelling entails varying coefficients by subject in order to account for the heterogeneity in carrying out causal inference. This dissertation unites hierarchical modelling with efficiency estimation by formulating a methodological research objective in the first paper. It verifies the potential of hierarchical modelling in efficiency aggregation by using hierarchical data. Further, in the second paper, the role of temporal land use and fertilizer as an input have been studied in determining the scope for sustainable intensification, with the use of hierarchical modelling. The third paper applies Stochastic Frontier Analysis for estimating the production frontier and technical efficiency with the aim of studying the elasticity of inputs. The modelling context as well as the results of the second and third paper are specific to paddy production in Bangladesh.

The first paper is a comparison of two independent modelling approaches in the aggregation of technical efficiency scores. The first approach is an estimation of a stochastic frontier model at the lowest level of observation (a plot). Technical efficiency scores obtained at this level are aggregated to the next level of observation (a farm) by applying composite indices such as the arithmetic mean and its weighted variants. The second approach is the direct estimation of technical efficiency scores at the farm level using a hierarchical model. With Monte Carlo simulation, the paper concludes that while both approaches can preserve the ranking of the true efficiency scores, it is only the hierarchical model which accurately estimates the level of the true efficiency scores.

The second paper is an empirical study of the potential of temporal land use and its interaction with fertilizer in bringing about sustainability in intensification as against the choice of keeping the plot fallow in paddy cultivation in Bangladesh. The question of whether fertilizer as an input, conditional on land use choice, is instrumental in ensuring sustainable intensification, is addressed with the application of hierarchical models. Specifically, the Random Intercept

and Random Coefficients Model (with cross level interaction) have been used to conclude that potato cultivation in Rabi season increases paddy production in the succeeding Boro season and also enhances the effect of fertilizer as an input in driving paddy production.

The third paper is an application of the Stochastic Frontier model to a large sample of plots which cultivate paddy in the immediate time span following the food price crisis of 2007-08. The estimation and inference indicated that since most of the plots operated close to the production frontier, there is a need to bring about a shift in the production frontier to increase production. Increased fertilizer and labour application is expected to bring gains in the quantity of output.

ACKNOWLEDGEMENTS

This research was supported by the German Research Association (DFG) Research Training Group 1644 "Scaling Problems in Statistics", grant no. 152112243.

I feel honoured to have been chosen for the prestigious position which I have held, of a doctoral candidate of the Research Training Group (RTG), "Scaling Problems in Statistics", and at the Chair of Agricultural Market Analysis, at University of Goettingen.

I am deeply grateful to Prof. Dr. Bernhard Bruemmer, for giving the opportunity to carry out doctoral research under his supervision, and his guidance and support as I worked on it. I thank Prof. Dr. Helmut Herwartz for his support in the same. I also thank Prof. Dr. Stephan von Cramon-Taubadel for being a part of my thesis committee and supporting my Ph.D.

My sincere gratitude to Ms Priscilla Wainaina for allowing me to use her data, collected in Kenya, for the purpose of analysis in the first paper (Chapter 2).

I thank Mr. Shalander Kumar and Mr. Eliaskhan Patan from ICRISAT, Hyderabad, India for helping me to gain a sound understanding of the data related to Bangladesh.

Ms Barbara Strauss and Ms Doerte Dede have facilitated my administrative tasks and ensured a smooth completion of my PhD program at the RTG. Likewise, Ms Nina Enke has helped me at the Chair. I thank them profusely. I also thank the staff at DARE-IT as they were readily available when I needed them many times.

My fellow colleagues at the Chair have been the best people I have dealt with as they have encouraged me and helped me in every possible way during the course of the program. I cherish their warmth and optimism. I also acknowledge the technical help which I received from my RTG colleagues, Peter Kramlinger and Paola Barajas.

My friends in Goettingen have made my experience here, the most memorable and enjoyable. I thank them for their warmth.

My father, mother and brother are the reason why I have come a long way in my studies. Their love and support have constantly nurtured all my academic pursuits. I will be eternally grateful to them.

I have learnt immensely in terms of gaining academic knowledge as well as imbibed noble values during the course of my stay at University of Goettingen. This dissertation is a manifestation of my immemorial want to learn boundarylessly, from academic study, and from people. I dedicate it to the pursuit of knowledge and excellence.

CONTENTS

0.1	Joint summary	iii
	Acknowledgements	v
	List of Tables	ix
	List of Figures	x
	Abbreviations	xi
1	General Introduction	1
2	Efficiency Aggregation in Stochastic Frontier Analysis with Hierarchical Data	4
2.1	Introduction	5
2.2	Methodological background	8
2.3	Methodology	10
2.3.1	Plot-level Stochastic Frontier Model	11
2.3.2	The Linear Mixed Effects Model	13
2.4	Variable Description	14
2.5	Data	16
2.6	Monte Carlo simulation	20
2.7	Empirical application	23
2.8	Conclusion	27
3	The role of temporal land use in driving sustainable intensification in Bangladesh: An application of hierarchical modelling	30
3.1	Introduction	31
3.1.1	Modelling context	33
3.2	The context of crop cultivation in Bangladesh	34
3.2.1	The physiography of Bangladesh	34
3.2.2	Exploratory analysis of the data	35
3.3	Methodology	38
3.3.1	Random Intercept model with interaction effects	39
3.3.2	Random Coefficients Model with cross-level interaction	40
3.4	Data	41
3.5	Results	46
3.6	Conclusion	52

4	Stochastic Frontier Analysis of Paddy Production in Bangladesh	54
4.1	Introduction	55
4.2	Methodology	55
4.3	Data	56
4.4	Results	59
4.5	Conclusion	64
5	General conclusion	67
5.1	Universal Conclusion	67
5.2	Limitations	67
5.3	Scope for further research	68
A	Chapter 1 of appendix	70

LIST OF TABLES

2.1	Count of plots by AEZ	17
2.2	Count of observations by season	17
2.3	Count of households by plot ownership	18
2.4	Descriptive Statistics	19
2.5	Monte Carlo simulation statistics for $\mu=0$	22
2.6	Monte Carlo simulation statistics for $\mu=2$	22
2.7	Plot stochastic frontier and LME estimates	24
2.8	Plot stochastic frontier and LME estimates (contd.)	25
2.9	Estimates of coefficients of plot-invariant variables	27
2.10	Model-specific estimates	28
3.1	Count of observations by group in the sample	42
3.2	Summary statistics of continuous variables	44
3.3	Percentage distribution of categorical variables	45
3.4	Estimates of the Random Intercept model	48
3.5	Variance estimates of the random intercept model	49
3.6	Estimates of the Random Coefficients model	50
3.7	Variance estimates of the random coefficient model	51
4.1	Summary statistics of continuous variables	57
4.2	Translog estimates for total production	62
4.3	Translog estimates for total production (contd.)	63
A.1	Count of observations by aggregated AEZs	70

LIST OF FIGURES

3.1	The distribution of plot size across all crops	36
3.2	The distribution of the proportion of irrigated plot area	36
3.3	Flexible rotation, cyclical, fixed rotation length	38
3.4	Caterpillar plot of Random effects of the RCM	52
4.1	Scatter plot of inputs with respect to Production	58
4.2	The distribution of rainfall by the seasons	59
4.3	Output elasticity of inputs	64
4.4	Output elasticity of rainfall by season	65
4.5	Kernel density plot of technical efficiency scores	65
4.6	Kernel density plot of scale elasticity	66

ABBREVIATIONS

AEZ	Agro-Ecological Zone
AM	Arithmetic Mean
BARC	Bangladesh Agricultural Research Council
BLUP	Best Linear Unbiased Predictor
CDF	Cumulative Density Function
cms.	centimeters
DAP	Diammonium Phosphate
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
DTWs	Deep Tubewells
FE	Fixed Effects
GBM	Ganges-Brahmaputra-Meghna
GM	Geometric Mean
HH	Household
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
Kg.	Kilograms
LLPs	Low Lift Pumps

LME	Linear Mixed Effects
MAR	Missing At Random
mm	Millimeters
No.	Number
NPK	Nitrogen, Phosphorous and Potassium
PSU	Primary Sampling Unit
RCM	Random Coefficients Model
RE	Random Effects
RI	Random Intercept
SD	Standard Deviation
SFA	Stochastic Frontier Analysis
STWs	Shallow Tubewells
TE	Technical Efficiency
VDSA	Village Dynamics of South Asia
WAM	Weighted Arithmetic Mean
WGM	Weighted Geometric Mean

CHAPTER 1

GENERAL INTRODUCTION

Stochastic Frontier Analysis (SFA), as a method of efficiency and productivity estimation, is suitable for application and inference at the level of a micro unit of analysis, which is a Decision Making Unit (DMU). The method uses data on production, mainly consisting of input usage and output. In the agricultural sector, surveys are conducted for procuring data regarding different aspects of production. A typical farm survey comprises of random sampling of households involved in the production of a particular crop in a given region. However, given the spatial dimension of agricultural production, random sampling of farm households is a process coupled with the need for complete representability of the households in the sample. This calls for consideration of the multiple levels of clustering of higher-level of units (to which the households may belong) before one can select the households. The villages to which the households belong make up the Primary Sampling Unit (PSUs). In many cases, Agro-Ecological Zones (AEZs) of the concerned region are delineated before the villages are sampled from each of them. Further, ownership of multiple plots by farm households extends the clustering in the opposite direction, with plots making up the lowest unit of observation. This sampling procedure gives rise to a hierarchical data structure, wherein, each input-output combination at the plot level can be traced to its farm household, village, and AEZ. According to Fitzmaurice et al. (2011), a hierarchical structure in data can emerge due to study design. The sampling procedure adopted to maintain representability of households in the sample is one case of study design which leads to a hierarchical data structure and this structure has to be addressed in statistical modelling for estimation of efficiency and productivity.

Just and Pope (2001) explain the atomistic, yet heterogeneous nature of the units of analysis involved in agricultural production. An aggregation of the output as a function of inputs fails to satisfy the regularity conditions and renders an unresolved question of incorporating stochasticity in the model. Stochasticity is an inherent characteristic of agricultural production, arising due to variation in soil, weather, and sometimes output and input prices. A major problem faced with aggregation of production technology is loss of information contained at the lower level unit. For example, plots exhibit varied geomorphological environments and the need for accounting for each of their distinct features is a matter of concern during aggregation. From a general perspective, the problems faced due to aggregation have been discussed in the theory of ecological inference. King (1997) discusses the aggregation bias which is caused when the information lost in the process of aggregation is systematic and characterised by selectivity.

Given hierarchy in the data structure, the potential of hierarchical modelling in accurately measuring technical efficiency is to be examined – a methodological objective. Chapter 2, "Efficiency Aggregation in Stochastic Frontier Analysis with Hierarchical data", addresses the aforementioned methodological objective.

One merit of hierarchical modelling is that it accounts for between-cluster heterogeneity and within-cluster covariance in a hierarchical data structure. The physiography of Bangladesh, given its extreme diversity across the landmass, offers itself to the application of hierarchical modelling in the assessment of farm efficiency and input productivity. Models such as the random coefficients model (a hierarchical model) are suitable for capturing heterogeneity at the desired units of analysis. Varying a coefficient by cluster is expected to explain

the differential effect of an input of interest on the output, thereby producing a cluster-specific causal relationship. In Bangladesh, such an analysis is of particular importance in the context of the effect of climate change. Excessive usage of environmentally detrimental inputs such as fertilizer, coupled with erratic rainfall patterns, and physiographic diversity make a sure case for accounting for maximum heterogeneity which may exist across spatial and temporal clusters.

Chapter 3, "The role of temporal land use in driving sustainable intensification in Bangladesh: An application of hierarchical modelling", studies the scope for sustainable intensification in paddy cultivation through the interactive roles of land use and fertilizer application.

Bangladesh, as a country which has been projected as susceptible to climate change, also faces pressure of increased population on land. This gives rise to the challenge of arranging for food security which can be brought about, partially, through increased production of its staple crop, paddy. Therefore, there is a need to estimate and infer about the extent to which farmers are meeting their true potential in production. Stochastic Frontier Analysis, as a parametric approach to efficiency estimation, facilitates the fitting of a frontier to data on production.

Chapter 4, "Stochastic Frontier Analysis of Paddy Production in Bangladesh" carries out efficiency analysis of paddy cultivation in Bangladesh.

Chapter 5 presents the overall conclusion of the dissertation, its limitations, and scope for further research.

CHAPTER 2
EFFICIENCY AGGREGATION IN STOCHASTIC FRONTIER ANALYSIS
WITH HIERARCHICAL DATA

Abstract

Data regarding agricultural production often have a natural hierarchical structure. Ownership of multiple plots by a farmer is one such case. When there is more than one level of observation at which technical efficiency can be estimated, the process of its aggregation from a micro unit of analysis to a higher, aggregate level, poses a topic for a methodological debate. Having used Stochastic Frontier Analysis on data concerning maize production, with a hierarchical structure, we compare scaling up of technical efficiency scores from a plot-level stochastic frontier model, and the Linear Mixed Effects model. With Monte Carlo simulation, we conclude that if monotonicity in the ranking of farm households is to be preserved, the Linear Mixed Effects Model performs slightly better than aggregation indices applied after plot-level estimation. In maintaining the Cumulative Density Function of the true aggregated efficiency scores, unlike aggregation indices, the Linear Mixed Effects Model performs accurately.

Keywords : Technical efficiency, Aggregation, Hierarchical modelling

2.1 Introduction

A hierarchical data structure is commonplace in agricultural surveys. Ownership of more than one plot by a farm household is one case, where, the collected data on production has a hierarchical structure. Notwithstanding that the data is cross-sectional, without a time dimension, a farm household can form a cluster, with data on the plots owned by it assuming the role of repeated measurements. Output and input usage are directly measured at the plot-level. Data on socio-economic variables which determine the efficiency of the household as a producer, are collected at the farm-level. These variables do not vary across plots which belong to the same household but vary between households. Plot-level technical efficiency can be directly estimated by classifying the production inputs as the explanatory variables of the frontier, and the household-level inputs as Z-variables, by using Stochastic Frontier Analysis (SFA).

Theoretically, the production frontier is attached to the notion of a producer as a Decision Making Unit (DMU). Quoting from Fried et al. (1993),

“..in practice, one has only data—a set of observations for each decision-making unit (DMU) corresponding to achieved output levels for given input levels”. (p. 121)

The inefficiency term in SFA captures the effect of managerial ability of the concerned DMU (here, the producer). Thus, estimation of technical efficiency should ideally produce the efficiency score of the producer. Given the naturally occurring hierarchical structure of survey data regarding production, there arises an estimation and inference anomaly with respect to technical efficiency. The estimated efficiency at the plot-level by SFA does not fit in the otherwise unanimous understanding of “efficiency”. Plot-level efficiency is, thus, difficult

to interpret and cannot directly be used to conclude about the performance of a DMU. This calls for a methodologically sound aggregation strategy for navigating from plot-level efficiency estimates to the higher level in the data hierarchy, who, in this case are the producers/farm households. One may also intend to make policy recommendations which affect DMUs so as to facilitate an improvement in their efficiency and this requires information about their efficiency performance.

Cook et al. (1998) recognize the need for an aggregation method in efficiency estimation when DMUs occur naturally in groups. They distinguish between pure hierarchies and levels. The former refers to hierarchies formed based on a particular attribute of the DMUs whereas the latter refers to groupings at one level which may be formed based on multiple attributes. They propose a method for synchronizing DMU ratings based on relative efficiency according to those received by their respective groups in the application of Data Envelopment Analysis (DEA). Blackorby and Russell (1999) extend this approach by deriving conditions under which efficiency index aggregation can be carried out consistently across different levels of DMUs, using DEA.

Brorsen and Kim (2013) examine the consequences of data aggregation on the estimation of a stochastic frontier, when dealing with hierarchical data, using a cost function approach. In the light of the closed skew normal being the true distribution of the aggregated data, they find that misspecification caused by using the standard stochastic frontier model leads to an incorrect conclusion of diseconomies of scale and higher inefficiency of smaller units. The hierarchical structure studied by them concerns aggregation across DMUs, whereas our study, which is regarding multiple plots per farm, is concerned with a hierarchy across parts of a DMU.

The need for consistent efficiency estimation at different levels of observation in a hierarchical data structure gives rise to the need for a method of aggregation. Using hierarchical maize production data of smallholders in Kenya, we address the unresolved question of being able to infer about aggregate efficiency from lower-level estimates from a data hierarchy, using SFA, by specifying two models: plot-level stochastic frontier model and the Linear Mixed Effects (LME) model. We verify if a fundamentally correct distribution of efficiency can be arrived at, at the household level from plot-level estimates of technical efficiency. We compare the performance of the plot stochastic frontier model and the LME model in deriving efficiency estimates at an aggregate level, with respect to estimating the true scores, using Monte Carlo simulation. We also examine the role of the plot-level statistical error term in maintaining robustness of the aggregation process.

The rest of the paper is organized as follows: In Section 2.2, the methodological background concerning hierarchical modelling is provided. Section 2.3 presents the methodology used for aggregation in the two models, after a description of their specification. Section 2.4 provides a description of the variables used in the study. In Section 2.5, we present the characteristics of the data used. Section 2.6 describes the procedure followed for Monte Carlo simulation, along with the result obtained from it. In Section 2.7, we present the empirical application of the two models on the existing data, along with a discussion of its result. Section 2.8 presents the conclusion.

2.2 Methodological background

Data arising from repeated measurements of different plots belonging to the same farmer has a clustered structure. It is a data structure in which a unit of observation is nested within another higher-level unit of observation. Such data can be analyzed using a cross-sectional multilevel modelling approach. In multilevel modelling, estimation and inference at one level of an observed unit often depends on the estimation and inference of parameters (random coefficients) at a higher level. It is this property of conditional modelling due to which it is also known as "hierarchical" (Gelman and Hill, 2007). It retains the identity of being cross-sectional but has a hierarchical structure.

Clark (2016) provides an explanation of the different modelling approaches in clustered data analysis. There are several terms used for models applied to clustered data: Variance components, Random Intercept, Random effects, Hierarchical model, Multilevel model, Mixed models, and so on. These refer to similar modelling approaches, viewed from a different objective of analysis and treatment of the random components. To start with, the classical linear model is expressed in terms of the data generating process, before considering the clustered structure of the data. Having introduced the cluster structure, a model is specified in which the coefficients vary by cluster. These coefficients can be varying intercepts as well as varying slopes. Since our study uses varying intercepts and not varying slopes, we focus on the typology of clustered data analysis with respect to a random intercepts model.

We use the "Multilevel model" classification (Clark, 2016), which is given as follows:

$$y_{ij} = \alpha_i + \beta X_{ij} + \epsilon_{ij} \quad (2.1)$$

$$\alpha_i = \beta_o + \gamma_{oi} \quad (2.2)$$

$$\gamma_{oi} \sim N(0, \tau^2)$$

In equation (2.1), y_{ij} is the dependent variable, which in our case is the output. x_{ijk} is the k^{th} covariate and the β s are the estimated coefficients corresponding to the covariates. γ_{oi} measures the extent to which the cluster i differs in its "base" level of production from the population fixed intercept β_o . Thus, the sum of β_o and γ_{oi} is the cluster-specific random intercept, denoted by α_i . The random deviation γ_{oi} is assumed to follow a normal distribution with zero mean and variance τ^2 . ϵ_{ij} is the random error term.

Substituting equation (2.2) in equation (2.1), we get equation (2.3) as follows:

$$y_{ij} = (\beta_o + \gamma_{oi}) + \beta X_{ij} + \epsilon_{ij} \quad (2.3)$$

Alternatively, the random deviation γ_{oi} can be summed with the random error term, ϵ_{ij} as given in equation (2.4).

$$y_{ij} = \beta_o + \beta X_{ij} + (\gamma_{oi} + \epsilon_{ij}) \quad (2.4)$$

Equations (2.3) and (2.4) correspond to the different treatment of the random component γ_{oi} in terms of the purpose of modelling. Equation (2.3) considers the random deviation as a variable of interest which forms the cluster-specific intercept. Equation (2.4) regards the random deviation as a nuisance parameter. In our study, we will be using cluster-specific intercepts for estimating technical efficiency. Hence, random effect is a substantive parameter, and we use the Multilevel model specification in equations (2.1) and (2.2).

Since we are interested in cluster-specific random intercepts, we acknowledge their process of prediction as the Best Linear Unbiased Predictor (BLUP).

Since the random deviation is a random variable, we “predict” them instead of “estimating” them. Its prediction is about realizing its conditional mean, based on the data at hand. For an explanation of the method of prediction of the random effects, one can refer to Fitzmaurice et al. (2011).

The multilevel modelling approach can also be viewed as a combination of regressions, conditional and marginal model, correlated error model, multivariate normal model, penalized regression and Bayesian mixed model. For a complete overview of the different modelling approaches, one can refer to Clark (2016).

2.3 Methodology

A variant of clustered data is longitudinal data, wherein, the ordering of the repeated measures is to be preserved for analysis (Fitzmaurice et al., 2011). Several methods of longitudinal data analysis come under the purview of those which are used in the more general case of clustered data, as given in Fitzmaurice et al. (2011).

Various models have been proposed for efficiency estimation with longitudinal data in SFA. Schmidt and Sickles (1984) provide a framework for estimating the production frontier, wherein, inefficiency is assumed to be time-invariant. It can be estimated by way of fixed effects or random effects. Models addressing time-varying efficiency estimation with longitudinal data were proposed thereafter - each building upon the previous in order to separate inefficiency, as distinct from heterogeneity. Inefficiency estimation has further been bifurcated into persistent and transient – giving rise to another class of models which estimate it.

We estimate technical efficiency scores at the farm level independently from two models: (i) Plot-level Stochastic Frontier, and (ii) the Random Intercept Model (henceforth, the Linear Mixed Effects (LME) Model) ¹. Having arrived at two efficiency scores for the same farm household from the two models, we measure Spearman’s rank correlation coefficient and Kolmogorov-Smirnov D statistic for comparing their performance. Estimation of the correlation coefficient is for checking if the ranking of farm households, based on their efficiency, is in accordance with the true ranking. The Kolmogorov-Smirnov D statistic, which measures the maximum difference between two distributions, has been estimated to check if the full distribution of the true efficiency can be arrived at.

2.3.1 Plot-level Stochastic Frontier Model

We use the stochastic frontier model proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977) for plot-level stochastic frontier estimation, as given by equation (2.5).

$$Y_j = \beta_o + \beta_k X_{jk} + v_j - u_j \quad (2.5)$$

$$j = 1, 2, \dots, n$$

$$v_j \sim N(0, \sigma_v^2)$$

¹The nomenclature and interpretation of “Fixed effects” and “Random effects” is not uniform and is often a source of confusion among researchers. An account of their various interpretations is given in the footnote of Gelman and Hill (2007), p. 245. They identify five different definitions of these terms, out of which, we adopt the first one. Accordingly, “fixed effects” are fixed across all individuals whereas random effects are individual-specific. In the context of our model, the “fixed effects” are parameters (β)s, including the coefficients of the X -covariates as well as the overall population intercept (which can be interpreted as the expected value of the random intercepts in multilevel modelling), estimated from regression, and “random effects” refer to the random deviation estimated to capture subject heterogeneity. These fixed and random effects, together, form the Linear Mixed Effects model.

$$u_j \sim N^+(0, \sigma_u^2)$$

j is an index for a plot. Y_j is the output of plot j . X_{jk} is the k^{th} input applied on plot j . β_k is the estimated coefficient corresponding to input k . u_j is the one-sided inefficiency at the level of plot j and v_j is the symmetric statistical noise term, which is meant to capture measurement error at the plot level. v_j is assumed to follow the normal distribution with zero mean and variance, σ_v^2 . u_j is assumed to follow a half-normal distribution with zero mean and variance, σ_u^2 .

From the plot-level Stochastic Frontier model, we estimate the technical efficiency of each plot based on Jondrow et al. (1982). For deriving efficiency estimates at the farm level from the plot efficiency scores, we use four composite indices: Arithmetic Mean (AM), Output-Weighted Arithmetic Mean (WAM), Geometric Mean (GM) and Output-Weighted Geometric Mean (WGM). The weight refers to the share of the plot's output in the total output of the respective farm household and is applied to the plot's efficiency score.

We estimate a Cobb-Douglas production frontier. Our main interest is in studying aggregation of technical efficiency, and less in the fit of the functional form. Previously, Ruggiero (1999), Ondrich and Ruggiero (2001), and Banker et al. (1993) have used the Cobb-Douglas specification for comparison of different methods of efficiency estimation. The dependent variable is output of maize of each plot.

2.3.2 The Linear Mixed Effects Model

The LME uses the hierarchical structure of the data concerning maize production on multiple plots owned by farm households. Hierarchical data tend to exhibit (positive) correlation within repeated measurements of a cluster. If the presence of this correlation is not accounted for, it leads to erroneous statistical inference as the resultant standard errors are too high. The statistics for hypothesis testing such as the p-value will be flawed (Fitzmaurice et al., 2011). The LME model remedies the problem with the help of a random effects induced covariance structure. It also facilitates the inclusion of covariates which vary at the household (cluster) level and not the plot level – the Z-variables specified in the plot-level stochastic frontier model can be classified as such group-level predictors.

The LME model has been estimated with the following specification:

$$y_{ij} = \beta_o + \beta_k X_{ijk} + \gamma_{oi} + \epsilon_{ij} \quad (2.6)$$

$$i = 1, 2, \dots, m$$

$$j = 1, 2, \dots, n_i$$

$$y_{ij} = \alpha_i + \beta_k X_{ijk} + \epsilon_{ij} \quad (2.7)$$

i refers to a farm household out of a total of m observed farm households. j refers to a plot which belongs to the farm household i . The total number of plots owned by a farm household, n_i , is not the same for all households. Hence, the subscript i has been assigned to n , for denoting the total number of plots owned

by a specific household i . y_{ij} is the maize output of plot j which belongs to farm household i , x_{ijk} is the k^{th} input applied on plot j of farm i and β is a vector of estimated coefficients corresponding to the inputs. γ_{oi} measures the extent to which household i differs in its "base" level of production from the population fixed intercept β_o . Thus, the sum of β_o and γ_{oi} is the household random intercept, denoted by α_i . ϵ_{ij} is the plot level random error.

Having estimated the random intercept for each farm household, we estimate the farm efficiency score, as proposed by Schmidt and Sickles (1984). Aggregation of efficiency indices is carried out as given in Equations (2.8) and (2.9).

$$u_i = \max(\alpha_i) - \alpha_i \quad (2.8)$$

$$TE_i = \exp(-u_i) \quad (2.9)$$

The random intercept estimated at the household level is transformed to arrive at household-level technical efficiency, denoted by TE_i .

2.4 Variable Description

The selection of inputs as frontier covariates follows Liu and Myers (2009), who, in a bid to introduce a model choice procedure across different specifications of the stochastic frontier model, also estimate the model for maize production from a survey of smallholders in Kenya. Similar to Liu and Myers (2009), we also distinguish between inputs which would determine the physical output of maize and those which are expected to affect production by operating as farm management characteristics i.e. Z variables. There is, however, some dissimilarity in the measurement of some variables as compared to our study.

The inputs which are included in the estimation of the production frontier are plot size, seed usage, labour (pre-harvest as well as post-harvest, family as well as hired) and the quantity of fertilizer, pesticide, and manure. An interaction term of seed usage and fertilizer application has been included for estimating the differential impact of fertilizer, given a unitary increase in seed usage. Additionally, we incorporate a dummy variable for the soil quality of each plot, viz., poor, medium and good. Medium soil type is the reference category and the effect of poor and good soil is captured through dummy variables. Similarly, the season of cultivation is controlled for by introducing a dummy variable for long rains (March-April, 2012). The season of short rains (October-November, 2011) is the reference category.

The six Agro-Ecological Zones (AEZs) to which the plots belong have been split into five dummy variables, with Coastal Lowland being the reference category. These AEZ dummies would account for the difference in environmental conditions, the omission of which, would result in an omitted variable bias as they determine input level decisions (Liu and Myers, 2009). Additional dummy variables for certain inputs have been included in order to accommodate for zero input values for some plots under a Cobb-Douglas specification (Battese, 1997).

The set of Z-variables comprises of the farm-level inputs which are expected to affect efficiency. The maximum level of education among the members of a household would affect the efficiency of the household in production. Similarly, the distance to the nearest agricultural extension service center from the household residence is expected to inversely affect production efficiency. The type of land ownership affects the incentive structure for investment through the notion of tenure security (Liu and Myers, 2009). Therefore, the proportion of land

owned out of the total land cultivated by the household has been included. The measurement of this variable differs from Liu and Myers (2009) as they create a dummy variable, depending upon whether the concerned field was owned or rented. We also include a dummy variable which is indicative of whether the farm household tried to avail credit and was unsuccessful in doing so, as this is expected to reduce efficiency by distorting the timing of input usage.

2.5 Data

The survey was concentrated in the areas which mainly grow maize, spread across the six AEZs of Kenya. The classification of AEZs is based on the one given by Hassan et al. (1998). These AEZs were the strata from which rural sublocations were sampled using the probability proportionate to size method. Households were randomly sampled from these sublocations. The reference year for recall was 2012. The data used is a subsample² comprising of 2799 plots, owned by 1050 households. The count of plots from each AEZ is given in Table 2.1. Some plots are repeated in the data in order to account for cultivation in two seasons, long and short rains. The number of observations according to the season is given in Table 2.2. The total number of households exceeds 1050 because there are some plots cultivated in both seasons by them.

²Plots which reported crop failure, as indicated by zero harvest of maize and zero harvest labour were excluded.

Table 2.1: Count of plots by AEZ

AEZ	No. of plots
Highland tropics	234
Moist transitional	578
Dry transitional	638
Dry mid-altitude	532
Moist mid-altitude	619
Lowland tropics	198
Total	2799

Table 2.2: Count of observations by season

Season	No. of plots	No. of households
Long rains (March-April)	1576	921
Short rains (Oct-Nov)	1223	734
Total	2799	1655

The count of households according to plot ownership is given in Table 2.3.

Table 2.3: Count of households by plot ownership

No. of plots	No. of households	
	Long rains	Short rains
1	481	404
2	301	213
3	87	84
4	37	26
5	10	6
6 or more	5	1
Total	921	734

The descriptive statistics are given in Table 2.4.

Table 2.4: Descriptive Statistics

Variable	Unit	Mean	SD	Min	Max
Dry Harvest	Kg	363.6	534.05	1	5490
Plot size	Acre	1.05	0.79	0.05	4
Seed	Kg	6.98	6.3	0.5	60
Fertilizer	Kg	26.15	55.94	0	600
Pesticide	Liters	0.1	0.5	0	6
Manure	Kg	268.6	539.34	0	7000
Labour	Person-days	22.53	20.48	1	210
Poor soil	Dummy	0.12	0.33	0	1
Good soil	Dummy	0.36	0.48	0	1
Max education	Years	10.89	2.87	0	18
Credit Shortage	Dummy	0.12	0.33	0	1
Distance to extension	Km	7.63	8.4	0	80
Female headed HH	Dummy	0.16	0.37	0	1
Own cultivation	Proportion	0.85	0.26	0	1

^a SD stands for the standard deviation.

The seed types were mainly recycled hybrids, local varieties or Open Pollinated Varieties (OPVs). Fertilizer mainly consists of quantities of DAP³ and variants of NPK⁴. Fertility of plot soil was self-reported by the farmers.

These data constitute a hierarchy, wherein a farm-household/producer owns cultivable plots. We use the single level structure of ownership of multiple plots (repeated measurements) by households (the group at a higher level).

³DAP stands for Diammonium Phosphate

⁴NPK stands for Nitrogen, Phosphorous and Potassium

2.6 Monte Carlo simulation

The purpose of carrying out Monte Carlo simulation is to check the performance of the two models in arriving at the true aggregated efficiency scores at the household level as well as observe the effect of changes in plot error on their performance.

We use an artificially created hierarchical set of data from our original data. We establish a balanced cluster of farm households by assigning them a random identification variable which is common across 3 plots per household.⁵ Thus, we have a cluster of 933 households who own 2799 plots, each supposed to be owning 3 plots. We generate the random deviation (γ_{oi}) at the household level from a skew-normal distribution with zero mean, standard deviation 1 and omega parameter as -2.

$$\gamma_{oi} \sim SN(0, 1, -2)$$

$$\alpha_i = 4.35 + \gamma_{oi} \tag{2.10}$$

$$u_i = \max(\alpha_i) - \alpha_i \tag{2.11}$$

Having assumed the fixed population intercept β_o as 4.35 (the average value returned in model estimation), we compute the unique random intercept specific to each household, as given in equation (2.10). We use equation (2.11) to arrive at household inefficiency estimates, and equation (2.12) to generate the true efficiency score for each household.

⁵We extended the analysis procedure to 9 plots per household and found identical patterns in the results.

$$TE_{true} = \exp(-u_i) \quad (2.12)$$

We generate random numbers for the plot error term v_j with different combinations of the parameters pertaining to the assumed normal distribution with parameters μ and σ_v^2 , as the mean and variance, respectively. Thus, σ_v indicates the standard deviation of the plot error.

$$v_j \sim N(\mu, \sigma_v^2)$$

We use two of our X-covariates from the data, plot-size and labour, with their respective elasticities, 0.45 and 0.35, and calculate the true values of y_j through the data generating mechanism, given by equations (2.13) and (2.14).

$$\log y_j = \alpha_i + 0.45 \log \text{plotsize}_j + 0.35 \log \text{labour}_j + \text{ploterror}_j \quad (2.13)$$

$$y_j = \exp(\log y_j) \quad (2.14)$$

We apply the two models, plot-stochastic frontier and LME model, in their original specification as given in Section 2.3.1 and Section 2.3.2, on this newly generated dependent variable, y_j and compute aggregated farm efficiency by their respective aggregation strategies. We use 500 replications of the simulation procedure and compare the farm efficiency scores generated thus, from the two models, with the true values, using Spearman's rank correlation coefficient (ρ) and Kolmogorov-Smirnov test statistic (D). Tables 2.5 and 2.6 present the mean⁶ of ρ and D , for different plot error parameter combinations, across the 500 replications. They are presented for the values, $\mu=0$ and $\mu=2$ ⁷.

⁶The standard deviation of ρ across 500 simulations was 0.00 and increased to a positive integer in the second decimal place as the plot error standard deviation increased.

⁷Further, Monte Carlo simulation was carried out for other values of the assumed mean of the plot error such as -2, 4, and -4. The pattern, as observed in Tables 2.5 and 2.6, did not change.

Table 2.5: Monte Carlo simulation statistics for $\mu=0$

	σ_v	ρ	D	σ_v	ρ	D	σ_v	ρ	D
AM	0.2	0.98	0.74	0.4	0.94	0.78	0.6	0.88	0.82
WAM	0.2	0.97	0.74	0.4	0.93	0.79	0.6	0.85	0.84
GM	0.2	0.98	0.73	0.4	0.94	0.77	0.6	0.88	0.81
WGM	0.2	0.97	0.74	0.4	0.93	0.79	0.6	0.86	0.84
LME	0.2	0.98	0.06	0.4	0.94	0.08	0.6	0.88	0.10

Table 2.6: Monte Carlo simulation statistics for $\mu=2$

	σ_v	ρ	D	σ_v	ρ	D	σ_v	ρ	D
AM	0.2	0.98	0.74	0.4	0.94	0.78	0.6	0.88	0.82
WAM	0.2	0.97	0.74	0.4	0.93	0.79	0.6	0.85	0.84
GM	0.2	0.98	0.73	0.4	0.94	0.77	0.6	0.88	0.82
WGM	0.2	0.97	0.74	0.4	0.93	0.79	0.6	0.86	0.84
LME	0.2	0.98	0.06	0.4	0.94	0.09	0.6	0.88	0.09

An overall comparison of the ρ and the D statistic between the plot stochastic frontier model and LME reveals that both are able to preserve the ranking of true efficiency scores, with minor differences between them. In each case, LME performs slightly better than the aggregation indices but the latter do produce high correlation as well. However, there is a stark contrast between the two models when one considers the D statistic. The aggregation indices lead to high values of the Kolmogorov-Smirnov D , most of them being close to one. The LME model produces low values of the D , most of them being close to zero. This indicates that LME model is well able to maintain the cumulative density of the true efficiency distribution.

An increase in the plot-level statistical error variability (σ_v) erodes the ranking of the efficiency scores, as ρ falls with an increase in the standard deviation. The plot stochastic frontier model as well as LME report a decrease in the correlation due to increase in σ_v .

As far as choosing between the different aggregation indices is concerned, the Arithmetic Mean produces the highest correlation, as compared to other indices. The Geometric Mean produces high correlation at lower levels of plot error variability but its performance drops to second to the Arithmetic Mean, when there is an increase in plot error standard deviation. However, none of the indices are an appropriate choice according to D . Also, irrespective of whether it is WAM or WGM, the application of weights reduce the correlation, as compared to the unweighted means.

2.7 Empirical application

This section applies the two models on the existing maize data from smallholders in Kenya. Estimates are presented in Tables 2.7 and 2.8.

Table 2.7: Plot stochastic frontier and LME estimates

Log(Harvest)	Plot Stochastic frontier		Linear Mixed Model
	Estimate	Estimate	t-value
Intercept	5.07*** (0.17)	4.35 (0.29)	15.14
Log(Size)	0.52*** (0.03)	0.4 (0.04)	9.23
Log(Seed)	0.13*** (0.04)	0.17 (0.04)	4.06
Log(Labour)	0.05 (0.03)	0.18 (0.04)	4.48
Fertilizer dummy	0.11 (0.07)	0.03 (0.1)	0.33
Log(Fertilizer)	0.07* (0.03)	0.09 (0.04)	2.16
Log(Seed)*Log(Fertilizer)	0.04*** (0.01)	0.02 (0.01)	1.81
Pesticide dummy	-0.08 (0.06)	-0.09 (0.1)	-0.97
Log(Pesticide)	0.02 (0.03)	-0.02 (0.06)	-0.28
Manure dummy	0.49*** (0.11)	0.04 (0.14)	0.31
Log(Manure)	0.07*** (0.02)	0.01 (0.02)	0.37

Table 2.8: Plot stochastic frontier and LME estimates (contd.)

Log(Harvest)	Plot Stochastic frontier		Linear Mixed Model
	Estimate	Estimate	t-value
Poor soil	-0.27*** (0.05)	-0.16 (0.07)	-2.15
Good soil	0.15*** (0.04)	0.08 (0.05)	1.73
Long rains dummy	0.05 (0.03)	-0.03 (0.03)	-1.01
High Tropics	0.68*** (0.09)	0.59 (0.17)	3.59
Moist Transitional	0.35*** (0.08)	0.36 (0.15)	2.33
Dry Transitional	0.2** (0.08)	0.1 (0.16)	0.62
Dry Mid-Altitude	0.2** (0.08)	0.03 (0.16)	0.2
Moist Mid-Altitude	0.37*** (0.08)	0.41 (0.15)	2.71

^a Figures have been rounded upto 2 decimal places.

^b *, **, *** correspond to 0.1, 0.05 and 0.01 level of significance, respectively.

^c Standard errors are reported in parenthesis.

In the estimation of plot stochastic frontier, monotonicity is globally satisfied as all output elasticities are positive. The largest output elasticity is that of the plot size, followed by seed. Soil fertility plays a major role in determining pro-

duction, as the coefficients of both, poor as well as good type are significantly different from medium category, which is the reference category. Also, they have opposite signs, as expected. The five AEZs included in the model perform significantly better than Coastal Lowland, which is the reference category. The LME model hints at a significant effect of labour on the frontier, with an output elasticity of 0.18 percent.

Table 2.9 provides the coefficients of Z-variables/group-level predictors, respectively, from plot stochastic frontier and the LME model. Education in the household and a female headed household are significant in explaining inefficiency, according to plot stochastic frontier. The former reduces inefficiency and the latter increases it, as expected. The negative effect of a female headed household is in lines with the result of Liu and Myers (2009). They explain this adverse effect on efficiency through the fact that it is difficult for women to possess land ownership rights, unlike men and this affects the incentive to work. The coefficients of group-level predictors in LME model indicate the effect of a unitary increase in the predictor on the household random deviation.

Table 2.10 presents model-specific results. In terms of AIC, BIC and Log-likelihood, the LME model fares better than the plot stochastic frontier. Further, the higher value of the standard deviation of the random effect, denoted by σ_{γ} , as compared to the residual standard deviation, confirms a high level of heterogeneity among the farm households in production. The estimates of Gamma and variance share of the inefficiency term in plot stochastic frontier confirm the existence and high level of inefficiency in production by the households.

Table 2.9: Estimates of coefficients of plot-invariant variables

Log(Harvest)	Plot Stochastic frontier	Linear Mixed Model	
	Estimate	Estimate	t-value
Max education	-0.21*** (0.05)	0.01 (0.01)	0.67
Female headed HH	0.81*** (0.23)	-0.15 (0.09)	-1.64
Cultivated land owned	-0.83 (0.49)	-0.36 (0.14)	-2.59
Distance to Extension	-0.01 (0.01)	0 (0)	-1.13
Credit Shortage	-0.53 (0.3)	0.13 (0.1)	1.26

^a Figures have been rounded upto 2 decimal places.

^b *, **, *** correspond to 0.1, 0.05 and 0.01 level of significance, respectively.

^c Standard errors are reported in parenthesis.

2.8 Conclusion

This study is a first in examining the performance of the LME model and aggregation indices in estimating technical efficiency when there is a data hierarchy, using Stochastic Frontier Analysis. We perform Monte Carlo simulations with replications of the data generating process, using different parameter combinations of the plot error and observe the mean of correlation and Kolmogorov-Smirnov statistic of plot-level stochastic frontier and the Linear Mixed Effects model, in order to compare the accuracy in efficiency estimation at an aggregate (farm household) level. We observe that both models maintain the ranking

Table 2.10: Model-specific estimates

	Plot Stochastic frontier	Linear Mixed Model
AIC	8117	7210
BIC	8272	7364
Log likelihood	-4033	-3579
σ_γ		1.03
Residual σ		0.577
Mean efficiency	0.446	
Gamma	0.967	
$\text{Var}(u) / \text{Var}(u)+\text{Var}(v)$	0.914	

^a σ_γ denotes the standard deviation of the household random deviation γ_{oi} .

^b Residual σ is the standard deviation of the residuals after LME estimation.

of households according to the true ranking. However, the LME also closely estimates the true efficiency distribution, unlike aggregation indices of the plot-level stochastic frontier. The variability of plot-level error plays a systematic role in affecting the performance of both models. As it increases, the comparability of both models with the true aggregate efficiency distribution is reduced progressively.

The empirical application of the two models on maize production data, collected from smallholders in Kenya, gives insight into the factors which play a role in determining production of maize and the efficiency of the concerned farm households. There is scope for increasing the production through improving the soil fertility. Higher education achieved in the household improves efficiency in production.

Further, one can explore the potential of the LME model by incorporating more levels in the data hierarchy and use Multilevel Modelling to check the robustness of the results of our study.

CHAPTER 3

THE ROLE OF TEMPORAL LAND USE IN DRIVING SUSTAINABLE INTENSIFICATION IN BANGLADESH: AN APPLICATION OF HIERARCHICAL MODELLING

Abstract

In the context of the possibly adverse effects of climate change on crop production in Bangladesh, this paper examines the effect of temporal land use and fertilizer application on paddy production, with the aim of studying the scope for sustainable intensification. In the backdrop of four agricultural seasons which are characterised by differing climate, the study finds that land use in the preceding seasons determines the present season's output of paddy and that it also determines the final effect of fertilizer on the present paddy crop. Sustainable intensification depends on the crop choice in different seasons, potato cultivation emerging as the preferable one. Fallow land use is preferable in Aus season for higher production in Aman season. Finally, there are significant interaction effects of fertilizer usage with the preceding season's land use on production.

Keywords : Sustainable intensification, Hierarchical modelling, Bangladesh

3.1 Introduction

In the wake of an increased demand for food due to an increase in the population of the world, the intensification of agriculture has been proposed as a solution to meet global food security. "Agricultural intensification can be technically defined as an increase in agricultural production per unit of inputs (which may be labour, land, time, fertilizer, seed, feed or cash)" (FAO, 2004, Agricultural intensification section, para. 1). However, the need for "sustainable" agricultural intensification has been emphasized in the light of the impending phenomenon of climate change. Petersen and Snapp (2015) summarize the different interpretations of the term, based on interviews with experts from different scientific disciplines, related to agriculture and ecology. The concept of sustainable agricultural intensification has been interpreted differently from different viewpoints. In a generic sense, it refers to the simultaneous achievement of food security and ensuring the least damage to the environment. According to the different steps which can be taken to bring about sustainable intensification, as mentioned by the experts, they categorise the approach to sustainable intensification into biological and intensification approaches. One intensification approach is to increase agricultural diversity in the production system, of which one practice is crop rotation.

Rice production is particularly envisaged to be vulnerable to the adverse effects of climate change. Wichelns (2016) summarized the potential effects on rice production systems in Asia, wherein, with a possible increase in temperature and erratic rainfall patterns (characterized by a change in the quantity and timing), climate change will lead to thwarting of plant growth and yield. Gatto et al. (2020) study the role of potato in determining the scope for sustainable intensification in Eastern Indo-Gangetic plains, where rice is the dominant pro-

duction system. Having considered expert opinions and results of simulations, they prescribe that intensification should be made sustainable through managing soil efficiently in the context of fertilizer use.

Rahman and Parkinson (2007) examine the nexus between soil fertility and the decision of paddy farmers regarding the employment of resources such as fertilizer, labour and animal power in selected agro-ecological zones of Bangladesh. While they find that soil fertility leads to an increase in productivity by way of an increase in the yield, they emphasize that a policy for soil nutrient management is desirable at the village level. They find that farmers respond and make amendments in input allocation based on their different soil regimes. This provides motivation for a targeted village-level analysis for identifying high and low potential areas in terms of paddy production as well as studying the differential effects of fertilizer, given the spatially diverse physiography.

Thomas (2003) developed a model for measuring crop rotation and relating it with fertilizer application at the farm level (in the absence of plot data). Having accounted for expected fertilizer prices in the future as well as the propensity of nitrogen carryover in inter-temporal applications, this model indicates that in the decision regarding substitution of fertilizer remnant for new application, the farmer considers the difference in their respective marginal products.

The potential role of fertilizer in ensuring sustainable intensification in a dominant rice production system specific to Bangladesh is of interest, given its vulnerability to climate change. This study examines the popular choices of paddy farmers regarding land use, as an intensification approach, and pits them against intermediate fallow land use, with respect to increasing production.

Further, it attempts to conclude whether sustainable intensification is driven by the crop choices which define intermediate land use, or given the choices, works through the controlled use of fertilizer.

The rest of the paper is organized as follows: Section 3.1.1 presents the modelling context. Section 3.2 sets the context for the analysis by describing its physiography as well as an exploration into the data in subsections. An overview of the cropping system has been provided for introducing the background for analysis in the latter subsection. Section 3.3 presents the methodology. Section 3.4 describes the data used in the study. The results of the study are presented and discussed in Section 3.5 and the final conclusion is presented in Section 3.6.

3.1.1 Modelling context

Just and Pope (2001) explain the factors which set the modelling of production technology in agriculture apart from that of conventional manufacturing. Production is distributed across time, with the weather driving the different growing seasons. The gestation period is long and is not easily observed. An important characteristic of input usage is that its timing is crucial due to its dependence on the weather. Thus, the timing of input usage such as that of fertilizer is also a decision variable. The spatial dimension of crop production on different plots implies competition and complementarity in crop choice, thereby determining the degree of diversification and/or specialization.

According to Just and Pope (2001), one feature of agricultural production is that the current state of technology depends on previous decisions regarding inputs and resource management practices. Crop rotation is one response for converting the technology into a favourable one for the crop in question. Therefore, they acknowledge the assessment of intrayear dynamics for a complete

understanding of production possibilities over time.

3.2 The context of crop cultivation in Bangladesh

The inter-seasonal dynamics in agriculture within a year of Bangladesh have been delineated as a prelude to the analysis of this study. Section 3.2 is a positive study which consists of a description of the physiography of Bangladesh and the cropping plan adopted by farmers across the seasons.

3.2.1 The physiography of Bangladesh

Bangladesh is a diverse country from the viewpoint of its physical geography. Brammer (2012) provides an overview, encompassing its physical environment as well as a description of its physiographic regions. Stark contrasts in land use and landscapes occur even in very short distances. It is, therefore, imperative to study its climate, landforms, hydrological endowments, as well as the distribution of its soil types, for the purpose of realizing its widespread heterogeneity.

Bangladesh is situated between 20°25' and 38°40' N latitudes and 88°01' and 92°40' E longitudes. It lies in the Ganges-Brahmaputra-Meghna (GBM) catchment area and is a conglomeration of the Himalayas, other hills, and plateaus. 79% of the total land area is covered by floodplains. The climate of Bangladesh is volatile in terms of various seasons within a year, and is also dynamic across years. Extreme weather events such as floods and droughts are frequent. The country also bears the brunt of periodic El Nino and La Nina phenomena. The amount of rainfall is also not certain (Brammer, 2012).

There are four main seasons in Bangladesh from the climatic perspective: Winter (December-February), Summer (March-May), Southwest monsoon (June-September) and Autumn (October-November) (Khatun et al., 2016). These seasons roughly correspond to four agricultural seasons (in terms of the sowing months), namely, Boro, Aus, Aman and Rabi, respectively. The main seasons of paddy production are Aman, which is completely rainfed, Boro, which is completely irrigated, and Aus.

3.2.2 Exploratory analysis of the data

Figure 3.1 provides the distribution of plot size across crops. The distribution is skewed with a large number of plots reporting very low area in terms of the total acres cultivated. Figure 3.2 provides the density of the proportion of irrigated plot area. The value 0 indicates a completely rainfed plot and 1 indicates a completely irrigated plot. Irrigation, however, is of relevance in Boro season, and to some extent, in Aus season (during the beginning of its growing months (Rashid et al., 2012)).

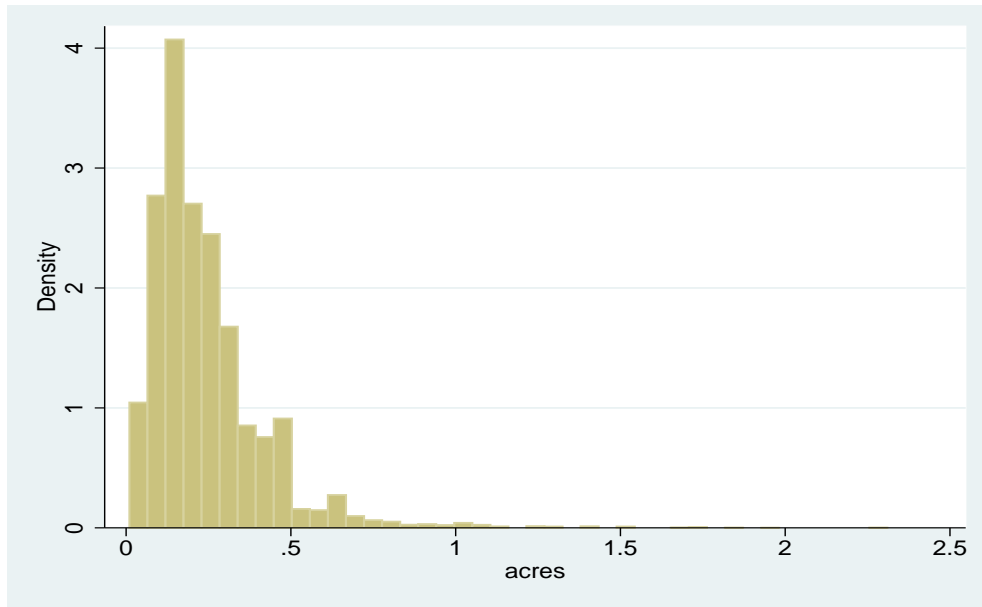


Figure 3.1: The distribution of plot size across all crops

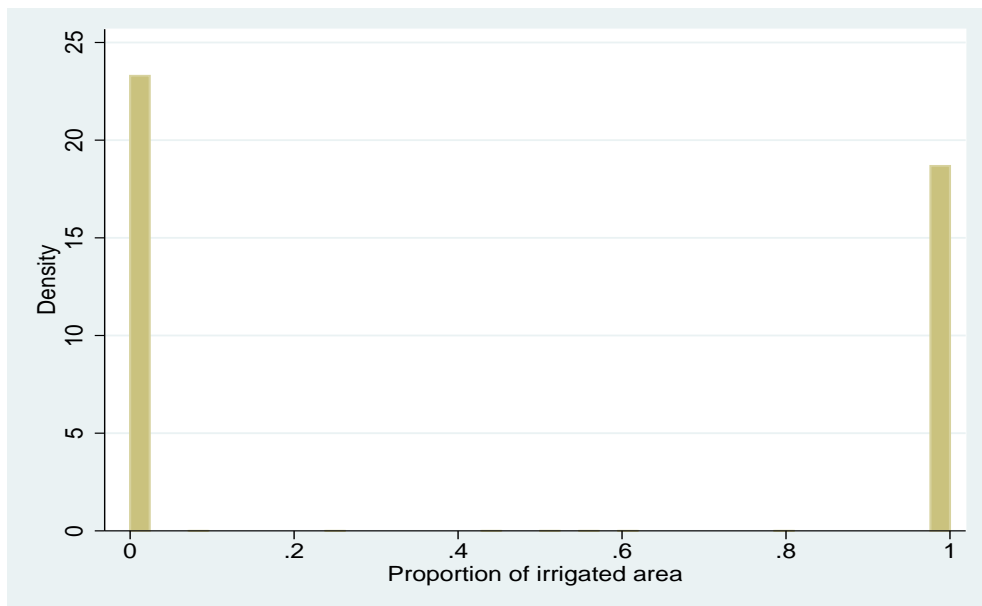


Figure 3.2: The distribution of the proportion of irrigated plot area

A large proportion of the irrigated area is served by Shallow TubeWells (STWs), Deep TubeWells (DTWs) and Low-Lift Pumps (LLPs). The dominant soil types are clay loamy, loamy, and sandy loamy. The dominant plot slopes

are high land, medium land and low land. The highest incidence of cropping intensity at the plot level is 2 crops per year, followed by 1 and 3, in that order.

Castellazzi et al. (2008) classify crop rotation into four categories, out of which, the cropping plan adopted by paddy farmers of Bangladesh fits into category b. This category entails the rotation type, in which the farmer, having completed the rotation cycle, will determine the land use with a possibility of deviation from the past crop choice after returning to a particular point of time. However, the rotation length is fixed, implying that the crop choices which are available to the farmer at a time period are fixed and are not interchangeable among the different time periods. Figure 3.3 is a representation of the rotation followed by the farmers across the four seasons. The first vertical set of blocks is related to the popular choices of crops in Rabi season. It consists of fallow, wheat, potato, mustard, and leguminous crops. The second set and the fourth set refer to the land use choice between keeping it fallow and paddy cultivation, in Boro and Aman seasons, respectively. Boro and Aman are the main seasons of paddy cultivation in Bangladesh. Boro paddy is an entirely irrigated crop, whereas Aman paddy is rainfed. The third vertical set of blocks in Figure 3.3 is related to Aus season, known for its distinct choice of jute cultivation. Although some paddy is cultivated as well, the main choice of land use in Aus is fallow.

Among the different crop choices made by farmers, there is a consistent pattern which was observed in the transition from one season to another. Mustard and potato cultivation in Rabi season will result in paddy cultivation or fallow land use in Boro season. However, if the farmers grow leguminous crops in Rabi season such as grasspea, lentil, pigeonpea, greengram etc., the plot will be kept fallow in the succeeding Boro season. Wheat cultivation in Rabi season is also followed by fallow land use in Boro season. Similarly, paddy cultivation in Aus

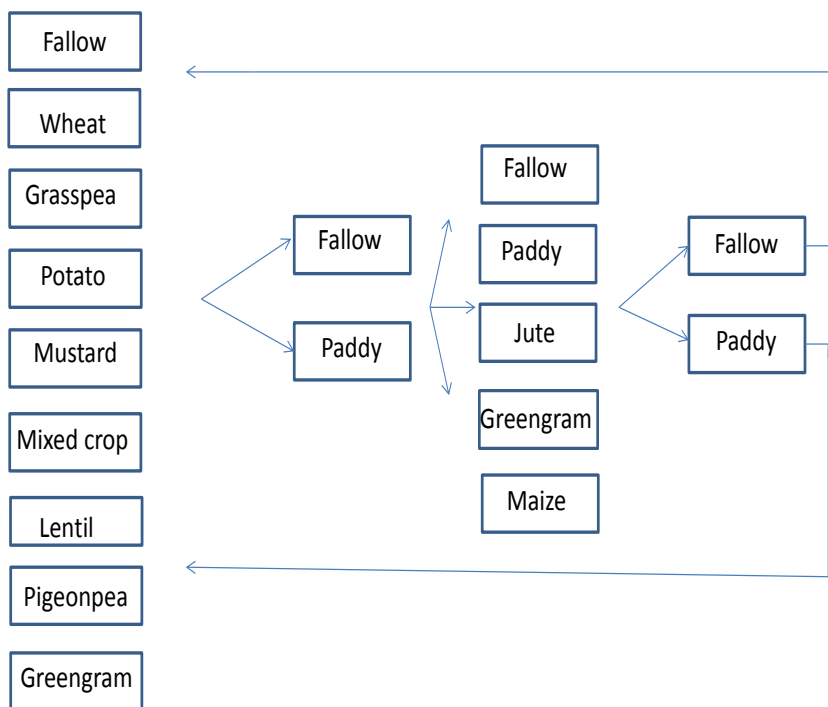


Figure 3.3: Flexible rotation, cyclical, fixed rotation length

season is followed by either paddy cultivation or fallow land use in the succeeding Aman season. If the farmer cultivates jute in Aus, the popular choice of land use in Aman is fallow. Each of the three seasons of paddy cultivation are characterised by distinct varieties of paddy in terms of adoption.

3.3 Methodology

We formulate two models for incorporating land use in the previous season and examining the role of fertilizer in ensuring sustainable intensification. Both models originate from the methods devised for multilevel modelling. The two

models are termed as: (i) Random Intercept (RI) model with interaction terms, and (ii) Random Coefficients Model (RCM) with cross-level interaction. In both models, the interaction terms are formed by season-specific dummy variables for different land uses (land use in seasons preceding the paddy growing seasons of Boro and Aman). In the RI model, the season-specific dummy variables for different land uses are interacted with the dummy variables of paddy growing seasons of Boro and Aman, and village-level random effects are accounted for by specifying the random intercept at the village level¹. In the RCM model, alongwith specifying the random intercept at the village level, fertilizer is specified as varying in slope by village, i.e. random slopes are specified for accounting for the fertilizer effect. In addition, season-specific dummy variables for different land uses in seasons preceding the paddy growing seasons of Boro and Aman are interacted with fertilizer, i.e. we specify a random slope with cross-level interaction effect. Both models implement the Cobb-Douglas specification.

3.3.1 Random Intercept model with interaction effects

Equation (3.1) presents the the random intercept model.

$$y_{vjs} = \beta_o + \beta_k X_{vjsk} + \gamma_{ov} + \epsilon_{vjs} \quad (3.1)$$

¹A peculiar feature of this model is that inspite of the inclusion of interactions between season-specific land use dummy variables and the paddy-growing seasons of Boro and Aman, the land use dummy variables have not been included as main effects in the regression. Inclusion of all the main effects, as separate from interaction terms, has been prescribed in statistical literature and there is considerable debate about the implications of their exclusion on inference. At the outset, the introduction of all the main effects alongwith interactions led to a high degree of multicollinearity and automatic exclusion of the interaction terms. Hence, a model with the main effects as dummy variables and without any interaction terms was estimated for verification. The results indicated that the only significant land use effect was that of potato cultivation in Rabi season and it was positive (0.19).

$$v = 1, 2, \dots, 11$$

$$j = 1, 2, \dots, n_v$$

$$s = 1, 2, 3, 4$$

$$y_{vjs} = \alpha_v + \beta_k X_{vjsk} + \epsilon_{vjs} \quad (3.2)$$

v refers to a village in the sample. j refers to a plot which is cultivated in season s . The total number of plots in a village, n_v , is not the same for all villages. Hence, the subscript v has been assigned to n , for denoting the total number of plots which constitute a village v . s is the notation for the four seasons, Rabi, Boro, Aus and Aman. y_{vjs} is the paddy production of plot j in village v , cultivated in season s . x_{vjsk} is the k^{th} input applied on plot j of village v in season s , and β is a vector of estimated coefficients corresponding to the inputs. γ_{ov} measures the extent to which village v differs in its "base" level of production from the population fixed intercept β_o . Thus, the sum of β_o and γ_{ov} is the village random intercept, denoted by α_v , given in equation (3.2). ϵ_{vjs} is the plot level random error in season s .

3.3.2 Random Coefficients Model with cross-level interaction

Kim et al. (2001) use a RCM for specifying variability in the fertilizer effect by plot. The random slope (or random coefficient) has been incorporated for capturing heterogeneity in production, arising from plot-specific attributes such as soil quality. We adopt a similar approach by specifying random slopes for fertilizer, varying by village, alongwith a village random intercept. Further, we introduce cross-level interaction terms which interact fertilizer application with land use dummy variables. Equation (3.3) presents the full form of the RCM model with cross-level interaction.

$$y_{vjs} = \beta_o + \gamma_{ov} + \beta_1 Fertilizer_{vjs} + \gamma_{1v} Fertilizer_{vjs} + \beta_k X_{vjsk} + \epsilon_{vjs} + Fertilizer_{vjs} : Landusedummy_{vjs} \quad (3.3)$$

$\beta_o + \gamma_{ov}$ represents the random intercept of village v , analogous to the random intercept model of equation (3.1) and (3.2). β_1 of $Fertilizer_{vjs}$ is the fixed coefficient of fertilizer in the regression. γ_{1v} is the village-specific deviation of the slope of fertilizer on production. Thus, $\beta_1 Fertilizer_{vjs} + \gamma_{1v} Fertilizer_{vjs}$ is the total village-specific effect of fertilizer on production. x_{vjsk} is the k^{th} input applied on plot j of village v in season s , and β is a vector of estimated coefficients corresponding to the inputs, analogous to equation (3.1). $Fertilizer_{vjs} : Landusedummy_{vjs}$ is the cross-level interaction of land use dummy variables with fertilizer slopes. The coefficient of this cross-level interaction measures the extent to which the effect of fertilizer on production is moderated or facilitated by land use in the previous season. Specific land use choices such as fallow or crops such as mustard, potato, and paddy have been modeled as dummy variables and interacted with fertilizer.

3.4 Data

We use panel data of International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), collected for the project, Village Dynamics of South Asia (VDSA). The data is specific to Bangladesh, and is related to the time period, 2009-2014. We analyse data on paddy production from eleven villages, which can be grouped into 8 AEZs, based on the classification by Quddus (2009). Table 3.1 presents the number of observations at different levels of analysis.

Table 3.1: Count of observations by group in the sample

AEZ	Village	No. of hhs.	No. of plots	No. of obs.
MMRF	Begumpur	31	232	615
MMRF	Bhabanipur	35	195	853
HPTF	Boikunthapur	40	387	811
GTF	Dakkin Kabir Kathi	41	334	875
KFAB	Dharikamari	31	402	1324
BJF	Konapara	39	187	656
HGRF	Niamatpur	38	190	591
BJF	Nishaiganj	15	43	100
LGRF	Paschim Bahadurpur	31	98	116
DHAKA	Patordia	30	133	211
HPTF	Rasun Shimulbari	35	253	1144
Total		366	2454	7296

The plots are repeated across the four seasons and six years under study, albeit not regularly.

The full form of the abbreviation of AEZ names is provided in Table A.1 of the Appendix.

Table 3.2 presents the summary statistics of the continuous variables in the study. Production refers to the output of rice at the level of a plot and has been converted from local area units into kilograms. The quantity of output was measured in *maund*² (a local unit of measuring the output) and quintals³. The area of a plot has been converted from *decimal* (a local unit of measuring the

²The applied conversion equation is 1 maund = 37.3242 kilograms.

³The applied conversion equation is 1 quintal = 100 kilograms.

area) to acres.⁴ Since the mean size of plot is 0.22 acres, and its maximum size is 0.72 acres, one can conclude that farmers in the sample are smallholders with less variation in plot size. Also, the distribution is highly skewed with a positive skewness of 1.08. The seeds were of the local, improved/HYV, and hybrid type. Urea, Urea Super Granule (USG), Diammonium Phosphate (DAP), Triple Superphosphate (TSP), Mureate of Potash (MP), Gypsum, Zinc and "mix fertilizer" are the main fertilizers used for cultivation. Since some observations report zero application of fertilizer, a dummy variable for the zero values was included (Battese, 1997). The labour hours in the sample refer to the total number of hours across family, hired as well as exchange labour in terms of male as well as female. They also account for bullocks and farm servants.

Precipitation refers to the total rainfall in the sowing months of paddy, of the respective season in a year, across the region. Accordingly, rainfall in Aman season is the total precipitation across the months of July and August, recorded by region, for the different years in the sample. Similarly, for Aus season, the sowing months were March and April. The sowing months for Boro paddy are December of each year, followed by the month of January of the succeeding year. Accordingly, rainfall was summed across December of each year with that in the month of January of the succeeding year. A dummy variable for zero values of precipitation was included (Battese, 1997). The depth of the soil in a plot is in centimeters (cms.) and covers most of the sample in the range of 12 to 15 cms. The value for some plots which reported missing values of soil depth was imputed using the modal value of soil depth in the village to which the plot belonged. Soil fertility has been excluded due to very little variation in the variable – most of the plots were reported to have good soil.

⁴The applied conversion equation is 1 decimal = 0.01 acre.

Table 3.2: Summary statistics of continuous variables

Variable	Unit	Mean	SD	Min	Max
Production	Kg.	363.01	292.84	1.75	2080
Size	Acres	0.22	0.13	0.02	0.72
Seed	Kg.	4.05	3.02	0.12	25
Fertilizer	Kg.	28.23	25.67	0	245
Labour	Hours	77.17	51.55	4.25	531
Precipitation	mm	301.5	342.53	0	1074
Soil depth	cms.	13.33	1.19	8	16

Table 3.3 presents the percentage distribution of the dummy variables which take on the value one. The depth of flooding in the plot during the peak of monsoon is divided into four categories: Little, knee-deep, chest-deep and more than chest-deep water in the peak of monsoon. This variable qualifies as an important determinant of production. Given that northeast monsoon accounts for only 2% of the total annual rainfall (Khatun et al., 2016), flooding in the peak of monsoon mainly refers to flooding in Aman season (i.e. Southwest monsoon). The dummy variables for fallow land use in the previous season, potato and mustard cultivation in Rabi season, and paddy cultivation in Aus season are the ones which have to be accounted for in the production technology of rice cultivation in Aman and Boro season, thereby including temporal dynamics. As Table 3.3 indicates, a popular choice at the seasonal level is fallow land use, accounting for 66.69% of the sample. Paddy cultivation in Aus season is a preferred choice for some farmers in the sample. Also, the main types of soil are clay loamy, loamy and sandy loamy, with about half of the sample reporting sandy loamy soil. A dummy variable for the year 2009 was added for capturing

the effect of El Nino, and for 2010 and 2011, the effect of La Nina.

Table 3.3: Percentage distribution of categorical variables

Variable	Percentage of total sample
Fertilizer dummy for zero value	0
Precipitation dummy for zero value	21.25
Depth of flooding during peak of monsoon:	
Little water	28.13
Knee-deep water	55.15
Chest-deep water	14.04
More than chest-deep water	2.65
Land use:	
Fallow in previous season	66.68
Potato in Rabi	1.87
Mustard in Rabi	2.34
Paddy in Aus	12.7
Soil type:	
Clay loamy	16.5
Loamy	28.86
Sandy loamy	52.38

"Fallow in previous season" accounts for plot kept fallow during the preceding season of any observation.

3.5 Results

We compare the results from the two estimation approaches, as described in the methodology.

We estimate a random intercept model by specifying the village as bringing in the random effects⁵. Table 3.4 presents the fixed effect estimates of the model. The plot size has been estimated to have a very high output elasticity of about 0.70. The magnitude of this elasticity is justified, given the low average size of plot. The elasticity of fertilizer and precipitation is positive and significant⁶. Labour elasticity is also positive, high, and significant. The dummy variables of flooding levels, as indicated by little water, knee-deep water, chest-deep water show negative and significant effect on output as compared to "more than chest-deep water", which is the reference category. The interpretation of the effect of an increase in the depth of the soil will result in that of a log-linear model. An increase by 1 centimeter increases the logged value of output by 0.01 units. The annual rate of technical change is high, estimated at approximately 4%. The occurrence of La Nina in 2011 had significantly reduced the output. The estimates of dummy variables (Aus being the reference category) for the three rice-producing seasons indicate a high and significant effect of Boro season on production.

We estimate the effect of land use in the previous season, in terms of whether it was kept fallow or used for the cultivation of a popular and seasonal crop,

⁵We began with estimating a nested random intercept model, with AEZ, Village, Household and Plot, in that order. The structure reported very low random effects variance at each level. Also, AEZ alone as a random intercept reported very low random effects variance, which is possible due to high heterogeneity within an AEZ.

⁶An interesting relationship could have been studied through an interaction between the dummy variable for zero precipitation and dummy variables for season (especially, with Aman season). Its estimated coefficient would have indicated the effect of less or no rainfall on production in the rainfed Aman season. However, the derived interaction variable had little variation as there are very few observations reporting zero rainfall in Aman season.

on production. For this purpose, we include season-specific interaction with dummy variables for land use. Accordingly, dummy variables for Aman and Boro were interacted with a dummy variable indicating that the plot was kept fallow in the previous season, namely, Aus and Rabi, respectively. As Table 3.4 indicates, keeping the plot fallow in Rabi season adversely affects production in Boro season, as indicated by a negative and significant coefficient. On the contrary, keeping the plot fallow in Aus season is conducive to rice production in Aman season, as indicated by a positive and significant coefficient. The estimate of the interaction between Boro season and potato cultivation in Rabi season is high in magnitude as well as significance.

Table 3.5 presents the variance parameters of the random intercept and residual. The variance of the village intercept is lower than the residual variance.

Table 3.4: Estimates of the Random Intercept model

Dependent variable: Log(Output)	Estimate	Std. Error	t value
Intercept	5.00	0.13	38.49
Log(Acres)	0.70	0.02	34.07
Log(Seed)	-0.01	0.01	-0.68
Log(Fertilizer)	0.08	0.01	8.88
Log(Precipitation)	0.03	0.01	5.67
Log(Labour)	0.25	0.02	13.59
Dummy for zero precipitation	0.06	0.02	4.01
Dummy for zero fertilizer	-0.51	0.16	-3.13
Little water	-0.05	0.02	-2.14
Knee-deep water	-0.07	0.02	-2.89
Chest-deep water	-0.05	0.02	-2.31
Soil depth	0.01	0.00	2.74
Mean-scaled year	0.04	0.01	6.69
El Nino_2009	0.00	0.03	0.16
La Nina_2010	-0.02	0.02	-0.91
La Nina_2011	-0.07	0.02	-4.52
Aman	0.01	0.02	0.25
Boro	0.69	0.03	25.80
Boro_fallowinPrevious	-0.05	0.02	-2.56
Aman_fallowinPrevious	0.04	0.01	2.89
Boro_potatoinRabi	0.13	0.03	4.00
Boro_mustardinRabi	-0.06	0.03	-1.94
Aman_paddyinAus	0.00	0.02	0.13

Figures have been rounded to two decimal places.

Table 3.5: Variance estimates of the random intercept model

Group	Variance	std.dev
Village Intercept	0.02528	0.1590
Residual	0.0761	0.2759

Table 3.6 presents the estimates of fixed effects of the random coefficients model. The estimates are close to those in Table 3.4, with the exception of the dummy variables included for flooding. Once fertilizer is specified as a random slope, flooding ceases to be significant. We introduce cross-level interaction between land use in the previous season and fertilizer application in order to measure the effect of land use as a predictor of the slope of fertilizer, which is varying by village. The two main takeaways regarding the cross-level interaction terms are that potato cultivation in Rabi season increases the marginal effect of fertilizer in Boro season. This finding is in tandem with that of what has been explained in Gatto et al. (2020). They explain that although potato cultivation per se requires higher application of fertilizer, it has a strong carry-over effect in rice systems as the next crop requires lesser fertilizer application (Ali et al. (2008) as cited in Gatto et al. (2020)). Paddy cultivation in Aus season reduces the marginal effect of fertilizer in Aman season. Table 3.7 presents the variance estimates of the RCM. The correlation of -0.68 implies that as one increases the village RI, the slope of the fertilizer reduces, implying that high potential villages exhibit lower effect of incremental fertilizer.

Table 3.6: Estimates of the Random Coefficients model

Dependent variable: Log(Output)	Estimate	Std. Error	t-value
Intercept	4.99	0.14	35.79
Log(Acres)	0.71	0.02	34.56
Log(Seed)	-0.01	0.01	-1.37
Log(Fertilizer)	0.08	0.02	4.72
Log(Precipitation)	0.03	0.01	5.12
Log(Labour)	0.26	0.02	14.08
Dummy for zero precipitation	0.06	0.02	3.54
Dummy for zero fertilizer	-0.50	0.16	-3.07
Little water	-0.02	0.02	-0.82
Knee-deep water	-0.04	0.02	-1.78
Chest-deep water	-0.03	0.02	-1.22
Soil depth	0.01	0.00	2.57
Mean-scaled year	0.05	0.01	7.29
El Nino_2009	0.00	0.03	0.13
La Nina_2010	-0.01	0.02	-0.48
La Nina_2011	-0.07	0.02	-4.54
Aman	0.03	0.02	1.74
Boro	0.64	0.02	31.33
Log(Fertilizer) : Fallow in Previous	0.00	0.00	-0.61
Log(Fertilizer) : Potato in Rabi	0.05	0.01	4.14
Log(Fertilizer) : Mustard in Rabi	-0.01	0.01	-1.11
Log(Fertilizer) : Paddy in Aus	-0.01	0.00	-2.79

Figures have been rounded to two decimal places.

Table 3.7: Variance estimates of the random coefficient model

	Variance	std.dev
Village Intercept	0.0483	0.2198
Log(Fertilizer)	0.0022	0.0469
Residual	0.0754	0.2747
Correlation	-0.68	

Figure 3.4 presents a caterpillar plot of the RCM, with village RI as well as fertilizer random slopes by village. The first section indicates that the first five villages are the high potential areas for paddy production with high "base" production levels, as their estimated random deviation is greater than zero. The rest of the villages are lower than average in terms of production potential. The second section gives the estimates of the random slopes of fertilizer by village. Although most of the estimates are close to zero, most of the high production potential villages report a negative specific deviation of their random slope. This implies a lower slope for the total village-specific slope of fertilizer for those villages, hinting at a lower than average effect.

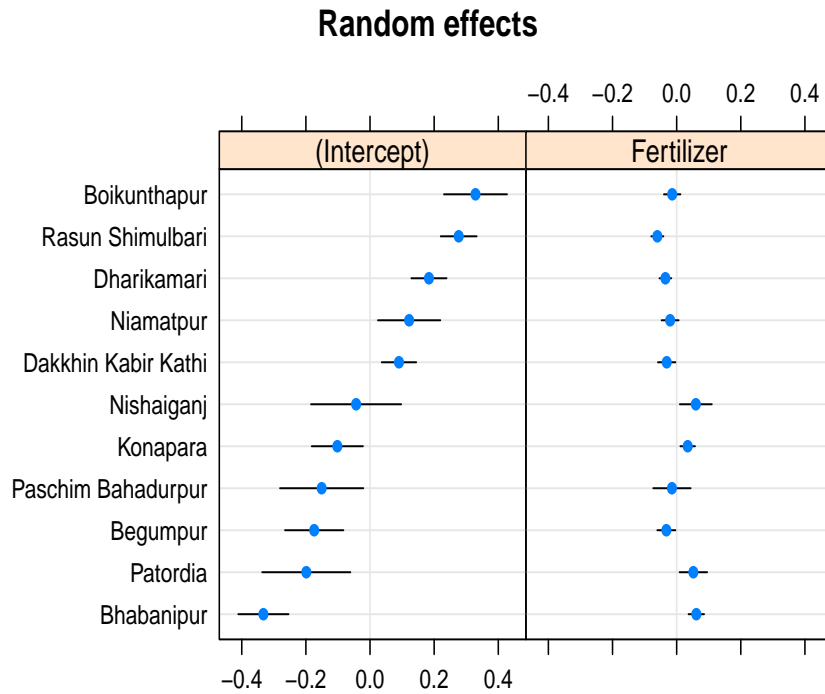


Figure 3.4: Caterpillar plot of Random effects of the RCM

3.6 Conclusion

In the backdrop of the projected vulnerability of Bangladesh to climate change, this paper has dual objectives: (i) to study the effect of temporal land use in terms of different crop choices as against the decision to keep the plot fallow on production, and (ii) to study the interactive role of fertilizer with temporal land use in bringing about sustainability in intensification. For these objectives, hierarchical modelling has been employed through the RI and RCM models. In both cases, village random effects have been specified for capturing heterogeneity which may arise due to diverse physiography across villages.

From the RI model, we can conclude that for increasing output of paddy in Aman season, it is preferable to keep the plot fallow in the preceding Aus

season. However, for higher output in Boro season, it is preferable to cultivate potato in the preceding Rabi season instead of keeping it fallow. From the RCM, we can conclude that potato cultivation in Rabi season increases the effect of fertilizer in Boro season, thereby leading to a lower requirement of the physical quantity of fertilizer in Boro paddy. Paddy cultivation in Aus season reduces the effect of fertilizer in Aman season, thereby leading to a higher requirement of fertilizer in terms of physical quantity in Aman. This implies that potato cultivation in Rabi ensures sustainable intensification and paddy cultivation in Aus reduces sustainability in intensification.

This paper has studied the potential of land use choices for sustainable intensification in Bangladesh. A comprehensive analysis of the intra-year and inter-seasonal dynamics of crop production in Bangladesh is required to prescribe an optimal and unit-specific cropping plan as an adaptation strategy to climate change. One can also extend the scope of the study to mitigation efforts for combating the adverse effects of events caused by climate change on production.

CHAPTER 4
STOCHASTIC FRONTIER ANALYSIS OF PADDY PRODUCTION IN
BANGLADESH

Abstract

An increase in the output of paddy is of importance in Bangladesh in the wake of growing population and the adverse effects of climate change. Due to limited land availability, the scope for increasing efficiency should be studied. This paper estimates a production frontier using data on paddy production by fitting a Stochastic Frontier model. Having estimated the input elasticity from a Translog model, it can be concluded that more research should be prioritised in bringing about technological improvement through improving the quality of seeds, and the availability of fertilizer and labour should be eased.

Keywords : Stochastic Frontier Analysis, Paddy production, Bangladesh

4.1 Introduction

Rice monoculture is the main productive activity of agricultural households in Bangladesh and is also their staple food. A large proportion of the total cropped area is encompassed by rice cultivation (Rashid et al., 2012). Given an increasing population and the possible adverse effects of climate change, food security in Bangladesh has come of importance. However, given limited land area as well as size of landholding, increasing paddy production implies an increase in the output per unit input usage.

There are three seasons in which paddy is cultivated, namely, Aman, Boro, and Aus. Each of these seasons is characterised by a unique climate and if an optimal production plan is to be prescribed for increasing the output, an analysis of the production frontier as well as technical efficiency is required. With the objective of estimating the output elasticity of the inputs and technical efficiency in paddy production, this paper uses Stochastic Frontier Analysis (SFA) on a sample of plots which belong to paddy-producing farmers in Bangladesh.

The rest of the paper is organized as follows: Section 4.2 presents the methodology in which the SFA model has been described. Section 4.3 explains the variables which have been used in the estimation of the frontier. The results are discussed in Section 4.4, and the conclusion is presented in Section 4.5.

4.2 Methodology

The Translog specification has been used for the production frontier as it is a more flexible functional form as compared to Cobb-Douglas specification. The model has been presented in Equation (4.1).

$$\ln y_i = \beta_o + \sum_i \beta_k \ln x_{ki} + 1/2 \sum_i \sum_j \beta_k \ln x_{ki} \ln x_{lj} + v_i - u_i \quad (4.1)$$

y_i is the output of paddy of plot i . β_o is the intercept. x_{ki} is the k^{th} covariate on plot i . " $\ln x_{ki} \ln x_{lj}$ " is the interaction between the continuous covariates x_i and x_j . v_i is the statistical, symmetric error term, assumed to have a normal distribution with mean 0 and variance σ_v^2 . u_i measures the inefficiency in production and is assumed to have a half-normal distribution with mean μ and variance σ_u^2 .

4.3 Data

The data consists of 7296 plots on which paddy is cultivated from eleven villages of Bangladesh. The time span is 2009-2014. The data was collected in the project, VDSA, of ICRISAT, India. The dependent variable is the production of paddy at the plot level. The independent variables are plot size (in acres), the quantity of seed, fertilizer, labour, precipitation and the depth of soil, as inputs in production. Dummy variables for zero precipitation and fertilizer are included as shifter variables (Battese, 1997). Mean-scaled year, dummy variables for Aman and Boro seasons, high and low slopes of the plot, depth of flooding in the plot during the peak of monsoon are also included as shifter variables ¹. The maximum level of education in the household and age of the head of the household have been included to capture the effect of education and experience, respectively. The descriptive statistics of the variables have been presented in Table 4.1. Figure 4.1 presents the scatter plots of the input quantities and output and they indicate a linear relationship with respect to the output. However, production is higher at lower levels of precipitation. Considering that lower

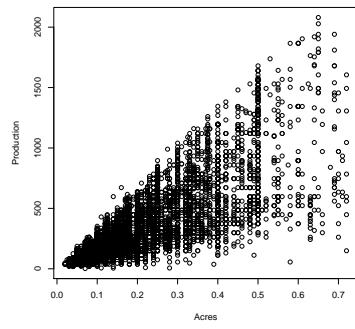
¹Shifter variables are variables which are not to be included in quadratic or interaction form.

precipitation is likely to occur in Boro season, which is the most productive paddy producing season, the plot is justified.

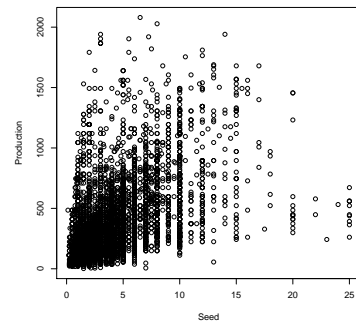
Table 4.1: Summary statistics of continuous variables

Variable	Unit	Mean	SD	Min	Max
Production	Kg.	363.01	292.84	1.75	2080
Size	Acres	0.22	0.13	0.02	0.72
Seed	Kg.	4.05	3.02	0.12	25
Fertilizer	Kg.	28.23	25.67	0	245
Labour	Hours	77.17	51.55	4.25	531
Precipitation	mm	301.5	342.53	0	1074
Soil depth	cms.	13.33	1.19	8	16

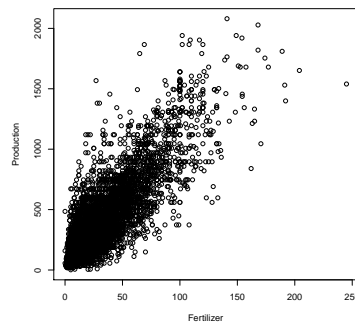
Agro-Ecological Zones (AEZs) exhibited low between-heterogeneity during the estimation of random intercept in the chapter 3, "The role of temporal land use in driving sustainable intensification in Bangladesh: An application of hierarchical modelling". However, given the diversity in the physiography at small distances, it is necessary to account for this source of heterogeneity. For this purpose, the slope as well the depth of soil has been included in the model, at the plot level. The combination of soil properties with seasonal variables such as dummy variables for seasons and the amount of precipitation is expected to substitute for the inclusion of AEZs in the model. Figure 4.2 presents the kernel density of the distribution of rainfall in millimeters (mm.) by the three paddy growing seasons. A comparison of the range of the horizontal X-axis across the subfigures indicates that the maximum rainfall occurs in Aman season, followed by Aus and Boro, in that order.



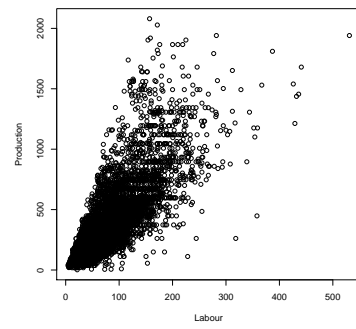
a) Plot size



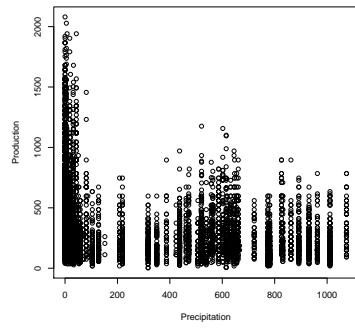
b) Seed



c) Fertilizer



d) Labour



e) Precipitation

Figure 4.1: Scatter plot of inputs with respect to Production

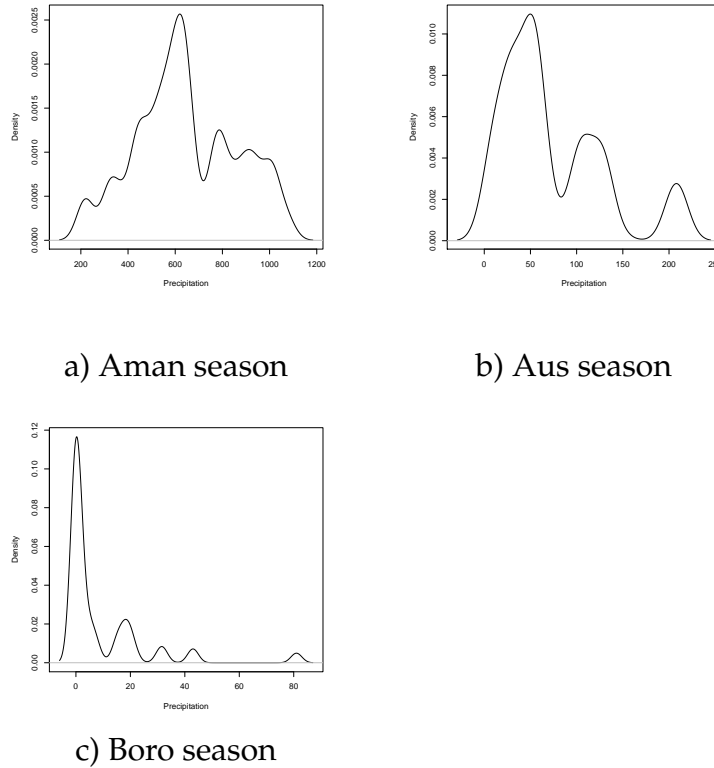


Figure 4.2: The distribution of rainfall by the seasons

4.4 Results

Table 4.2 presents the estimates of the Translog production frontier of total production. The distribution of output elasticities² with respect to plot size, seed, fertilizer, labour and precipitation is given in Figure 4.3. Plot size (acres) reports a very high output elasticity of 1% at the sample mean, given the low average size of the plot. The distribution of the elasticity also indicates a range of high increase. The distribution of elasticity of seed contains positive as well as negative values, and at the sample mean, the value is negative at -0.05%. The coefficient of the interaction between seed and all inputs are negative except with labour. The elasticity of fertilizer is 0.06% at the sample mean, and is significant. Also,

²All inputs except precipitation were mean scaled. Hence, the interpretation of the first order coefficients is at the sample mean.

the coefficient of fertilizer_squared is positive and significant, indicating that there is a large gain in output by increasing the quantity of fertilizer. Labour elasticity is in most cases is positive, as given in Figure 4.3 (d). The coefficient of squared labour is also very high and significant at 0.35, indicating large gains in production with increased labour. The values in the range of the distribution are high, upto 0.70%. At the sample mean, the elasticity is significant at 0.07%. Precipitation elasticity at the sample mean, as given in Table 4.2, is positive and significant. In Figure 4.3 (e), the output elasticity of precipitation is the aggregate of the three seasons of paddy cultivation. The distribution is split according to the three seasons in Figure 4.4. Figure 4.4 (a) presents the elasticity distribution in Aman season, all values of which are negative. Given that most of the rainfall in a year occurs in Aman season, a further increase of one percentage will reduce the output.

The coefficient for the dummy variable for zero precipitation is positive and significant. This can be explained by the fact that most of the zero precipitation values occur in the Boro season, which is the most productive among the three paddy producing seasons. The high productivity is due to the availability and application of irrigation. The rate of technical change, as indicated by the coefficient of Mean-scaled Year is 4%. This high rate of technical change follows 2007-08 food price crisis. The price crisis coincides with Sidr cyclone which had caused widespread floods and reduced production. At the same time, India resorted to an export ban on rice, which restricted imports in Bangladesh. In response to rising prices as a result of the three aforementioned factors, the government of Bangladesh had taken several measures for arresting the spike in prices such as increasing the procurement prices, fertilizer subsidy, as well as improving the availability of electricity, fuel, and seeds. Rural irrigation was

prioritised and was granted an improved electricity distribution (Wiggins et al., 2010). This led to record levels of output after the food price crisis and may explain an improvement in technology.

The coefficients of dummy variables for Aman and Boro seasons are positive and significant, as compared to the reference category, Aus season. They are also high in magnitude. Low slope of the plot is significant in increasing production as compared to the reference, medium slope.

Table 4.2: Translog estimates for total production

Dependent variable: Log (Production)	Estimate	Std. Error
Intercept	-0.2***	0.04
Log(Acres)	1***	0.01
Log(Seed)	-0.05**	0.02
Log(Fertilizer)	0.06***	0.01
Log(Labour)	0.07**	0.02
Log(Precipitation)	0.03**	0.01
Acres_squared	0	0.01
Acres_Seed	-0.02*	0.01
Acres_fertilizer	0.01	0.01
Acres_Labour	0.07***	0.01
Acres_Precipitation	0.	0
Seed_squared	0.02	0.02
Seed_Fertilizer	0	0.01
Seed_Labour	0.09***	0.02
Seed_Precipitation	-0.01*	0
Fertilizer_squared	0.06***	0.02
Fertilizer_Labour	-0.14***	0.02
Fertilizer_Precipitation	0	0
Labour_squared	0.35***	0.05
Labour_Precipitation	0.04***	0
Precipitation_squared	-0.01**	0
Soil depth	-0.01*	0

Education and experience, as given in Table 4.3, increase technical efficiency,

as Z-variables of the production frontier.

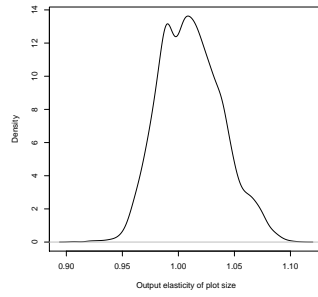
Table 4.3: Translog estimates for total production (contd.)

Dependent variable: Log (Production)	Estimate	Std. Error
Dummy variable for zero precipitation	0.12***	0.01
Dummy variable for fertilizer precipitation	-0.29	0.17
Mean-scaled Year	0.04***	0
Aman season dummy	0.25***	0.02
Boro season dummy	0.69***	0.01
High slope of the plot	-0.02	0.01
Low slope of the plot	0.05***	0.01
Little water flooding during monsoon	0.02	0.02
Knee deep water flooding during monsoon	-0.01	0.02
Chest deep water flooding during monsoon	-0.01	0.02
Maximum education in the household	-0.07***	0.01
Age of the head of the household	-0.01***	0
Sigma squared	0.55***	0.07
Gamma	0.97***	0

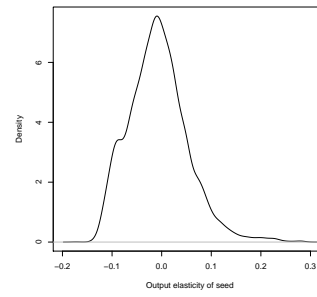
Figure 4.5 presents the kernel density plot of the technical efficiency scores. The mean efficiency is 0.78 and the distribution is highly skewed to the left.

The kernel density distribution of elasticity of scale³ is presented in Figure 4.6. A large part of the sample exhibits decreasing returns to scale, with a value between 0 and 1. Scale inefficiency was also concluded in Rahman (2011).

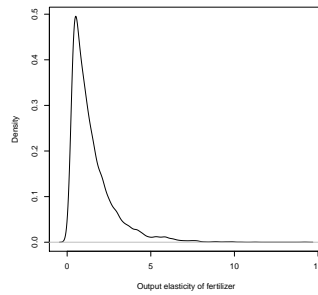
³The inputs used to calculate elasticity of scale were plot size, seed, fertilizer and labour.



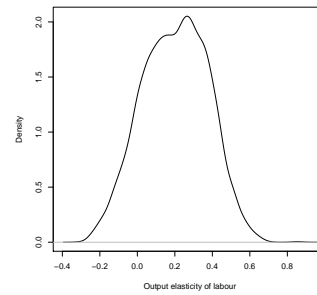
a) Plot size



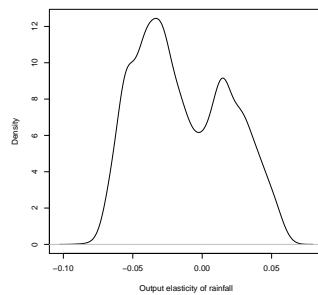
b) Seed



c) Fertilizer



d) Labour

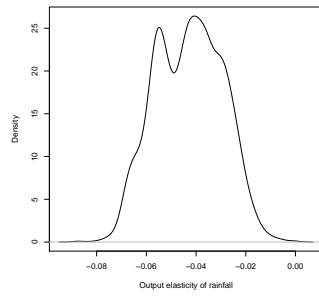


e) Precipitation

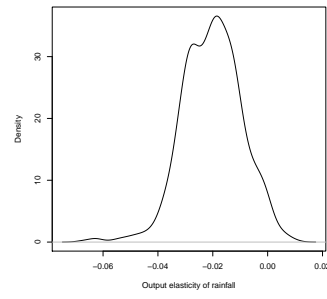
Figure 4.3: Output elasticity of inputs

4.5 Conclusion

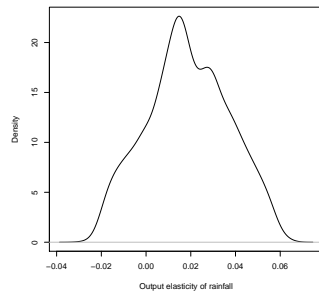
This study examines paddy production in Bangladesh in the time period which immediately followed the food price crisis of 2007-08. It estimates the output elasticity of different inputs as well as that of scale. An increase in the plot size is beneficial to increase production and a consolidation of landholdings may be a solution for improving the status of food security.



a) Aman



b) Aus



d) Boro

Figure 4.4: Output elasticity of rainfall by season

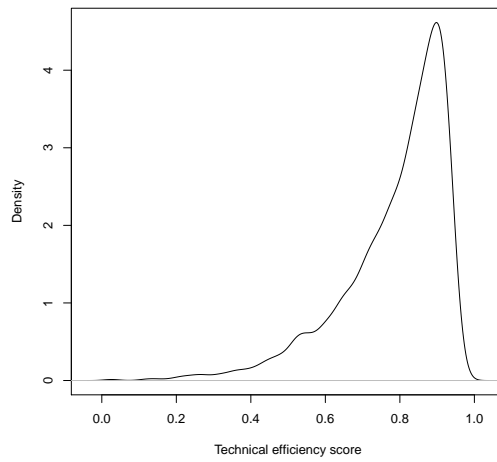


Figure 4.5: Kernel density plot of technical efficiency scores

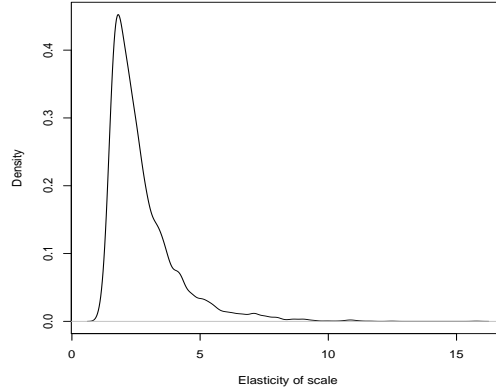


Figure 4.6: Kernel density plot of scale elasticity

Any increase in the quantity of seed will reduce output. Hence, to increase output, a further improvement in the seed type is needed. An increase in fertilizer and labour usage reports high potential for an increase in production and measures should be taken to ease their availability.

Most of the plots operate close to their production frontier with high levels of technical efficiency. This implies that there is a need to shift the production frontier for higher levels of output to ensure food security.

CHAPTER 5

GENERAL CONCLUSION

5.1 Universal Conclusion

This dissertation employs hierarchical modelling for the purpose of verifying its performance in aggregation of technical efficiency across different levels of the units of analysis. Further, hierarchical modelling has been applied for examining the effect of land use choice and fertilizer usage in the extent to which sustainability can be practiced while intensifying paddy production in Bangladesh. Finally, SFA has been applied to elicit the elasticity scenario of different inputs in paddy production in Bangladesh.

Based on the analysis of the chapter 2 and 3, one can conclude that there is scope for integrating hierarchical models in the analysis of production. They exhibit potential in aiding the development of methods for aggregation of efficiency estimates. Further, by estimating varying coefficients by cluster, they aid in the analysis of the marginal product of inputs, thereby prescribing their optimal usage. The analysis of chapter 4 indicates that there is variation in the optimality status of input usage across different plots. Policies affecting the availability of each input have to be sub-region specific and cannot assume universal application across the country.

5.2 Limitations

This section describes the possible boundaries of the applicability of the studies.

In chapter 2, "Efficiency Aggregation in Stochastic Frontier Analysis with Hierarchical Data", all plots which reported zero production and zero harvest labour on account of crop failure have been excluded from the study. To that extent, the total number of plots owned by a household may be under-reported. If decision making regarding cultivation on such plots or if the plots themselves are systematically different from the ones included in the study and are Missing At Random (MAR), the estimated regression parameters β may be biased.

In chapter 3, "The role of temporal land use in driving sustainable intensification in Bangladesh: An application of hierarchical modelling", in spite of a preferably higher variance of the village RI as compared to that of AEZs, the former variance is still lower than the residual variance. Also, the variance of fertilizer as a random slope varying by village is low. This hints at the possibility that there may be other levels of the units of observation at which heterogeneity is higher and has to be accounted for. This possibility is acknowledged due to the diverse physiography of Bangladesh. An ancillary estimation of varying fertilizer slope by households led to higher variance than residual variance of the household RI. For the purpose of completeness in model selection, this chapter demands more attention.

5.3 Scope for further research

The three studies potentially guide further research which can be carried out for gaining comprehensive understanding of their results.

In chapter 2, "Efficiency Aggregation in Stochastic Frontier Analysis with Hierarchical Data", we have exploited the hierarchical structure of the data at the

single level of ownership of multiple plots by households. Given the sampling process, our survey data assumes a multilevel structure, with sublocations and AEZs making up the groups at higher level in the hierarchy. It would be useful to check if one can generalize the results of this study further by incorporating more levels of the data structure and applying our methodology to the same. Also, the inclusion of plot-specific weather information in the analysis will help to account for more variation in production than what is currently explained by the models. A consistent pattern which was observed in the empirical application of the plot-level stochastic frontier model as well as the result of Monte Carlo simulations is that the output weights reduce correlation when testing for the maintenance of ranking of technical efficiency scores, as compared to unweighted means. The cause of this pattern remains to be found and can be used as a question for further research.

In chapter 3, "The role of temporal land use in driving sustainable intensification in Bangladesh: An application of hierarchical modelling", the data on crop production incorporates an additional level of observation at the seasonal level. There is a possibility that a level of observation at the seasonal level is Bangladesh-specific due to three seasons of paddy cultivation. This data structure is suitable for methodological research questions related to aggregation of technical change as well as technical efficiency change to an annual level from estimates at the seasonal level. Also, although the variance of fertilizer random slope is low, its inclusion as a random slope, varying at the household level instead of the village level, had substantially increased the variance of the random intercept (larger than the residual variance). This gives rise to scope for further application of hierarchical modelling at the household level.

APPENDIX A
CHAPTER 1 OF APPENDIX

Table A.1: Count of observations by aggregated AEZs

AEZ	Abbreviation	#
Old Himalayan Piedmont Plain and Tista Floodplain	HPTF	1955
Karatoya Floodplain and Atrai Basin	KFAB	1324
Brahmaputra - Jamuna Floodplain	BJF	756
High Ganges River Floodplain	HGRF	591
Low Ganges River Floodplain	LGRF	116
Ganges Tidal Floodplain	GTF	875
Middle Meghna River Floodplain	MMRF	1468
Greater Dhaka	DHAKA	211
Total		7296

BIBLIOGRAPHY

- Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6:21–37.
- Ali, Y., Waddington, S., Hodson, D., Timsina, J., and Dixon, J. (2008). *Maize-Rice Cropping Systems in Bangladesh : Status and Research Opportunities*. CIMMYT, Mexico, D. F.
- Banker, R. D., Gadh, V. M., and Gorr, W. L. (1993). A Monte Carlo comparison of two production frontier estimation methods : Corrected ordinary least squares and data envelopment analysis. *European Journal of Operational Research*, 67:332–343.
- Battese, G. E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics*, 48(2):250–252.
- Blackorby, C. and Russell, R. R. (1999). Aggregation of Efficiency Indices. *Journal of Productivity Analysis*, 12:5–20.
- Brammer, H. (2012). *The physical geography of Bangladesh*. The University Press Limited.
- Brorsen, B. W. and Kim, T. (2013). Data aggregation in stochastic frontier models: The closed skew normal distribution. *Journal of Productivity Analysis*, 39(1):27–34.
- Castellazzi, M. S., Wood, G. A., Burgess, P. J., Morris, J., Conrad, K. F., and Perry, J. N. (2008). A systematic representation of crop rotations. *Agricultural Systems*, 97:26–33.

- Clark, M. (2016). Thinking About Mixed Models. https://m-clark.github.io/docs/mixedModels/mixedModels.html#mixed_model_5b:_multivariate_normal_model, last accessed on 2020-04-08.
- Cook, W. D., Chai, D., Doyle, J., and Green, R. (1998). Hierarchies and Groups in DEA. *Journal of Productivity Analysis*, 10(2):177–198.
- FAO (2004). *The ethics of sustainable agricultural intensification*. Rome, Italy. <http://www.fao.org/3/j0902e/j0902e03.htm#bm03>, last accessed on 2020-04-08.
- Fitzmaurice, G., Laird, N., and Ware, J. (2011). *Applied Longitudinal Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, second edition.
- Fried, H., Lovell, C. A. K., and Schmidt, S., editors (1993). *The Measurement of Productive Efficiency Techniques and Applications*. Oxford University Press.
- Gatto, M., Petsakos, A., and Hareau, G. (2020). Sustainable Intensification of Rice-Based Systems with Potato in Eastern Indo-Gangetic Plains. *American Journal of Potato Research*, 97:162–174.
- Gelman, A. and Hill, J. (2007). *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press.
- Hassan, R. M., Njoroge, K., Njore, M., Otsyula, R., and Laboso, A. (1998). *Adoption Patterns and Performance of Improved Maize in Kenya in Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*. CAB International.
- Jondrow, J., Knox Lovell, C. A., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2-3):233–238.

- Just, R. and Pope, R. (2001). The agricultural producer: Theory and statistical measurement. In *Handbook of Agricultural Economics*, volume 1A, chapter 12, pages 629–741.
- Khatun, M. A., Rashid, B., and Hygen, H. O. (2016). MET report: Climate of Bangladesh. Technical Report 08, Norwegian Meterological Institute.
- Kim, K., Barham, B. L., and Coxhead, I. (2001). Measuring soil quality dynamics. A role for economists , and implications for economic analysis. *Agricultural and Resource Economics Review*, 25:13–26.
- King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual behaviour from Aggregate Data*. Princeton University Press, Princeton, New Jersey.
- Liu, Y. and Myers, R. (2009). Model selection in stochastic frontier analysis with an application to maize production in Kenya. *Journal of Productivity Analysis*, 31(1):33–46.
- Meeusen, W. and van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, 18(2):435–444.
- Ondrich, J. and Ruggiero, J. (2001). Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research*, 129:434–442.
- Petersen, B. and Snapp, S. (2015). What is sustainable intensification ? Views from experts. *Land Use Policy*, 46:1–10.
- Quddus, M. A. (2009). Crop production growth in different agro-ecological zones of Bangladesh. *Journal of Bangladesh Agricultural University*, 7(2):351–360.

- Rahman, S. (2011). Resource use efficiency under self-selectivity : the case of Bangladeshi rice producers. *The Australian Journal of Agricultural and Resource Economics*, 55:273–290.
- Rahman, S. and Parkinson, R. J. (2007). Productivity and soil fertility relationships in rice production systems , Bangladesh. *Agricultural Systems*, 92:318–333.
- Rashid, A., Alam, K., and Gow, J. (2012). Exploring the relationship between climate change and rice yield in Bangladesh : An analysis of time series data. *Agricultural Systems*, 112:11–16.
- Ruggiero, J. (1999). Efficiency estimation and error decomposition in the stochastic frontier model: a Monte Carlo analysis. *European Journal of Operational Research*, 115(3):555–563.
- Schmidt, P. and Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics*, 2(4):367–374.
- Thomas, A. (2003). A dynamic model of on-farm integrated nitrogen management. *European Review of Agricultural Economics*, 30(4):439–460.
- Wichelns, D. (2016). Managing Water and Soils to Achieve Adaptation and Reduce Methane Emissions and Arsenic Contamination in Asian Rice Production. *Water*, 8(4).
- Wiggins, S., Compton, J., Keats, S., and Davies, M. (2010). Country responses to the food price crisis 2007 / 08. Case studies from Bangladesh, Nicaragua, and Sierra Leone. Technical report, Overseas Development Institute.