



139  
829  
THS

This is to certify that the  
dissertation entitled

PAPERS OF AGRICULTURAL INSURANCE AND FARM  
PRODUCTIVITY

presented by

Yanyan Liu

has been accepted towards fulfillment  
of the requirements for the

doctoral

degree in

Agricultural Economics;  
Economics

*R. J. Myer*

Major Professor's Signature

11/25/2006

Date

*MSU is an Affirmative Action/Equal Opportunity Institution*

LIBRARY  
Michigan State  
University

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

# PAPERS ON AGRICULTURAL INSURANCE AND FARM PRODUCTIVITY

By

Yanyan Liu

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics  
Department of Economics

2006



# **ABSTRACT**

## **PAPERS ON AGRICULTURAL INSURANCE AND FARM PRODUCTIVITY**

By

Yanyan Liu

This dissertation is composed of two distinct papers. The first one is a theoretical paper on agricultural insurance pricing. The second paper studies model selection in stochastic frontier analysis with an application to maize production in Kenya.

The first paper reviews existing agricultural insurance valuation models and provides a new model. The new model takes explicit account of the non-diversifiable market risk inherent in offering insurance contracts, and demonstrates how capital markets can facilitate risk spreading and diversification. The analysis suggests that present value models may provide appropriate insurance valuations in some circumstances, but the standard Black-Scholes model has deficiencies for pricing agricultural insurance. Other existing methods for pricing the market risk in agricultural insurance contracts are logically consistent and potentially useful. However, the heterogeneous agent equilibrium model developed here is easy to use, amenable to empirical estimation, and provides a simple and intuitive way to value market risk in agricultural insurance contracts.

The second paper shows how to estimate the quantitative magnitude of partial effects of exogenous firm characteristics on technical efficiency (along with their standard errors) under a range of popular stochastic frontier model specifications. An  $R^2$ -type measure

is also derived to summarize the overall explanatory power of the exogenous factors on firm inefficiency. The paper also applies a recently developed model selection procedure to choose among alternative stochastic frontier specifications using data from household maize production in Kenya. The magnitude of estimated partial effects of exogenous household characteristics on inefficiency turns out to be very sensitive to model specification, and the model selection procedure leads to an unambiguous choice of best model. Bootstrapping is used to provide evidence on the size and power of the model selection procedure. The empirical application also provides further evidence on how household characteristics influence technical inefficiency in maize production in developing countries.

To my husband, Shanjun Li,  
my parents, Guiying Liu and Shujun Chen,  
my brother, Weiwei Chen,  
and my grandparents, Honglan Li and Honglin Liu

## ACKNOWLEDGMENTS

I would like to thank my committee, Drs. Robert Myers (Co-Chair), Peter Schmidt (Co-Chair), Steven Hanson, Jack Meyer, and Roy Black for providing me with valuable insights in their respective areas of expertise. My deepest gratitude and respect go to Drs. Myers and Schmidt for being wonderful mentors and for constant support throughout my PhD program. I will be forever indebted to them.

My sincere thanks also go to my master's advisor, Dr. Scott Swinton, for his genuine support during my stay at Michigan State University, and to Dr. Roy Black who is always willing to go out of his way to help me. My sincere thanks also extend to Dr. Steven Hanson, who helped support my research program, and Dr. Eric Crawford, for his help and support to my husband and me throughout our time at Michigan State. I would also like to thank Dr. Thomas Jayne for generously providing me the data in the second paper of my dissertation.

Thanks also to all my fellow graduate students for their friendship and encouragement. A special mention goes to Ren Mu, Zhiying Xu, Wei Zhang, Christopher Wright, Denys Nizalov, Lesiba Bopape, Antony Chapoto, Elliot Mghenyi, Kirimi Sindi and Gerald Nyambane.

Last but not the least, I thank my husband, Shanjun Li, my parents, Guiying Liu and Shujun Chen, my brother, Weiwei Chen, and my grandparents, Honglan Li and Honglin Liu for their love and support.

# TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	<b>viii</b>
<b>LIST OF FIGURES</b> . . . . .	<b>ix</b>
<b>1 INTRODUCTION</b> . . . . .	<b>1</b>
<b>2 HOW SHOULD WE PRICE THE MARKET RISK IN AGRICULTURAL INSURANCE CONTRACTS?</b> . . . . .	<b>4</b>
2.1 Introduction . . . . .	4
2.2 A Simple Insurance Model . . . . .	5
2.3 Existing Approaches to Valuing Agricultural Insurance Contracts . . . . .	6
2.3.1 Present Value Models . . . . .	6
2.3.2 Arbitrage-Based Option Pricing Models . . . . .	9
2.3.3 Lucas General Equilibrium Models . . . . .	11
2.3.4 The Chambers State-Contingent Approach . . . . .	13
2.4 An Alternative Valuation Model . . . . .	14
2.4.1 Mutual Fund Model . . . . .	15
2.4.2 Discussion . . . . .	18
2.4.3 Insurance Industry Model . . . . .	19
2.5 Simulation Results . . . . .	21
2.6 Conclusions . . . . .	25
<b>3 MODEL SELECTION IN STOCHASTIC FRONTIER ANALYSIS: MAIZE PRODUCTION IN KENYA</b> . . . . .	<b>27</b>
3.1 Introduction . . . . .	27
3.2 Stochastic Production Frontier Models . . . . .	30
3.2.1 Alternative Model Specifications . . . . .	32
3.3 Empirical Application . . . . .	34
3.3.1 Data . . . . .	35
3.3.2 Variables in the Production Frontier . . . . .	36
3.3.3 Exogenous Factors Affecting Efficiency . . . . .	38
3.4 Estimation Results from Competing Models . . . . .	40
3.5 Model Selection . . . . .	46
3.5.1 Empirical Model Selection . . . . .	46

3.5.2	A Bootstrap Evaluation . . . . .	47
3.6	Post-Estimation Analysis . . . . .	50
3.7	Conclusion . . . . .	53
<b>APPENDICES . . . . .</b>		<b>54</b>
A	Derivation of Equation (2.2) . . . . .	55
B	Derivation of Equation (2.10) . . . . .	56
C	Derivation of Equation (2.13) . . . . .	58
D	Estimating Partial Effects of Exogenous Factors and their Standard Errors for the General Model . . . . .	59
E	Results of Specifying Relevant Explanatory Variables . . . . .	62
<b>BIBLIOGRAPHY . . . . .</b>		<b>65</b>

## LIST OF TABLES

2.1	Summary of the parameters in simulation model . . . . .	22
3.1	Descriptive statistics for the variables in the production frontier . . . . .	38
3.2	Descriptive statistics for the exogenous variables in the efficiency model . . .	40
3.3	Estimates for the production frontier in alternative models . . . . .	43
3.4	Estimates for the inefficiency components in alternative models . . . . .	44
3.5	Correlation of efficiency estimates among alternative models . . . . .	44
3.6	Partial effects of exogenous factors, evaluated at the sample mean . . . . .	45
3.7	Average partial effects of EDUHIGH on $E(-u_i x_i, z_i)$ , for the observations within each of the four quartiles based on efficiency levels predicted in KGMHLBC model . . . . .	46
3.8	Correlation of partial effects of EDUHIGH on $E(-u_i x_i, z_i)$ among alternative models . . . . .	46
3.9	Results of specification tests for model selection . . . . .	47
3.10	Partial effects of the exogenous factors on $E(-u_i x_i, z_i)$ and their 90% confidence intervals based on bootstrap and the delta method in the KGMHLBC model, evaluated at the sample mean . . . . .	50
3.11	Output elasticity with respect to inputs for local seed users and hybrid seed users, evaluated at the sample means . . . . .	50
E.1	Specifying variables in the production frontier using OLS . . . . .	62
E.2	Tests results for specifying the exogenous factors in the efficiency component	62
E.3	Tests results for specifying explanatory variables in the production frontier .	63

## LIST OF FIGURES

2.1	Insurance premiums versus interest rate . . . . .	24
2.2	Insurance premiums versus market risk factor . . . . .	24
2.3	Insurance premiums versus share of farm income in aggregate consumption .	25
3.1	Location of sample villages in Kenya . . . . .	35
3.2	Kernel density estimate based on Battese and Coelli technical efficiency estimates . . . . .	51



# CHAPTER 1

## INTRODUCTION

This dissertation deals with two important issues in agricultural production and risk management. The first paper is on how to price the market risk in agricultural insurance contracts and the second paper identifies and quantifies the determinants of technical efficiency in agricultural production using stochastic frontier analysis.

Agriculture is a highly risky industry. Producers are constantly faced with market price risk as well as production risk caused by weather variations. Since its birth in 1938 in the United States, agricultural insurance has played an increasingly significant role in risk management for farmers. In 2004, the coverage of agricultural insurance in the US was over 46 billion dollars and the premium subsidy was about 2.5 billion dollars. Although subject to many criticisms, the contribution of agricultural insurance in promoting farmer welfare and agricultural development is difficult to refute. Therefore, agricultural insurance programs are being expanded both in the U.S. and other countries. Like all types of insurance, a central question of agricultural insurance is how to price insurance contracts. In recent years, traditional actuarial pricing methods have been gradually overtaken by financial valuation models, particularly in academic research. Financial valuation models are considered superior to traditional actuarial pricing models because they take into account the role played by diversification and portfolio management in determining the market price of the risk

embodied in insurance contracts.

The first paper on agricultural insurance has two objectives. The first is to review existing approaches to valuing agricultural insurance contracts and point out some of their advantages and disadvantages. The second objective is to develop a new valuation model which is more transparent about the way in which market risk is valued, and which therefore provides additional insights into the valuation of agricultural insurance contracts. The new model takes explicit account of some important institutional features of agricultural insurance contracting, such as the fact that only farmers can buy insurance and only insurance firms or the government offer insurance. Using simulation we also show why, and under what conditions, alternative modeling approaches give very different answers to the insurance valuation question.

The second paper in the dissertation is motivated by the widespread hunger problem in Kenya. The rapidly growing population and declining growth rate of the agricultural sector has resulted in a pressing problem of food security in Kenya. Maize is the primary staple food in the Kenyan diet, and about 75% of farmers are engaged in maize production. Since the area of arable land is stagnant, promoting farm productivity in the maize sub-sector has attracted increased attention due to the significance of maize in food security. The factors that possibly influence farm efficiency include infrastructure, education, information technology, access to extension services, and so on. This paper aims to identify the determinants of farm efficiency and to quantify the effects of these factors.

The methodology used is stochastic production frontier analysis allowing for specific exogenous influences on farm efficiency. The stochastic frontier model was first developed in 1977 by two groups of researchers independently: Meeusen and van den Broeck in a paper published in the *International Economic Review*, and Aigner, Lovell and Schmidt in a paper published in *Journal of Econometrics*. Since then, it has been a major tool in productivity analysis and widely used in various industries including agriculture, banking, electric utilities, railways, and so on. Compared with traditional methods of production

function estimation, the stochastic frontier model relaxes the assumption that all producers are successful in producing maximal possible output given their input levels. The distance between the actual output and the maximal possible output is called technical inefficiency. This inefficiency may exist due to explicit or implicit constraints faced by producers, or simply because producers make mistakes in arranging their production activities. Thus, the stochastic frontier model allows for differentiation in efficiency levels across producers. By associating inefficiency with exogenous factors, the stochastic frontier model can be used to identify the sources of inefficiency.

Besides addressing the empirical question of the determinants of technical efficiency in maize production in Kenya, this paper also makes some methodological contributions to the stochastic frontier literature. First, we provide a method for estimating the quantitative magnitude of the partial effects of exogenous firm characteristics on firm inefficiency, show how to put standard errors around these partial effects, and propose an  $R^2$ -type measure to summarize the overall explanatory power of the exogenous factors on inefficiency. Second, we show that while alternative models of the relationship between household characteristics and technical inefficiency tend to provide the same direction of the influence of household characteristics, the magnitudes of the partial effects on firm inefficiency are quite sensitive to model selection. Third, we show how a recently developed model selection procedure (Alvarez, Amsler, Orea, and Schmidt, 2006) can be used to choose among competing models. A novel detail is that we use bootstrapping to provide evidence on the power and size of this procedure. The model selection procedure gives an unambiguous choice of best model for our application to Kenyan maize production. This is important because if different models give different results, and we cannot distinguish statistically among the models, we do not know which set of results should be used. But if we can pick a clearly best model, then the fact that inferences and conclusions are sensitive to model selection is not such a problem.

# **CHAPTER 2**

## **HOW SHOULD WE PRICE THE MARKET RISK IN AGRICULTURAL INSURANCE CONTRACTS?**

### **2.1 Introduction**

Valuation of agricultural insurance contracts has become an increasingly important issue in recent years as the Risk Management Agency (RMA) continues to expand the range of insurance products offered to farmers. Traditional farm-based multiple peril yield insurance has been supplemented with area-based yield insurance, revenue insurance products (both farm-based and area-based), and catastrophic risk coverage. The number of eligible commodities has also expanded to include non-traditional and specialty crops, as well as livestock products. Alternative approaches to valuing agricultural insurance contracts have therefore come under increased scrutiny in both the academic literature and in the applied construction of premium schedules for new insurance products (e.g. [46]; [49]; [57]; [48];

[41]; [7]; [15]).

This paper has two objectives. The first is to review existing approaches to valuing agricultural insurance contracts and point out some of their advantages and disadvantages. The second objective is to develop a new valuation model which is more transparent about the way in which market risk is valued, and which therefore provides additional insights into the valuation of agricultural insurance contracts. The new model takes explicit account of some important institutional features of agricultural insurance contracting, such as the fact that only farmers can buy insurance and only insurance firms or the government offer insurance. The new model also accounts for the presence of uninsurable background risk which is a pervasive feature of most agricultural insurance environments. Using simulation we also show why, and under what conditions, alternative modeling approaches give very different answers to the insurance valuation question. The intention is to improve our understanding of the theoretical issues underlying agricultural insurance valuation, as well as to provide a new approach that may eventually lead to improved applied ratings of agricultural insurance contracts.

To begin, we fix ideas by outlining a simple generic agricultural insurance contract. Next we review existing approaches to valuing the generic contract and discuss their advantages and disadvantages. Then our alternative model is presented, followed by simulation results which highlight model differences and show why different valuation models can lead to such different results. Finally there are some concluding comments.

## 2.2 A Simple Insurance Model

To fix ideas we focus on index-based farm revenue insurance contracts. Farmers are indexed by  $i = 1, 2, \dots, n_Y$  and endowed with random farm income  $Y_i$  which may be correlated across farms. There are  $m$  regional farm revenue indices  $Z_j$  for  $j = 1, 2, \dots, m$  that are correlated with the  $Y_i$  and which form the basis for insurance contracts. If  $m = n_Y$  and

$Z_j = Y_i$  for all  $i = j$  then the insurance is individual farm revenue insurance. But generally  $m < n_Y$  and the indices will be based on regional or area farm revenue as opposed to individual farm revenue. This index approach makes the insurance more consistent with actual area-based farm revenue insurance programs and eliminates the need to account for moral hazard or adverse selection.<sup>1</sup>

An insurance contract on  $Z_j$  is a contingent claim that costs  $P_j(G_j)$  and pays off  $V_j(G_j) = \max(G_j - Z_j, 0)$ , where  $G_j$  is a guaranteed level of regional farm revenue and  $P_j(G_j)$  is a premium schedule. We are interested in valuing  $P_j(G_j)$  taking proper account of the non-diversifiable market risk embodied in offering insurance contracts.

## 2.3 Existing Approaches to Valuing Agricultural Insurance Contracts

We now examine several existing approaches to valuing agricultural insurance within the context of the simple insurance model of the preceding section.

### 2.3.1 Present Value Models

The simplest and still most common approach to valuing agricultural insurance contracts is to set the premium schedule equal to the present value of the expected indemnity:

$$P_j(G_j) = \beta E[V_j(G_j)] = \beta \int_0^{G_j} (G_j - Z_j) f(Z_j) dZ_j, \quad (2.1)$$

where  $\beta = 1/(1 + r)$  is a discount rate based on the risk-free rate of interest  $r$ ,  $E$  is expectation conditional on information available when the contract is valued (sold), and  $f(\cdot)$  is the density function for the relevant insurance index. For most U.S. agricultural insurance contracts administered through the RMA, premium payment is not required

---

<sup>1</sup>That is, we assume individual farms cannot influence regional revenues and that the probability distribution of regional revenues is known to all.

until the index is realized and claims are paid.<sup>2</sup> In this case we would set  $\beta = 1$  because no discounting would be necessary.

The present value model (2.1) has the advantage of being easy to calculate under a wide range of potential probability distributions for  $Z_j$  because it is straightforward to evaluate the integral in (2.1) numerically for quite flexible probability distributions. Not surprisingly, this has led to a significant amount of research focused on what the right probability distribution for the integration should be ([35]; [34]; [19]; [25]). The present value model also has the advantage that the formula is independent of the returns to other assets in the economy (other than  $r$ ).

For ease of comparison with other models outlined below, we evaluate the integral in (2.1) assuming the insurance index is lognormally distributed. Using results from the appendix of Rubinstein ([43]) (reproduced in appendix A) this gives:

$$P_j(G_j) = \beta \left[ G_j N \left( \frac{g_j - \mu_j}{\sigma_j} \right) - e^{\mu_j + 0.5\sigma_j^2} N \left( \frac{g_j - \mu_j - \sigma_j^2}{\sigma_j} \right) \right], \quad (2.2)$$

where  $N(\cdot)$  is the cumulative distribution function for the standard normal,  $g_j = \ln(G_j)$ , and  $\mu_j = E(z_j)$  and  $\sigma_j^2 = Var(z_j)$  are the mean and variance of  $z_j = \ln(Z_j)$ , conditional on information available when the insurance is purchased. This formula has the advantage that it is easy to compute without resorting to numerical integration or Monte Carlo methods because  $N(\cdot)$  is already compiled and available in most computational software programs.

It should be immediately obvious that the present value model (2.1) and (2.2) values the market risk inherent in issuing insurance contracts at zero. Put another way, the present value model is consistent with an insurance market equilibrium in which insurers act as if they are risk neutral and incur zero operating costs (see [42]). There are two situations in which valuing the market risk at zero might be appropriate. The first is if the risks being insured are fully diversifiable. For example, this might be reasonable in the case of

---

<sup>2</sup>This is a very unusual insurance feature because most insurance schemes require up-front payment of premiums to ensure compliance. However, it is a common feature of agricultural insurance used to encourage broader farmer participation, and therefore needs to be accounted for in agricultural insurance valuation.

auto insurance where a large number of independent risks are being pooled across insureds leaving negligible aggregate risks for insurers. It is generally acknowledged, however, that agricultural revenue risks are covariate because individual contract losses have a tendency to move up and down together. In this case, even pooling large numbers of contracts together leaves aggregate market risk remaining (see [33]; and [17]). Of course, this still leaves the possibility of pooling aggregate agricultural risks with other types of (uncorrelated) insured risks via the reinsurance market. However, there are a number of reasons why reinsurance markets may not be able to diversify all of the aggregate risk inherent in holding agricultural insurance portfolios (see [29]; [23]; [45]; [17]).

The second situation in which valuing the market risk of agricultural insurance at zero might be appropriate is if the insurance is being underwritten by the government which spreads the indemnity risk across all taxpayers. If the number of taxpayers is large relative to the size of the risk then it has been argued that the risk per individual taxpayer is negligible and the insurance should be priced as if the market risk is zero ([6]). However, while this argument may be reasonable for agricultural insurance in a very large developed economy like the U.S., it is unlikely to hold in developing countries where a much larger proportion of the total population is engaged in agriculture. Furthermore, even if agricultural insurance is being underwritten by the government and spread across a large number of taxpayers, it might still be useful to value the market risk that would have occurred without the government underwriting, because this will provide a more complete picture of the degree of subsidy that is inherent in the underwriting.

For all of these reasons there have been several attempts to relax the assumption of no market risk, and begin to build more general, market-based models of agricultural insurance valuation.



### 2.3.2 Arbitrage-Based Option Pricing Models

The contingent payoffs on insurance contracts are identical to the payoffs on a put option written on the underlying insurance index  $Z_j$ .<sup>3</sup> This insight has led a number of researchers to apply arbitrage-based option pricing models to value agricultural insurance contracts ([53]; [52]; and [49]). The advantage of the arbitrage-based approach is that valuation is based on a fully specified equilibrium asset pricing model that implicitly includes a market value for the risk involved in holding portfolios of insurance contracts and other assets.

Initial applications of the option pricing approach used the Black-Scholes ([11]) formula which, using our notation and under the usual assumption of lognormal  $Z_j$ , can be expressed as:

$$P_j(G_j) = \beta G_j N\left(\frac{g_j - z_j^0 - r + 0.5\sigma_j^2}{\sigma_j}\right) - Z_j^0 N\left(\frac{g_j - z_j^0 - r - 0.5\sigma_j^2}{\sigma_j}\right), \quad (2.3)$$

where  $Z_j^0$  is the “initial value” of the index when the insurance is taken out and  $z_j^0 = \ln(Z_j^0)$ . While this formula does include an implicit value for the market risk from trading the option, it suffers from two weaknesses for valuing agricultural insurance. First, equation (2.3) assumes that the option is paid for when it is purchased but most U.S. agricultural insurance does not require premium payment until maturity when the index is realized and claims paid. Of course, this problem is trivial and can be overcome by simply compounding the formula through to the maturity date (i.e., multiplying the formula by  $1/\beta$ ). A second and more important weakness for pricing agricultural insurance is that the Black-Scholes model prices the option (insurance) by assuming that the underlying index  $Z_j$  is the price of a tradable asset. The formula is then derived by constructing a time-varying portfolio consisting of the underlying asset and a risk-free bond that exactly replicates the time-varying return on the option (insurance contract). Imposing the no arbitrage condition (i.e., equating the returns on the portfolio and the option) under an assumption of lognormally

---

<sup>3</sup>A put option gives the owner the right, but not the obligation, to sell a unit of an underlying asset whose price becomes  $Z_j$  at maturity, at a given strike price  $G_j$  determined when the option is written. The payoff on the option therefore becomes  $\max(G_j - Z_j, 0)$  which is equivalent to the payoff on an insurance contract written on  $Z_j$ .

distributed  $Z_j$  then leads to (2.3). This approach may be reasonable for financial options written on tradable assets such as stocks but in most agricultural insurance applications  $Z_j$  will be an index that is not the price of a tradable asset, and so the no arbitrage argument which forms the basis of the Black-Scholes formula breaks down (see [49]; [57]; [48]).

The non-tradability of agricultural insurance indices has led more recent applications of arbitrage based option pricing models to use a variant of the Black-Scholes model that allows the option to be written on a non-tradable asset ([16]; [51]). The pricing formula for options written on non-tradable assets or indices, again assuming a lognormally distributed index, can be written using our notation as:

$$P_j(G_j) = \beta \left[ G_j N \left( \frac{g_j - \mu_j + \lambda \sigma_j}{\sigma_j} \right) - e^{\mu_j + 0.5 \sigma_j^2 - \lambda \sigma_j} N \left( \frac{g_j - \mu_j - \sigma_j^2 + \lambda \sigma_j}{\sigma_j} \right) \right], \quad (2.4)$$

where  $\lambda$  is the so-called “market price of risk” for the insurance contract. Again, this formula assumes the option or insurance is paid for at the time it is taken out so, for the case of agricultural insurance where premiums are paid at maturity, the formula would need to be compounded to the maturity date (i.e., multiplied by  $1/\beta$ ). It is interesting to note that if the market price of risk is zero ( $\lambda = 0$ ) then (2.4) gives exactly the same formula as the present value model under lognormality (2.2). This shows that (2.4) is just a generalization of the present value model (2.2) that incorporates a value for the market risk from offering insurance contracts.

Nevertheless, there are two major weaknesses in using (2.4) to value agricultural insurance contracts. First, while a value for the market risk is included explicitly, the market price of risk  $\lambda$  remains an undetermined coefficient. It has been suggested that  $\lambda$  can be estimated using an auxiliary asset pricing model such as the capital asset pricing model (CAPM) or arbitrage option pricing theory (APT) (see [57]; and [48]). However, computing an appropriate value for the market price of the risk embodied in the insurance index  $Z_j$  using the CAPM or APT may not be easy and, even if it can be done, why not just use the CAPM or APT to value the insurance product directly?<sup>4</sup> Second, the arbitrage-based formula (2.4)

---

<sup>4</sup>Of course, the reason why the CAPM or APT are not used to price the option (insurance) directly

for pricing an option written on a non-tradable asset is constructed by assuming that, even though  $Z_j$  is not the price of a tradable asset, there exists a portfolio of tradable assets whose risk spans the risk in  $Z_j$  (i.e. whose returns are perfectly correlated with  $Z_j$ ). Put another way, it implicitly assumes that the option (insurance contract) is a redundant asset because its returns can be replicated using a time-varying portfolio of the spanning portfolio and a risk-free bond.<sup>5</sup> There are many assets, such as commodity futures and options contracts, stocks of agribusiness and food companies, etc., that might be included in a spanning portfolio for  $Z_j$ . Yet to our knowledge there has never been any convincing evidence presented that such a spanning portfolio exists for agricultural insurance contracts. Indeed, all of the recent interest in developing new agricultural insurance products would suggest that agricultural insurance contracts are not redundant assets.<sup>6</sup>

### 2.3.3 Lucas General Equilibrium Models

Another approach that has been used to price agricultural insurance contracts is the representative agent general equilibrium asset pricing model of Lucas ([30]). This model prices contingent claims (including insurance) using an equilibrium pricing kernel derived from a dynamic optimization problem and the imposition of market clearing conditions. The equilibrium pricing formula takes the form:

$$P_j(G_j) = \delta E \left[ \frac{U'(C)}{U'(C_0)} V_j(G_j) \right], \quad (2.5)$$

---

is that they are viewed as having more restrictive assumptions than those underlying the arbitrage-based models. But if these more restrictive assumptions have to be invoked to obtain a value for the market price of risk then they are, at least to some extent, implicitly imbedded in the arbitrage based model for nontradables anyway.

<sup>5</sup>This assumption can be relaxed slightly by allowing the return on the spanning portfolio to track the risk in  $Z_j$  with error, provided the tracking error is completely diversifiable using existing asset markets [see [32]]. Nevertheless, the valuation approach is still based on the notion that there exists a “spanning portfolio” that generate returns that are (almost) perfectly correlated with the with movements in  $Z_j$ .

<sup>6</sup>Conversely, one might argue that the fact that farmers seem unwilling to buy agricultural insurance contracts without sizable subsidies indicates that the contracts may be redundant. However, this argument is incomplete at best because even if the insurance is redundant risk-averse farmers should be willing to purchase actuarially fair insurance.

where  $U(\cdot)$  is a concave von Neumann-Morgenstein utility function,  $\delta$  is a time preference parameter,  $C$  is consumption in the period when  $Z_j$  is realized and insurance indemnities are paid, and  $C_0$  is consumption in the previous period when the insurance is taken out. In equilibrium, the representative agent consumes all of the economy's endowments (see [30]).

The Lucas model is a representative agent model and the utility function and time preference parameter are that of the representative agent. Notice that if the representative agent is risk neutral then (2.5) just reduces to the present value formula (2.1), with discounting occurring at the rate of time preference parameter  $\delta$ .<sup>7</sup> But if the representative agent is risk averse then the ratio of marginal utilities represents an adjustment factor that compensates investors for taking on the risk of issuing the insurance contracts. This approach has been used to price weather insurance products for agriculture (see [12] and [41]).

There are several problems with using the Lucas model (2.5) to price agricultural insurance products. First, whose consumption should be used to compute the marginal utility of consumption? Should it be aggregate consumption in the economy or consumption of farmers or some local agricultural region?<sup>8</sup> There is no reason why it should be farm or regional consumption because all agents in the economy can potentially invest in firms offering agricultural insurance. And yet aggregate economy-wide consumption tends to have low correlation with agricultural crop revenues in the U.S. and so applying (2.5) using aggregate economy-wide consumption tends to just reduce to the risk neutral present value formula (2.1). Another problem in computing the marginal utility of consumption is that a specific form must be assumed for the utility function. Clearly, insurance pricing results may be sensitive to which utility function and consumption variable are included in the analysis. More importantly, while the Lucas model (2.5) does not require the existence of a spanning portfolio, it is a single good representative agent exchange model in which no trades of

---

<sup>7</sup>Of course, this implies that insurance premiums are paid when the insurance is taken out, rather than when  $Z_j$  is realized and indemnities are paid. If premiums only need to be paid at the time indemnities are paid out then the Lucas model becomes  $P_j(G_j) = E \left[ \frac{U'(C)}{E[U'(C)]} V(G_j) \right]$ .

<sup>8</sup>It have been done both ways (see [12] and [41]).

any commodity or asset occur in equilibrium (no trades being necessary because all agents are identical). Hence the model generates equilibrium pricing kernels for contingent claims but the claims are redundant and never traded in equilibrium. Put another way, the Lucas general equilibrium asset pricing model is implicitly a complete markets model and so again prices insurance contracts assuming they are redundant assets.

### 2.3.4 The Chambers State-Contingent Approach

A more recent contribution to agricultural insurance pricing is the state-contingent approach of Chambers ([15]). The Chambers model focuses on the farmer's role as a producer rather than as a consumer, and derives the value of an insurance contract to farmers using the notion that farmers will never forego an opportunity to raise profit risklessly, regardless of their risk preferences. The resulting valuation formula takes the form:

$$P_j(G_j) = E \left[ \frac{\partial k(a, \tilde{b})}{\tilde{d}} V_j(G_j) \right], \quad (2.6)$$

where  $k(a, \tilde{b})$  is the producer's cost of producing state contingent output  $\tilde{b}$  given input prices  $a$ ,  $\partial k(a, \tilde{b})$  is the Gateaux derivative of  $k$ , and  $\tilde{d}$  is the state contingent output price. The basic idea is that the return on any asset held by the farmer can be replicated physically by investing in alternative stochastic output levels, and this equivalence can be used to value the farmer's willingness to pay for agricultural insurance.

The Chambers model takes a novel perspective that provides some new insights into how farmers value insurance. Furthermore, it has the advantage that the valuation formula is independent of risk preferences (though it does depend on the form of the cost function). Nevertheless, it is not yet a complete general equilibrium model because it only focuses on the farmer's decision and the farmer's willingness to pay for insurance. Equilibrium insurance valuation will also depend on the decisions of those offering insurance and this side of the market is not treated in the Chambers paper. So while the Chambers approach gives some interesting insights into farmer willingness to pay for insurance, it is not yet

an equilibrium valuation model in the same sense as the other valuation models discussed above.<sup>9</sup>

## 2.4 An Alternative Valuation Model

Our alternative valuation model is a general equilibrium approach in the spirit of Lucas ([30]) but we allow for agent heterogeneity and uninsurable background risk. Hence, insurance contracts are not redundant assets and trade in insurance contracts actually takes place. We also allow for some important institutional features of agricultural insurance, such as that only farmers residing in a region can purchase that region’s area-based insurance contract, and only insurance companies or the government offer insurance (insurance contracts not offered or traded among individual agents).

We use the area-based insurance contract and notation outlined in the beginning of the paper but now need to add more detail to the economy. Each of the  $n_Y$  farmers in the economy are indexed by  $i = 1, 2, \dots, n_Y$  and have two endowments of income—farm income  $Y_i$  and non-farm income  $W_i$ , which could consist of wage income from off-farm employment, returns from stock market and other investments, net borrowing etc. Both  $Y_i$  and  $W_i$  are random at the time farmers have the opportunity to buy insurance, and their joint distribution is heterogeneous across farmers and allows the two income sources to be correlated. There are also  $n - n_Y$  “wage-earners” in the economy who are indexed by  $i = n_Y + 1, n_Y + 2, \dots, n$  and only have endowments of non-farm income  $W_i$ , which again is random at the time insurance decisions are made. Insurance contracts are available on the regional revenue indexes  $Z_j$  but there is uninsurable background risk generated by the imperfect correlation between  $Z_j$ ,  $Y_i$ , and  $W_i$ . We begin with a mutual fund model and then introduce some common features of insurance industries to examine an alternative insurance industry equilibrium.

---

<sup>9</sup>Of course, the Chambers model could be turned into an equilibrium model by making additional assumptions regarding sellers of insurance and the trading of insurance contracts.

### 2.4.1 Mutual Fund Model

Suppose there exist  $m$  insurance contracts on regional revenue indices  $Z_j$  for  $j = 1, 2, \dots, m$  with prices  $P_j(G_j)$  and indemnities  $V_j(G_j) = \max(G_j - Z_j, 0)$  where  $G_j$  is the guarantee level defined earlier. These contracts can be traded freely and costlessly among all agents in the economy in any continuous amount desired. Short and long sales are allowed (i.e. all agents are allowed to offer as well as purchase insurance). The budget constraint of an arbitrary agent  $i$  can then be expressed as:<sup>10</sup>

$$C_i \leq Y_i + W_i + \sum_{j=1}^m X_{ij} [V_j(G_j) - P_j(G_j)] \quad \forall i = 1, 2, \dots, n, \quad (2.7)$$

where  $C_i$  is agent  $i$ 's consumption and  $X_{ij}$  is the amount of insurance on index  $Z_j$  purchased (sold if negative) by agent  $i$ . Notice that this is essentially a mutual fund model where optimal sharing of the insurable risks can be obtained via unconstrained and competitive trade in insurance contracts. Also, we are assuming premium and indemnity payments both occur at the same time (as is typical in agricultural insurance) and so they have a net effect on income available for current consumption.

Each agent's decision problem is to choose insurance amounts to maximize the expected utility of consumption,  $E[U_i(C_i)]$  subject to (2.7), where each  $U_i$  is an increasing and concave von Neumann-Morgenstern utility function. Necessary conditions for a maximum are:<sup>11</sup>

$$E\{U'_i(C_i)[V_j(G_j) - P_j(G_j)]\} = 0 \quad \forall i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m. \quad (2.8)$$

This condition is virtually identical to that obtained from Lucas ([30]) under the assumption that premiums are collected at the time indemnities are paid (see footnote 7). The only difference is that we allow for heterogeneous preferences, endowments, and probability distributions while the Lucas model is a representative agent model.

---

<sup>10</sup>For "wage earner" agents  $i = n_Y + 1, n_Y + 2, \dots, n$ , then  $Y_i$  will be zero with probability one.

<sup>11</sup>Second-order conditions are satisfied by the concavity of  $U_i$ .

We close the model by assuming the  $m$  insurance markets all clear,  $\sum_{i=1}^n X_{ij} = 0 \quad \forall j$  which implies from (2.7) that the aggregate budget constraint is given by:

$$\sum_{i=1}^n C_i = \sum_{i=1}^n Y_i + \sum_{i=1}^n W_i \Rightarrow C = Y + W \quad (2.9)$$

where  $C$ ,  $Y$ , and  $W$ , are economy aggregate consumption, farm income, and non-farm income respectively. A mutual fund equilibrium is a set of consumption choices, insurance choices, and pricing functions  $P_j(G_j)$  that satisfy (2.8) and (2.9).

To impose more structure on the model we need to make additional assumptions on preferences and probability distributions. The most common assumptions in applying the valuation models discussed earlier are constant relative risk aversion (CRRA) utility functions and lognormally distributed random variables. So we assume that all agents have the same CRRA utility function  $U_i(C_i) = C_i^{1-\alpha}/(1-\alpha)$  where  $\alpha$  is the coefficient of relative risk aversion, and that all of the  $C_i$  and  $Z_j$  are joint lognormally distributed.<sup>12</sup> Using these assumptions and results from the appendix of Rubinstein ([43]) (reproduced in appendix B), (2.8) can be solved for  $P_j(G_j)$  and expressed as:

$$P_j(G_j) = G_j N\left(\frac{g_j - \mu_j + \alpha \sigma_{ij}}{\sigma_j}\right) - e^{\mu_j + 0.5\sigma_j^2 - \alpha \sigma_{ij}} N\left(\frac{g_j - \mu_j - \sigma_j^2 + \alpha \sigma_{ij}}{\sigma_j}\right) \quad (2.10)$$

where  $\sigma_{ij} = Cov(c_i, z_j)$  for  $c_i = \ln(C_i)$ . In equilibrium, (2.10) must hold for every  $i$  because all agents are free to trade insurance contracts. Therefore, if all agents know the joint probability distributions of each  $z_j$  and their own  $c_i$ , then  $\sigma_{ij} = \sigma_j^c$  must be equal for all  $i$ . Then (2.10) can be expressed as:

$$P_j(G_j) = G_j N\left(\frac{g_j - \mu_j + \alpha \sigma_j^c}{\sigma_j}\right) - e^{\mu_j + 0.5\sigma_j^2 - \alpha \sigma_j^c} N\left(\frac{g_j - \mu_j - \sigma_j^2 + \alpha \sigma_j^c}{\sigma_j}\right). \quad (2.11)$$

It is interesting to note that if agents are risk neutral ( $\alpha = 0$ ), or the risk from holding insurance contracts is completely diversifiable via the mutual fund ( $\sigma_j^c = 0$ ), then the market risk is valued at zero and (2.11) simplifies to the present value formula under lognormality

---

<sup>12</sup>This assumption requires agents to have homogeneous preferences but still allows them to have different endowments, face different risks, and consume different amounts.



(2.2).<sup>13</sup> But if agents are risk averse and the insurance risk is not completely diversifiable, then (2.11) gives a generalized valuation formula that incorporates a value for the non-diversifiable market risk embodied in the insurance contracts. It is also interesting to note that for  $\beta = 1$  (premiums and indemnities paid at the same time) then (2.11) is exactly the same as the arbitrage based option pricing model (2.4) for options on non-tradable indexes, as long as  $\lambda\sigma_j = \alpha\sigma_j^c$ . This shows that the arbitrage model (2.4) and the equilibrium based model (2.11) are identical, except for the way in which they define and characterize the market risk term embodied in the valuation formula.

For (2.11) to be a useful formula we need a way of characterizing the magnitude of the market risk factor  $\alpha\sigma_j^c$ . A useful property of the lognormal distribution is that if  $c_i$  and  $z_j$  are joint normal then  $Cov(C_i, z_j) = E(C_i)Cov(c_i, z_j)$ . Using this property, and assuming aggregate incomes  $Y$  and  $W$  are lognormally distributed, we can apply the covariance operator with respect to  $z_j$  to the aggregate budget constraint (2.9) to get:

$$E(C)\sigma_j^c = E(Y)Cov[\ln(Y), z_j] + E(W)Cov[\ln(W), z_j]. \quad (2.12)$$

A property of the lognormal distribution shown in appendix C is that  $Cov[\ln(Y), z_j] \approx \rho_{YZ_j}CV_YCV_{Z_j}$  and  $Cov[\ln(W), z_j] \approx \rho_{WZ_j}CV_WCV_{Z_j}$  where  $\rho_{YZ_j}$  and  $\rho_{WZ_j}$  are correlation coefficients between aggregate farm income and the insurance index, and between aggregate non-farm income and the insurance index, respectively; and  $CV_Y$ ,  $CV_W$ , and  $CV_{Z_j}$  are coefficients of variation for aggregate farm income, aggregate non-farm income, and the insurance index, respectively. Substituting these expressions into (2.12) and solving for  $\sigma_j^c$  gives:

$$\sigma_j^c \approx S_Y\rho_{YZ_j}CV_YCV_{Z_j} + S_W\rho_{WZ_j}CV_WCV_{Z_j} \quad (2.13)$$

where  $S_Y = E(Y)/E(C)$  and  $S_W = E(W)/E(C)$  are the expected shares of aggregate farm and non-farm income in aggregate consumption.

---

<sup>13</sup>Of course,  $\beta = 1$  in (2.11) because we assume premiums and indemnities are paid at the same time.

### 2.4.2 Discussion

Equation (2.13) provides a way of characterizing and understanding the way in which market risk influences the valuation formula. In economies like the U.S., where the share of farm income in total consumption is small, and the correlation between agricultural insurance indices and non-farm incomes is low, then both terms in (2.13) will be small and the market risk associated with agricultural insurance will be correspondingly small, irrespective of the degree of risk aversion of agents. In this case, the equilibrium price of insurance contracts will be close to actuarially fair and the present value formula (2.2) or, more generally (2.1), may be quite appropriate. The reason this occurs is that the agricultural insurance risk is easily diversified by spreading it across a large non-agricultural sector with incomes that are largely uncorrelated with changes in insurance index values.

Alternatively, in developing economies (or developed economies that have a much higher share of farm income in total consumption) then the market risk may be much higher if aggregate farm income is highly variable and strongly correlated with the insurance index. We would generally expect aggregate farm income to be both variable and highly correlated with the insurance index because otherwise the insurance contract would be of little use to farmers. Of course, the other way in which the market risk can be high is if there is a strong correlation between non-farm incomes and the insurance index, which seems less likely to occur in general.

The overall conclusion is that the value of the market risk factor in particular situations and for particular insurance contracts is an empirical question and indiscriminant use of the present value model would be unwise. Equation (2.13) provides a simple but useful way of thinking about the factors contributing to the magnitude of the market risk. When combined with empirical estimates and an assumption (or estimate) of the CRRA parameter  $\alpha$ , equation (2.13) also may lead to a reasonable estimate of the value of market risk for particular agricultural insurance contracts in particular situations.

### 2.4.3 Insurance Industry Model

A potential criticism of the mutual fund model is that the institutional structure of actual agricultural insurance industries does not conform to the mutual fund structure. Agricultural insurance contracts can only be purchased by farmers and are only sold by insurance companies (or government agencies). That is, there is no free and unregulated direct trade (allowing both short and long selling) of agricultural insurance contracts among farmers and wage earners. In this section we develop an insurance industry model that embodies some of these key institutional features of actual insurance markets and show that the equilibrium (and hence the insurance valuation formula) remains identical to that obtained in the mutual fund model.

The setup and notation remain as in the mutual fund model except that now only farmers can buy agricultural insurance, and each farmer can only buy contracts based on his or her relevant regional revenue index. The contracts are sold by an insurance industry, which might be a set of competitive firms or it might be a government agency. The insurance industry acts as a broker by selling contracts to farmers and then repackaging and securitizing the contracts as an asset that is then resold on financial markets. Under this setup the farmer budget constraints become:

$$C_i \leq Y_i + W_i + X_{ik}[V_k(G_k) - P_k(G_k)] + \sum_{j=1}^m (R_j - 1)A_{ij} \quad \forall i = 1, 2, \dots, n_Y \quad (2.14)$$

where the  $k$  subscript on the insurance contract indicates the revenue index for the region where farmer  $i$  is located,  $A_{ij}$  is the amount of the repackaged asset based on insurance contract  $j$  that is purchased (sold if negative) by farmer  $i$ , and  $R_j$  is the gross rate of return on the repackaged asset based on insurance contract  $j$ . Similarly, the wage earners budget constraint becomes:

$$C_i \leq W_i + \sum_{j=1}^m (R_j - 1)A_{ij} \quad \forall i = n_Y + 1, n_Y + 2, \dots, n. \quad (2.15)$$

Notice that only farmers can buy insurance, and only on the revenue index for the region in which they are located. We also impose the constraint  $X_{ik} \geq 0$  (farmers can only buy not

sell insurance). However we allow all agents to buy or sell any perfectly divisible amount of the repackaged asset offered by the insurance brokerage industry.

Because the insurance industry only acts as a broker and therefore bears no risk, a natural equilibrium condition is to set insurance industry profits on each insurance contract to zero which implies:

$$P_j(G_j) \sum_{i=1}^n X_{ij} = - \sum_{i=1}^n A_{ij} \quad \forall j = 1, 2, \dots, m, \text{ and} \quad (2.16)$$

$$-V_j(G_j) \sum_{i=1}^n X_{ij} = R_j \sum_{i=1}^n A_{ij} \quad \forall j = 1, 2, \dots, m. \quad (2.17)$$

Equation (2.16) shows that in the aggregate agents will be net short each repackaged asset (the insurance industry collects premiums and uses the proceeds to buy the repackaged asset from individuals, who are selling short to the industry in aggregate). Equation (2.17) shows that when the indemnities are paid by the industry they are financed by selling the industry's net long position in the repackaged asset, while accounting for the equilibrium return on investment,  $R_j$ . Together, (2.16) and (2.17) imply:

$$R_j = V_j(G_j)/P_j(G_j). \quad (2.18)$$

Now we turn to the necessary conditions for individual choice of investment in the repackaged asset. Because all individuals can trade the asset, and short sales are allowed, the necessary conditions for expected utility maximization include:

$$E\{U'_i(C_i)[R_j - 1]\} = 0 \quad \forall i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m. \quad (2.19)$$

But substituting (2.18) into (2.19) and multiplying through by  $P_j(G_j)$  gives exactly the same necessary condition (2.8) as in the mutual fund model. Furthermore, aggregating the budget constraints (2.12) and (2.14), and imposing the insurance industry equilibrium conditions (2.16) and (2.17), the aggregate budget constraint for the industry equilibrium model simplifies to exactly the same aggregate budget constraint (2.9) as in the mutual fund model. Together, these two results show that equilibrium in the mutual fund model and the

insurance industry model are formally equivalent and so, under the usual assumptions of CRRA preferences and lognormally distributed random variables, the mutual fund insurance valuation formulas (2.11) and (2.13) also apply in the case of the insurance industry model. Of course, the reason this occurs is because even though risks cannot be diversified via trade in the insurance contracts themselves, competitive trade in the repackaged asset allows the optimal amount of insurance risk diversification and spreading to occur.

## 2.5 Simulation Results

It will be useful to examine just how different the results from alternative insurance pricing models can be under various conditions. Here we address this issue using a simulation model. We use the present value model under lognormality (2.2) as a baseline and set  $\beta = 1$  (premiums paid at the same time as indemnities) and  $\sigma_j = 0.3$ .<sup>14</sup> We also set  $G_j = e^{\mu_j}$ , indicating that the guarantee level is set equal to the median value of  $Z_j$ . We then find the value of  $\mu_j$  that makes  $P(G_j) = 1$  using the present value formula (2.2). In this way, the present value baseline has been normalized to give an insurance value of one, which makes it easy to compare the relative effects of using alternative valuation formulas. Values and ranges for all parameters used in the simulation are provided in table 2.1.

It is conceptually difficult to apply the Black-Scholes formula (2.3) to price agricultural insurance because it relies on the assumption that  $Z_j$  is the continuously observed price of a tradable asset that follows a geometric Brownian motion. When applying the model to a regional revenue index which is observed discretely, the assumption that seems to make the most sense is to think of the underlying continuous variable as the expectation of  $Z_j$  conditional on information available at different points in time. Then we can think

---

<sup>14</sup>It is well known that if  $Z_j$  is lognormal then  $\sigma_j$  is approximately equal to its coefficient of variation. Hence,  $\sigma_j$  is a relative volatility measure and setting  $\sigma_j = 0.3$  indicates an approximate 30% coefficient of variation for the uncertainty surrounding the value of the realized regional revenue index when indemnities are paid, given information available when the insurance is taken out. The 30% figure is arbitrary but seems a reasonable starting point for regional crop revenue indexes.

Table 2.1. Summary of the parameters in simulation model

Parameter	Notation	Base value	Range
$\alpha$	CRRA parameter	2	[0, 3]
$r$	Risk-free interest rate	.05	[0, .15]
$\mu_j$	Mean of the logarithm of insurance index $j$	2.3	
$\sigma_j$	Standard deviation of the logarithm of insurance index $j$	.3	
$S_Y$	Expected share of farm income in aggregate consumption	.2	[.05, .70]
$\rho_{YZ_j}$	Correlation coefficient between aggregate farm income and the insurance index $j$	.7	[.5, .9]
$\rho_{WZ_j}$	Correlation coefficient between aggregate non-farm income and the insurance index $j$	.1	[0, .2]
$CV_{Z_j}$	Coefficient of variation for insurance index $j$	.3	
$CV_Y$	Coefficient of variation for aggregate farm income	.2	[0, .4]
$CV_W$	Coefficient of variation for aggregate non-farm income	.2	[0, .4]
$\alpha\sigma_j^c$	Market risk factor	.026	[0, .166]

of the conditional expectation  $E_t(Z_j)$  changing continuously over time as new information becomes available about the realization of  $Z_j$ . Under this interpretation it would be natural to set  $Z_j^0 = E_0(Z_j)$  where  $t = 0$  indicates the time period when insurance is taken out. By the properties of the lognormal distribution this would imply:

$$\ln(Z_j^0) = z_j^0 = \mu_j + 0.5\sigma_j^2. \quad (2.20)$$

In simulating the Black-Scholes formula (2.3) we use the same values of  $g_j$ ,  $\mu_j$ , and  $\sigma_j$  that were used to compute the present value formula, and then use (2.20) to obtain  $z_j^0$  and  $Z_j^0$ . The only remaining unknown in the Black-Scholes formula is the risk-free rate of interest  $r$ . We simulated the Black-Scholes insurance price under different interest rates ranging from 0% to 15% (  $r = 0$  to  $r = 0.15$ ).<sup>15</sup>

We also simulate our alternative insurance pricing formula (2.11) and (2.13) under a range of assumptions. We begin by setting  $g_j$ ,  $\mu_j$ , and  $\sigma_j$  to the same values that were used to compute the present value price (see table 2.1). Then we examine a range of alternative values for the market risk factor  $\alpha\sigma_j^c$  and apply (2.11). We let the coefficient of relative risk aversion  $\alpha$  range between 0 and 3. Then using (2.13) we let the share of farm income in total aggregate consumption  $S_Y$  range from 5% (highly developed economies) to 70%

<sup>15</sup>To keep the valuation consistent with the present value calculation we assume the premium is agreed to at the time the insurance is taken out but not paid until the index has been realized and indemnities paid. This requires multiplying (2.3) through by  $1/\beta = 1 + r$ .

(developing economies) and set  $S_W = 1 - S_Y$ . The correlation between aggregate farm income and the insurance index is assumed to range from 0.5 to 0.9, and the correlation between aggregate non-farm income and the insurance index is assumed to range from 0 to 0.2. The coefficient of variation for the insurance index is set at 30%, consistent with the earlier assumption that  $\sigma_j = 0.3$ , and the coefficients of variation for farm and non-farm incomes are assumed to range between 0 and 40%. With these assumptions, the possible values for  $\alpha\sigma_j^c$  range from 0 to approximately 0.166 (see table 2.1).

The first simulation sets all parameters except the interest rate to their base values and lets the interest rate range from 0% to 15%. Insurance values from the present value model, the Black-Scholes model, and our market risk adjusted model are then graphed over alternative interest rates (see figure 2.5). These simulations highlight three important points. First, as long as premiums are not paid until revenue has been realized and indemnities paid, then the interest rate has no effect on the present value model insurance price, or on the market risk adjusted price. Second, our market risk adjusted price adds a premium of approximately 11% to the risk-neutral present value price at base parameter values. Third, when the interest rate is zero the Black-Scholes value equals the present value model, but as the interest rate rises the Black-Scholes price falls below the present value price (negative risk premium). This happens because as long as we continue to assume  $Z_j^0 = E_0(Z_j)$  then a higher interest rate implies the insurance premium must continue to fall in order equate returns from the insurance and the dynamically adjusted portfolio containing  $Z_j$  and a risk free bond. This is a nonsensical result and simply emphasizes that the Black-Scholes model is entirely inappropriate for valuing claims on non-traded assets.

The second simulation keeps the interest rate at its base value of 5% and allows all other parameters to vary over their entire range, which implies the market risk factor  $\alpha\sigma_j^c$  ranges from 0 to approximately 0.166. Results are shown in figure 2.5. In this case, the Black-Scholes price is 18% lower than the present value price, again indicating problems with the Black-Scholes approach. The market risk adjusted price equals the present value price when

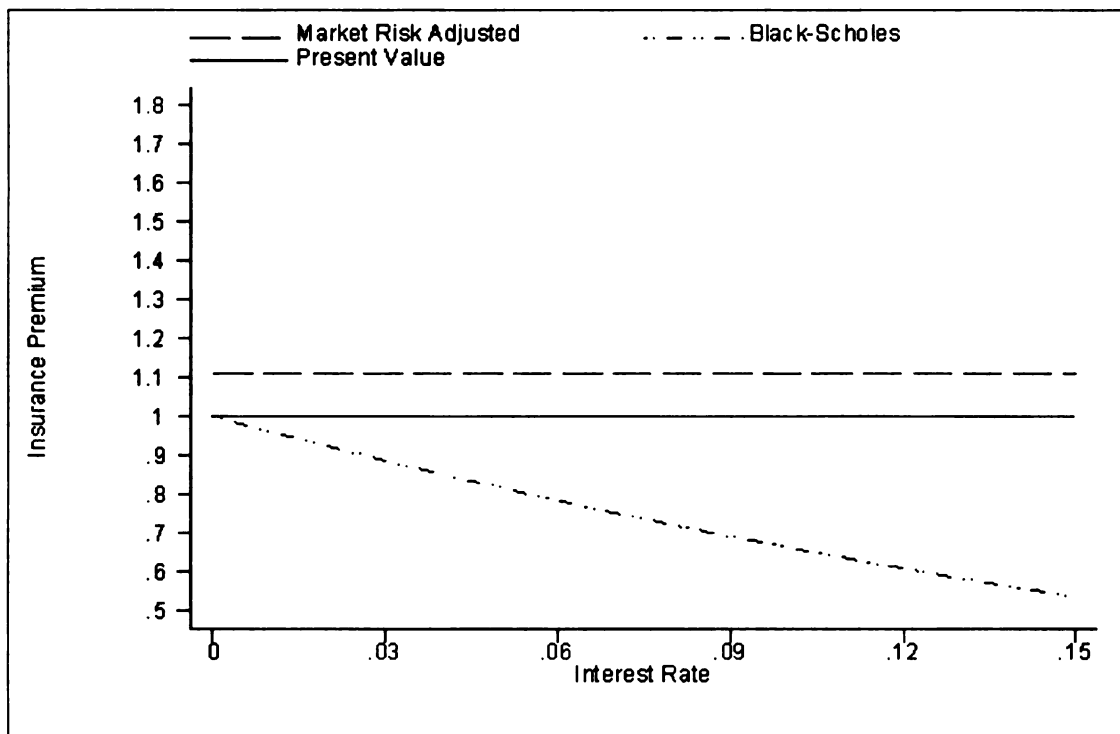


Figure 2.1. Insurance premiums versus interest rate

there is no market risk but may range up to 78% higher than the present value price under extreme parameter values (see figure 2.5). This shows that market risk can have nontrivial effects on insurance valuation, and the differential will be higher the more risky is farm income, the more important agriculture is in the overall economy, the more risk averse are agents, and the greater the correlation is between insurance indices and farm incomes.

For the third and final simulation we keep all parameters except the share of farm income in aggregate consumption constant at their base values and allow the farm income share to vary over its range from 5% to 70%. Results in figure 2.5 highlight that the relative importance of agriculture and farming in the economy can have a sizable effect on insurance valuation in the market risk adjusted model. While the risk premium is only around 5% in economies where agriculture and farming are relatively unimportant, this proportion rises to around 30% in economies that are heavily dependent on agriculture.



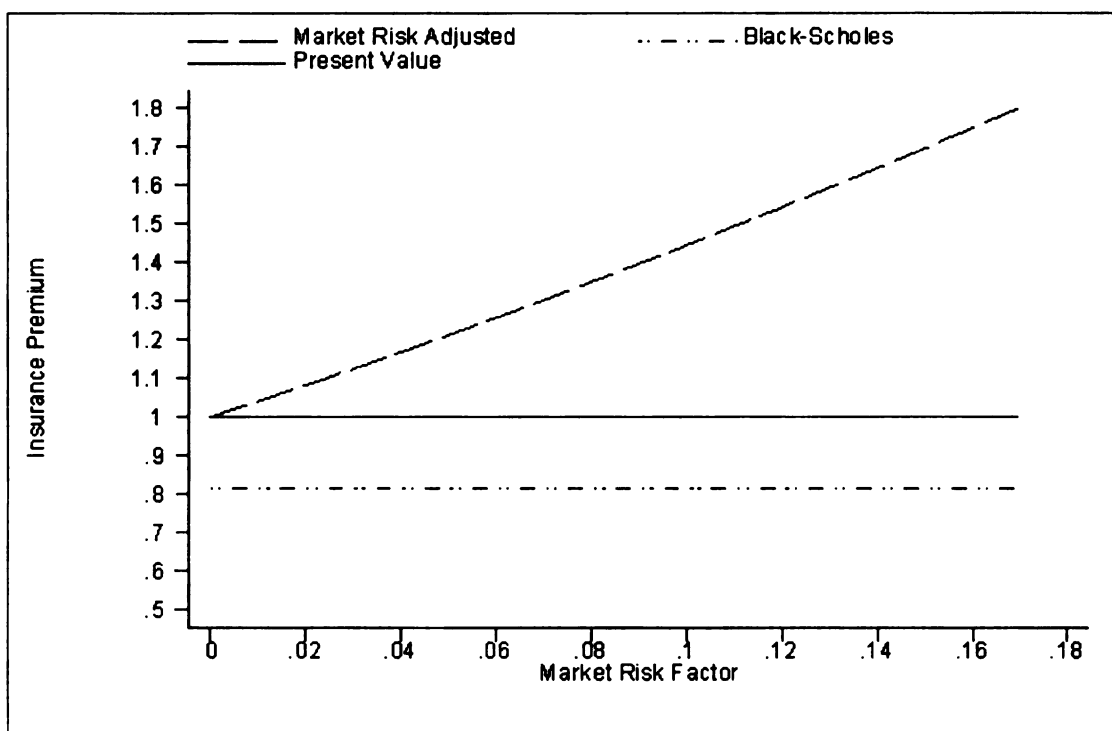


Figure 2.2. Insurance premiums versus market risk factor

## 2.6 Conclusions

This paper evaluates existing methods for valuing agricultural insurance contracts and then develops a new valuation model that offers some intuitive insights into the factors that determine the value of market risk. Under the assumptions of CRRA preferences and lognormal probability distributions, the new model also provides a simple, practical way of estimating the impact of market risk for particular insurance contracts in particular situations and environments.

Simulation results suggest that simple present value models of insurance pricing, which assume no market risk, may be quite appropriate in large developed economies that can spread and diversify agricultural risks over a large non-agricultural sector either through capital markets in the case of an insurance industry equilibrium, or through the tax system in the case of government insurance programs. In economies where the share of farm income

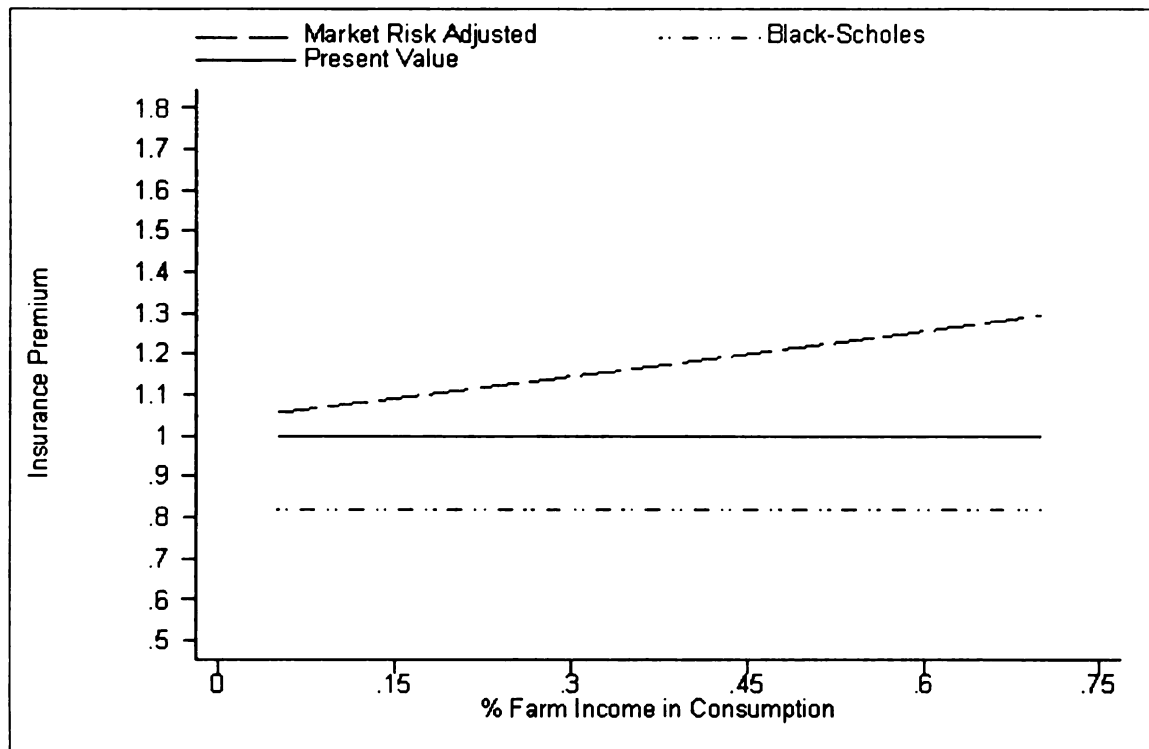


Figure 2.3. Insurance premiums versus share of farm income in aggregate consumption

in total consumption is much higher, however, the market risk may also be higher and have a significant impact on insurance valuation.

Overall, our analysis suggests that present value models may be a reasonable approximation in some circumstances but, as is now becoming well-known, the standard Black-Scholes model has a number of deficiencies for pricing agricultural insurance. Other methods for characterizing the market price of risk in agricultural insurance contracts, such as arbitrage-based pricing model for options written on non-tradable assets, the Lucas representative agent model, and the Chambers state-contingent cost function approach are logically consistent and potentially useful. However, the simple heterogeneous agent equilibrium model developed here is easy to use, amenable to empirical estimation, and provides some useful insights into the factors determining the effect of market risk on agricultural insurance valuation.

# CHAPTER 3

## MODEL SELECTION IN

## STOCHASTIC FRONTIER

## ANALYSIS: MAIZE PRODUCTION

## IN KENYA

### 3.1 Introduction

Stochastic production frontier analysis has been widely used to study technical inefficiency in various settings since its introduction by Aigner, Lovell, and Schmidt ([2]), and Meeusen and van den Broeck ([31]). The approach has two components: a stochastic production frontier serving as a benchmark against which firm inefficiency is measured, and a one-sided error term which captures technical inefficiency and is usually assumed to be identically and independently distributed across firms. Recent studies have generalized the one-sided error term to allow its distribution to be heterogeneous and depend on various firm characteristics (see [9]; [14]; [54]; [55]).

Allowing inefficiency to depend on firm characteristics allows researchers to identify the

determinants of inefficiency, and to suggest interventions which might improve efficiency. However, this approach has been hampered by two problems. First, the relationship between firm characteristics and the extent of technical inefficiency can be sensitive to the model used to incorporate firm characteristics and little is known about how to choose among competing models. This model uncertainty makes policy recommendations quite tenuous. Second, existing studies have mostly focused on the *directions* of the influence of the firm characteristics on technical inefficiency while generally overlooking the *magnitudes* of the partial effects. This makes it difficult to determine which type of policy intervention will have the largest impact on inefficiency. The second problem is somewhat surprising given that the magnitudes of the effects of explanatory variables on dependent variables are often the focal point in other types of regression analyses.

In this paper, we make three contributions to the stochastic frontier literature. First, we provide a method for estimating the quantitative magnitude of the partial effects of exogenous firm characteristics on firm inefficiency, show how to put standard errors around these partial effects, and propose an  $R^2$ -type measure to summarize the overall explanatory power of the exogenous factors on inefficiency. Second, we examine maize production in Kenya and show that while alternative models of the relationship between household characteristics and technical inefficiency tend to provide the same direction of the influence of household characteristics, the magnitudes of the partial effects on firm inefficiency are quite sensitive to model selection. Third, we show how a recently developed model selection procedure (Alvarez, Amsler, Orea, and Schmidt, [4], hereafter AAOS) can be used to choose among competing models, and we use bootstrapping to provide evidence on the power of this procedure. The model selection procedure gives an unambiguous choice of best model for our application to Kenyan maize production. This is important because if different models give different results, and we cannot distinguish statistically among the models, we do not know which set of results should be used. But if we can pick a clearly best model, then the fact that inferences and conclusions are sensitive to model selection is not such a

problem.

Our empirical analysis is on maize production in Kenya. The problem of hunger in Kenya remains widespread and its economy depends heavily on agriculture with 75 percent of Kenyans making their living from farming. Maize is the primary staple food and most farmers are engaged in maize production. In recent years, total maize output has not kept pace with the growing population and demand, largely due to falling land productivity: average national maize yields have fallen from over 2 tons per hectare in the early 1980's to about 1.6 tons per hectare recently ([36]). The technical efficiency level of Kenya maize production is therefore an important economic and policy issue.

The empirical analysis uses detailed household survey data from Kenya and investigates determinants of productivity and inefficiency using stochastic frontier analysis. The variables used to explain inefficiency are related to the educational background of the household, rural infrastructure, land tenure, credit constraints faced by the household, and farm size. These explanatory factors go well beyond those used to study production inefficiency in most other studies of agriculture (see [27]; [21]; [4]).

In the remainder of the paper, we first review the standard stochastic frontier production model commonly used in the literature and then extend it to allow estimation of the magnitude of firm characteristic effects on technical inefficiency, put standard errors around these estimates, and develop an  $R^2$ -type measure of overall explanatory power. Next we describe our data and variables used in the empirical analysis, followed by estimation results from alternative model specifications. Results for the magnitude of partial effects of household characteristics on inefficiency are quite sensitive to model specification. Then we carry out specification tests to choose our final model. A novel approach is that we use the bootstrap to examine the reliability of these specification tests in choosing the correct model. The final sections contain an analysis of technical inefficiency in maize production in Kenya based on the final model, and some concluding comments.

## 3.2 Stochastic Production Frontier Models

The basic setup and notation follow Wang and Schmidt (2002) and AAOS. Firms are indexed by  $i = 1, \dots, N$ . Let  $y_i$  be log output;  $x_i$  be a vector of inputs; and  $z_i$  be a vector of exogenous variables that exert influence on farm inefficiency. Let  $y_i^*$  be the unobserved frontier which is modeled as

$$y_i^* = x_i' \beta + v_i, \quad (3.1)$$

where  $v_i$  is distributed as  $N(0, \sigma_v^2)$  and is independent of  $x_i$  and  $z_i$ , and  $\beta$  is a parameter vector. The actual log output level  $y_i$  equals  $y_i^*$  less a one-sided error,  $u_i$ , whose distribution depends on  $z_i$ . The full model is written as

$$y_i = x_i' \beta + v_i - u_i(z_i, \theta), \quad u_i(z_i, \theta) \geq 0, \quad (3.2)$$

where  $\theta$  is a vector of parameters. It is assumed that  $u_i$  and  $v_i$  are independent of one another and that  $u_i$  is independent of  $x_i$  (conditional on  $z_i$ ). The model is usually implemented by assuming  $u_i$  is distributed as  $N(\mu_i, \sigma_i^2)^+$  with various specifications (discussed below) used to model  $\mu_i$  and  $\sigma_i$ . The frontier function and the inefficiency part are generally estimated in one step using maximum likelihood estimation (MLE) to achieve both efficiency and consistency.<sup>1</sup>

Indexing exogenous factors with  $k = 1, \dots, K$ , we take expectations conditional on  $x_i$  and  $z_i$ , and then take partial derivatives with respect to  $z_{ik}$  on both sides of equation (3.2), to get

$$\partial[E(y_i|x_i, z_i)]/\partial z_{ik} = \partial[E(-u_i|x_i, z_i)]/\partial z_{ik}. \quad (3.3)$$

Here,  $\partial[E(-u_i|x_i, z_i)]/\partial z_{ik}$  can be interpreted as the partial effect of  $z_{ik}$  on efficiency  $-u_i$ , and is also the partial effect on  $y_i$ . Because  $y_i$  is log output,  $\partial[E(-u_i|x_i, z_i)]/\partial z_{ik}$  is the semi-

---

<sup>1</sup>Some studies use a two-step procedure where the frontier function is estimated first, and then the inefficiency term is regressed on exogenous variables in the second step. This procedure is biased for two reasons. The first and more obvious reason is the possible correlation between the input variables in the frontier function and the variables in the inefficiency term. The second reason is that the inefficiency term from the first step is measured with error and the error is correlated with the exogenous factors. See Wang and Schmidt ([56]) for an extensive discussion and evidence from Monte Carlo experiments.

elasticity of output (efficiency) with respect to the exogenous factors, i.e., the percentage change in expected output when  $z_{ik}$  increases by one unit. Similarly, taking conditional variances we have

$$\partial[V(y_i|x_i, z_i)]/\partial z_{ik} = \partial[V(u_i|x_i, z_i)]/\partial z_{ik}. \quad (3.4)$$

So  $\partial[V(u_i|x_i, z_i)]/\partial z_{ik}$  is the partial effect of  $z_{ik}$  on the variance of both the inefficiency term  $u_i$  and  $y_i$ . It can be interpreted as an estimator of the partial effect of  $z_{ik}$  on production uncertainty.

The measures  $\partial[E(u_i|x_i, z_i)]/\partial z_{ik}$  and  $\partial[V(u_i|x_i, z_i)]/\partial z_{ik}$  were first proposed and used in Wang (2002) and Wang (2003), but for a different purpose than they are used here.<sup>2</sup> Here, we interpret  $\partial[E(-u_i|x_i, z_i)]/\partial z_{ik}$  as the semi-elasticity of output with respect to exogenous factors so that not only its sign but also its magnitude is of economic interest. In appendix D we provide formulas for computing estimates of  $\partial[E(-u_i|x_i, z_i)]/\partial z_{ik}$  and  $\partial[V(u_i|x_i, z_i)]/\partial z_{ik}$ , along with their standard errors using the delta method for several popular model specifications for  $\mu_i$  and  $\sigma_i$ .

It will often be useful to measure how well the vector of exogenous factors,  $z$ , explains inefficiency,  $u$  in a data sample. Surprisingly, this has not been addressed in the previous literature. We suggest a statistic  $R_z^2$ , to summarize the explanatory power of  $z$ . To motivate the measure, the variance of the inefficiency term  $u_i$  can be decomposed as

$$V(u_i) = V_z[E(u_i|z_i)] + E_z[V(u_i|z_i)], \quad (3.5)$$

where  $V_z[E(u_i|z_i)]$  is the variance of the conditional mean function over the distribution of  $z_i$ , and  $E_z[V(u_i|z_i)]$  is the expected variance around the conditional mean of  $u_i$ . The fraction of variation in  $u_i$  that is explained by  $z_i$  is  $V_z[E(u_i|z_i)]/V(u_i)$ . Thus a natural measure of explanatory power over the sample would be

$$R_z^2 = \frac{\sum_{i=1}^n \left[ \hat{E}(u_i|z_i) - \frac{1}{n} \sum_{i=1}^n \hat{E}(u_i|z_i) \right]^2}{\sum_{i=1}^n \left[ \hat{E}(u_i|z_i) - \frac{1}{n} \sum_{i=1}^n \hat{E}(u_i|z_i) \right]^2 + \sum_{i=1}^n \hat{V}(u_i|z_i)}, \quad (3.6)$$

---

<sup>2</sup>Wang uses the two measures to show the property of monotonic/non-monotonic efficiency effect associated with alternative model specifications.

where  $\hat{E}$  and  $\hat{V}$  indicate sample estimates of the mean and variance of  $u_i$  conditional on  $z_i$ . Letting  $R_1 = \mu_i/\sigma_i$ ,  $R_2 = \phi(R_1)/[\Phi(R_1)]^{-1}$ , and  $R_3 = -R_2^2 - R_1 R_2$ , where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the density and cumulative density functions for the standard normal, then the mean and variance of  $u_i$  conditional on  $z_i$  can be expressed as

$$E(u_i|x_i, z_i) = \sigma_i \cdot (R_1 + R_2) \quad (3.7)$$

$$V(u_i|x_i, z_i) = \sigma_i^2 \cdot (1 + R_3). \quad (3.8)$$

So all that remains to compute  $R_z^2$  is to estimate  $\hat{\mu}_i$  and  $\hat{\sigma}_i$  for a specific model specification (see the model specification section below) and then substitute these estimators for the population values  $\mu_i$  and  $\sigma_i$  in (3.7) and (3.8) to get the sample estimates of  $\hat{E}(u_i|z_i)$  and  $\hat{V}(u_i|z_i)$ .

Similar to  $R^2$  in an ordinary least squares regression,  $R_z^2$  can be called the “goodness of fit” of the efficiency component, and it can be interpreted as the fraction of the sample variation in  $u$  that is explained by  $z$ .

### 3.2.1 Alternative Model Specifications

In the original specification of stochastic frontier functions, Aigner, Lovell, and Schmidt ([2]) and Meeusen and van den Broeck ([31]) assumed an identical and independent half-normal distribution for the one-sided error terms  $u_i$ . Subsequent studies have generalized the model to allow for heterogeneity in the distribution of the inefficiency term while maintaining the assumption that each  $u_i$  is half normal. Kumbhakar, Ghosh, and McGuckin ([27]), Huang and Liu ([21]), and Battese and Coelli ([9]) allow the mean of the pre-truncated normal distribution of  $u_i$  to depend on a set of exogenous factors. Reifschneider and Stevenson (1991), Caudill and Ford ([13]), Caudill et al. ([14]) and Hadri ([20]) allow the variance of the pre-truncated normal distribution of  $u_i$  to depend on the exogenous factors. Wang ([55]) allows both the mean and the variance of the pre-truncated distribution of  $u_i$  to depend on exogenous factors.



Regardless of whether we allow the mean, the variance, or both the mean and the variance of the *pre-truncated* normal to depend on exogenous factors, both the mean and the variance of the *truncated* half normal will always depend on the exogenous factors. These are sometimes called models of heteroscedasticity, but the fact that the mean also changes makes this terminology potentially misleading. Whereas heteroscedasticity affects only the efficiency of estimation in a standard linear model, in a stochastic frontier model with heterogeneity in the distribution of the inefficiency term, failure to model the exogenous factors appropriately leads to biased estimation of the production frontier model, and of the level of technical inefficiency, hence leading to poor policy conclusions (see [13]; [14]; [20]; [55]).

With different specifications available to model heterogeneity, it is unclear which should be used in particular applied settings. The choices made in many past studies seem to be somewhat arbitrary. However, a carefully specified model might help to increase estimation efficiency and remove sources of potential bias and inconsistency ([55]). Moreover, there has been little investigation on how the choice of model specification influences the estimation results. In order to deal with the model specification problem, researchers usually do sensitivity analysis using competing models. But if the competing models give very different results, it is difficult to pick one and discard the others. Wang ([55]) treats this problem by specifying a flexible model that nests most of the usual model specifications for  $\mu_i$  and  $\sigma_i$ . However, a more flexible model has more parameters, which impose a higher computational burden and reduces degrees of freedom. Given that large samples are typically difficult to obtain in stochastic frontier models, some relevant parameters may be estimated imprecisely in flexible model specifications.

AAOS suggest a procedure for selecting a model for the one-sided error term. First, assume the general model of inefficiency ([55]) in which  $u_i$  is distributed as  $N(\mu_i, \sigma_i^2)^+$ , with  $\mu_i = \mu \cdot \exp(z_i' \delta)$  and  $\sigma_i = \sigma_u \cdot \exp(z_i' \gamma)$ . This general model nests several simpler models, many of which have been used in previous studies. In particular, the following six models are special cases of the general model, as outline in AAOS.

1. Scaled Stevenson model: Let  $\delta = \gamma$ . Then the distribution of  $u_i$  becomes  $\exp(z_i'\delta) \cdot N(\mu, \sigma_u^2)^+$ , which is used in Wang and Schmidt ([54]).
2. KGMHLBC model: Let  $\gamma = 0$ . Then the distribution of  $u_i$  becomes  $N(\mu \cdot \exp(z_i'\delta), \sigma_u^2)^+$ , which has been considered in Kumbhakar, Ghosh, and McGuckin ([27]), Huang and Liu ([21]), and Battese and Coelli ([9]).
3. RSCFG- $\mu$  model: Let  $\delta = 0$ . Then the distribution of  $u_i$  becomes  $N(\mu, \sigma_u^2 \cdot \exp(2z_i'\gamma))^+$ .
4. RSCFG model: Let  $\mu = 0$ . Then the distribution of  $u_i$  becomes  $\exp(z_i'\gamma) \cdot N(0, \sigma_u^2)^+$ , which is considered in Reifschneider and Stevenson ([40]), Caudill and Ford ([13]), and Caudill, Ford, and Gropper ([14]).
5. Stevenson model: Let  $\delta = \gamma = 0$ . Then the distribution of  $u_i$  becomes  $N(\mu, \sigma_u^2)^+$ , which is the model of Stevenson ([47]).
6. ALS model: Let  $\mu = \gamma = 0$ . Then the distribution of  $u_i$  becomes  $N(0, \sigma_u^2)^+$ , which is the model of Aigner, Lovell, and Schmidt ([2]).

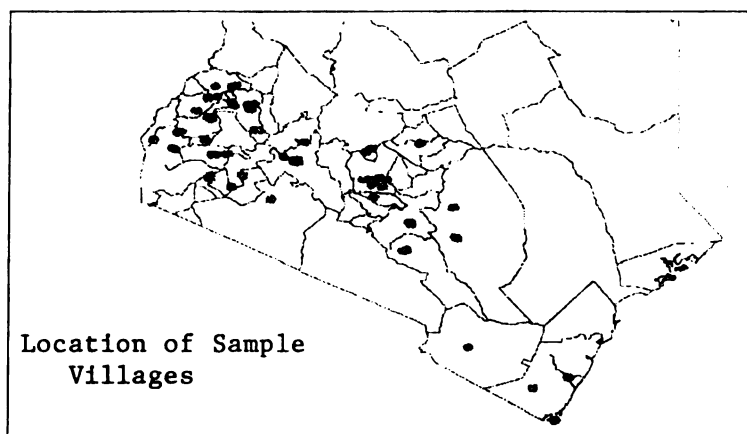
Among the six models, the scaled Stevenson, KGMHLBC and RSCFG- $\mu$  models have the same number of parameters. The RSCFG model is nested by the scaled Stevenson model and the RSCFG- $\mu$  model. Also notice that the Stevenson model and the ALS model do not contain any variables ( $z_i$ ) that influence the distribution of inefficiency. AAOS show how to use likelihood ratio (LR) tests, LM tests and Wald tests to test the above restrictions, and hence to choose a plausible model for inefficiency.

### 3.3 Empirical Application

The empirical application is to maize production in Kenya using detailed household survey data.

### 3.3.1 Data

The data are from a rural household survey of about 1100 households planting maize in the main season of 2003-2004 in Kenya.<sup>3</sup> The survey was designed and implemented under the Tegemeo Agricultural Monitoring and Policy Analysis Project, a collaboration among Tegemeo Institute of Egerton University, Michigan State University, and the Kenya Agricultural Research Institute. Figure 3.3.1 is a map of Kenya with the round dots representing sampled villages. These villages were chosen randomly from each of eight pre-determined agro-economic zones and then households were sampled randomly from each selected village.



Source: Suri Tavneet (2005).

Figure 3.1. Location of sample villages in Kenya

Field level data are available for each sampled household and some households planted maize in more than one field. The survey includes not only detailed field production information but also rich demographic and infrastructure characteristics of each household. The production data for each field include size of the field, yield, labor input, fertilizer application, and seed usage. The demographic information includes the age, gender and education level of each household member; how far a household is from a bus stop, a motorable road, a telephone booth, mobile phone service, and extension service; whether a household mem-

---

<sup>3</sup>See Suri ([50]) for a study of the adoption decisions of hybrid seed by maize producers in Kenya using the same data set.

ber has non-farm income; whether a household receives loans; how much land a household owns, and land tenure. Rainfall and soil quality data are also available at the village level.

### 3.3.2 Variables in the Production Frontier

In the production frontier part of the model, the output variable is maize yield per acre, and the input variables are applied fertilizer nutrients, labor, maize seeds and machine usage. Since both the output and inputs are in per acre terms, land is not explicitly included as an input. Most of the maize fields are inter-crop fields where more than one type of crop is planted in the same season. Because most inputs (land, fertilizer and labor) are at the field level and cannot be separately allocated to maize production only, we generate an output index for inter-crop fields using:

$$Y_i = \left( \sum_j Y_{ij} P_j \right) / P_1, \quad (3.9)$$

where  $Y_i$  is the output index,  $P_j$  is the market price of crop  $j$ ,  $Y_{ij}$  is the yield of crop  $j$  in field  $i$ , and crop 1 is maize. Fields with more than three types of crops are deleted because we want to focus on the fields where maize is the major crop.<sup>4</sup> Only pre-harvest labor input (LABOR) is included because harvesting and post-harvest activities have little effect, if any, on yield. The unit of labor is person-hours. One person-hour of labor from children younger than 16 is transformed to 0.6 person-hours of adult labor. Nitrogen (FERTILIZER), the most important nutrient in maize growth, is computed from fertilizer application data according to the quantity and composition of each type of fertilizer used.<sup>5</sup> Maize seeds can be separated into hybrid seeds and local seeds. All fields used either hybrid seeds or local seeds (no combinations in the same field). These seed inputs are captured by two variables, SEED measures the amount of (hybrid or local) seed per acre applied to the

---

<sup>4</sup>637 out of the total 1718 fields are dropped.

<sup>5</sup>More than 20 types of fertilizers were applied. While some of these use nitrogen, phosphorous, and other nutrients in various proportions, nitrogen is usually the major nutrient deficiency and many of the major fertilizers use nitrogen and phosphorous in fixed proportions. Therefore, the level of applied nitrogen should give a reasonably accurate measure of the impact of fertilizer on yields.

field, and HYBRID is a dummy variable measuring one for hybrid seeded fields and zero otherwise. We also use a dummy variable MONO as an indicator for mono-crop fields because these might be expected to have systematically different yields than multi-crop fields. Tractor usage in land preparation is the only machine used for pre-harvest activities. This is captured by a dummy variable TRACTOR with one indicating that a tractor was used and zero otherwise.

Environmental variables are also included on the right hand side of the frontier production function. Failure to control for environmental variables may cause a correlation between some inputs and unobserved factors in the error term (for example, if a farmer makes input decisions based on soil properties that also affect maize yield) and therefore may bias estimates of the production frontier and inefficiency level ([44]). In order to control for environmental conditions, we include seven dummy variables indicating the different agro-economic zones. Farms in the same zone share similar terrain and climate conditions. We also include three village level variables: DRAINAGE, DRAINAGE<sup>2</sup> and STRESS. DRAINAGE captures the drainage property of the soil. It is a categorical variable ranging from one to ten where one indicates the least and ten the highest drainage. DRAINAGE<sup>2</sup> is the square of DRAINAGE. We include a quadratic term because yield is expected to increase in DRAINAGE at lower drainage levels and decrease at higher levels. Rainfall is a very important factor in maize production in Kenya because all of the maize fields are rain-fed and drought is the usual cause of yield loss. We use the variable STRESS to capture the moisture stress in maize growth. STRESS is computed as the total fraction of 20-day periods with less than 40 millimeters of rain during the 2003-2004 main season. This is a better measure for moisture conditions than total rainfall because total rainfall does not reflect the distribution of rainfall over time, which is very important in maize growth.

Any observations with missing values were discarded. Because of potential measurement errors, we also drop any observation that satisfies one of the following conditions: 1) yield lower than 65 kg per acre or higher than 4580 kg per acre, 2) seed usage less than two kg

per acre or more than 20 kg per acre, and 3) labor input less than 40 person-hours per acre or more than 2200 person-hours per acre. After these filters were applied, there are 815 fields (observations) remaining. The 815 fields were managed by 660 households. Table 3.1 summarizes the descriptive statistics for the variables included in the frontier production function (excluding zone dummies).

Table 3.1. Descriptive statistics for the variables in the production frontier

Variable	Notation	Mean	Std. Dev.	Min	Max
YIELD	Maize yield index (kg/acre)	1071	726	69	4410
LABOR	Pre-harvest labor input (person-hour/acre)	344	271	40	2160
FERTILIZER	Nitrogen fertilizer application (kg/acre)	11	12	0	63
SEED	Maize seed quantity (kg/acre)	8.5	3.3	2.5	18.8
TRACTOR	If tractor used in land preparation (1=yes, 0=no)	0.28	0.45	0	1
MONO	If mono-crop field (1=yes, 0=no)	0.11	0.31	0	1
HYBRID	If hybrid seed (1=yes, 0=no)	0.72	0.45	0	1
STRESS	Moisture stress (0-1)	0.14	0.21	0	1
DRAINAGE	Drainage of soil (categorical 1-10)	7.2	2.1	1	10

### 3.3.3 Exogenous Factors Affecting Efficiency

Previous studies have identified numerous factors that may limit farm productivity and efficiency. Education is arguably an important factor and Kumbhakar, Biswas, and Bailey ([26]) find that education increases the productivity of labor and land on Utah dairy farms while Kumbhakar, Ghosh, and McGuckin ([27]) also show that education affects production efficiency. Huang and Kalirajan ([21]) find that average household education level is positively correlated with technical efficiency levels for both maize and rice production in China. Here we measure education with EDUHIGH, the highest level of education among all household members.<sup>6</sup> We also investigate gender effects by including a dummy variable for female-headed households (FEMHEAD).

Physical and social infrastructure, such as road conditions, access to telephone and mobile phone service, access to extension service, etc., have also been mentioned for their role in rural development and farm productivity. Jacoby ([22]) examines the benefits of rural

---

<sup>6</sup>EDUHIGH may capture the effects of education on efficiency for a household better than the average education level or the education level of the household head, in that the one who receives the highest education can help the household head and the other household members in making production decisions.

roads to Nepalese farms and suggests that providing road access to markets would confer substantial benefits through higher farm profits. Karanja, Jayne, and Strasberg ([24]) show that distance to the nearest motorable road and access to extension services have positive effects on maize productivity in Kenya. More developed infrastructure helps farmers to obtain more information and thus may improve technical efficiency. Here we use three infrastructure variables to account for these effects on efficiency—DISTBUS, distance of the house from the nearest bus stop;<sup>7</sup> DISTPHONE, distance of the house from the nearest telephone or mobile phone service; and DISTEXTN, distance from the nearest extension service office.

Land tenure is another element that affects farm performance. Secure tenure may induce more investment (such as soil conservation) and increase farm productivity in the long run. Place and Hazell ([38]) suggest land tenure is important to investment and productivity in Rwanda. Puig-Junoy and Argiles ([39]) show that farms with a large proportion of rented land have low efficiency in Spain. Here we use a dummy variable (OWNED) with one indicating that the field is owned by the household and zero indicating the field is rented.

Financial constraints, such as limited access to credit, might also affect farm input decisions and efficiency. Ali and Flinn ([3]) show that credit non-availability is positively and significantly related to profit inefficiency for rice producers in Pakistan. Parikh, Ali, and Shah ([37]) find that farmers with larger loans are more cost efficient in Pakistan. The effects of financial constraints on technical efficiency seems to be unexamined to date but may be important because the timing of input usage can be an important factor influencing yields. So farms that face financial constraints may not be able to optimize production because inputs were not applied at the right times. We attempt to capture this effect using CRDCSTR (a dummy variable with one indicating the household has unsuccessfully pursued credit and zero otherwise), and RNFINC (the proportion of household members that

---

<sup>7</sup>We use DISTBUS instead of how far a household is from a motorable road, because only a very small proportion of the households in Kenya own motorable transportation tools (like tractors), and bus and bicycles are the major transportation tools there.

have non-farm income).

The relationship between farm productivity and farm size has been a long-standing empirical puzzle in development economics since Sen (1962) (see [10]; [8]; [28]). Empirical results on the relationship between efficiency and farm size have been mixed. Kumbahakar, Ghosh, and McGuckin ([27]) show that large farms are relatively more efficient both technically and allocatively. Ahmad and Bravo-Ureta ([1]) find a negative correlation between herd size and technical efficiency, while Alvarez and Arias ([5]) find a positive relationship between technical efficiency and size of Spanish Dairy farms. Huang and Kalirajan ([21]) show that the size of household arable land is positively related to technical efficiency in maize, rice and wheat production in China. Parikh, Ali, and Shah ([37]) find that cost inefficiency increases with farm size. Hazarika and Alwang ([18]) show that cost inefficiency in tobacco production is negatively related to tobacco plot size but unrelated to total farm size in Malawi. Here we include farm size (TTACRES) and field size (ACRES) as measures of the size effect. Descriptive statistics for the household survey data used to define the exogenous factors affecting efficiency are summarized in table 3.2.

Table 3.2. Descriptive statistics for the exogenous variables in the efficiency model

Variable	Notation	Mean	Std Dev	Min	Max
EDUHIGH	# school years for the highest educated member	12	5.5	0	24
FEMHEAD	If the household head is female (1=yes, 0=no)	0.19	0.39	0	1
DISTBUS	Distance to the nearest bus-stop (km)	2.4	2.4	0	20
DISTPHONE	Distance to the nearest phone service (km)	0.78	1.6	0	15
DISTEXTN	Distance to the nearest extension service (km)	5.2	4.5	0	33
OWNED	If the field owned by the household (1=yes, 0=no)	0.86	0.35	0	1
CRDCSTR	If pursued credits and was rejected (1=yes, 0=no)	0.08	0.27	0	1
RNFINC	% of members that have non-farming income	0.20	0.19	0	1
TTACRES	Total acres of land owned by the household	7.46	10.9	0.13	110
ACRES	Acres of the field	1.46	2.01	0.03	27

### 3.4 Estimation Results from Competing Models

In this section, we report the estimation results under alternative model specifications for the inefficiency component of the model. We use a flexible translog functional form for FERTILIZER, LABOR, and SEED in the frontier production function. We also inter-



act the dummy variable for hybrid maize (HYBRID), and the variable for moisture stress (STRESS), with FERTILIZER, LABOR, and SEED because there may be important interactions between these variables. With these choices, and given the constant and seven agro-economic zone dummy variables and  $\sigma_v^2$ , there are 30 parameters in the frontier production function. Furthermore, if we use the AAOS general model, there are another 22 parameters models used to specify  $\mu_i$  and  $\sigma_i$ . So the total dimension of the parameter space is 52. Even for the simpler models, such as the scaled Stevenson model, the KGMHLBC model, and the RSCFG model, the total parameter dimension is still large (42 parameters). To maximize a likelihood with such a high dimension can be computationally difficult given the complexity and non-regularity of the likelihood function. We do not want to eliminate potentially important variables *ex-ante*, nor do we want to sacrifice the flexibility of the general inefficiency model or use a less flexible frontier production function, such as Cobb-Douglas.

In our analysis we follow a three-step procedure to restrict the dimensionality of the parameter space:

1. Estimate the most general form of the production frontier using OLS and excluding the inefficiency component. Then drop jointly insignificant variables using F tests. This yields a reduced parameter space for the production frontier component.
2. Estimate the production frontier and the inefficiency component jointly with MLE using the general model for  $\mu_i$  and  $\sigma_i$ , but using the reduced set of explanatory variables for the production frontier obtained in step 1. Then drop jointly insignificant exogenous factors from the inefficiency component using likelihood ratio (LR) and Wald tests. This gives a reduced set of exogenous factors in the efficiency component.
3. Estimate the production frontier with the full set of explanatory variables and the efficiency component with the reduced set of exogenous factors after step 2 in one step using MLE. Then using LR and Wald tests, test the joint significance of the

explanatory variables dropped in the first step.

The OLS estimates in the first step are inconsistent if some of the  $x_i$ 's are correlated with  $\epsilon_i = v_i - u_i$ , or if the distribution of  $\epsilon_i$  is heterogeneous. In this study, we do not assume away either of these two possibilities. Therefore, our procedure to reduce the dimension of production frontier based on the OLS results in the first step may be inconsistent. To overcome this weakness, we take the third step of undertaking the same exclusion restrictions with a consistent estimator. Complete results from applying this procedure to the Kenya maize data are available in appendix E.

Using this procedure to impose zero restrictions, and then estimating the zero-restricted model for alternative specifications for the inefficiency part using MLE, leads to the estimation results provided in tables 3.3 and 3.4. Table 3.3 contains results for the frontier parameters and table 3.4 contains the inefficiency parameters, each under alternative model specifications for the inefficiency part. In the frontier model the variable selection criteria led seven parameters to be restricted to zero (second order effects for LABOR and SEED, all interaction effects among FERTILIZER, LABOR, and SEED, as well as interaction effects for SEED and HYBRID, and FERTILIZER and STRESS—see table 3.3). An additional two of the zone dummy variables (not reported in table 3.3) were also restricted to zero for a total of nine restrictions. In the inefficiency model, the variable selection criteria led to four effects being eliminated in the general model (DISTPHONE, DISTEXTN, CRDCSTR, and ACRES). But because these four effects enter both the mean and the variance terms in the general model this amounts to a total of eight zero restrictions—see the first column of table 3.4.

The parameter estimates for the frontier part of the model are very similar across alternative models for the inefficiency component (see table 3.3). Furthermore, Both the LR test and Wald test reject the null hypothesis that all the exogenous factors have zero effect on inefficiency at the 1% significance level in each of the five models (see table 3.4). Hence, it seems clear that the exogenous factors have a statistically significant effect irrespective

Table 3.3. Estimates for the production frontier in alternative models

LYIELD	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
LFERTILIZER	.15 (.020)	.15 (.020)	.15 (.020)	.15 (.020)	.15 (.020)
LLABOR	.33 (.050)	.33 (.052)	.33 (.049)	.33 (.052)	.33 (.052)
LSEED	.33 (.048)	.32 (.050)	.33 (.048)	.32 (.050)	.32 (.050)
LFERTILIZER <sup>2</sup>	.025 (.004)	.026 (.004)	.026 (.004)	.026 (.004)	.026 (.004)
LLABOR <sup>2</sup>	0	0	0	0	0
LSEED <sup>2</sup>	0	0	0	0	0
LFERTILIZER $\times$ LLABOR	0	0	0	0	0
LFERTILIZER $\times$ LSEED	0	0	0	0	0
LLABOR $\times$ LSEED	0	0	0	0	0
LFERTILIZER $\times$ HYBRID	-.062 (.016)	-.063 (.016)	-.063 (.016)	-.063 (.016)	-.063 (.016)
LLABOR $\times$ HYBRID	-.16 (.059)	-.15 (.061)	-.16 (.059)	-.16 (.061)	-.15 (.060)
LSEED $\times$ HYBRID	0	0	0	0	0
LFERTILIZER $\times$ STRESS	0	0	0	0	0
LLABOR $\times$ STRESS	-.23 (.14)	-.29 (.014)	-.26 (.14)	-.29 (.14)	-.29 (.14)
LSEED $\times$ STRESS	-.29 (.17)	-.28 (.19)	-.29 (.17)	-.27 (.20)	-.29 (.19)
HYBRID	.19 (.063)	.20 (.059)	.20 (.063)	.20 (.059)	.20 (.059)
STRESS	-.38 (.18)	-.36 (.18)	-.39 (.18)	-.36 (.18)	-.37 (.18)
MONO	-.22 (.059)	-.21 (.060)	-.23 (.058)	-.21 (.060)	-.21 (.60)
DRAINAGE	.15 (.056)	.13 (.056)	.15 (.055)	.13 (.057)	.13 (.056)
DRAINAGE <sup>2</sup>	-.012 (.005)	-.001 (.005)	-.011 (.005)	-.001 (.005)	-.001 (.005)
TRACTOR	.15 (.056)	.15 (.051)	.15 (.057)	.14 (.050)	.15 (.051)
Constant, Zone Dummies			not reported		
$\sigma_u^2$	.16 (.023)	.14 (.023)	.15 (.020)	.15 (.022)	.13 (.021)

Note: LYIELD is log YIELD. LFERTILIZER, LLABOR and LSEED are defined similarly. Standard errors are in parentheses.

of the model specification employed to model inefficiency. The Battese and Coelli efficiency estimates are computed for each observation in all the models and their correlations across alternative models are reported in table 3.5. The lowest correlation is 0.97. Therefore, all five models yield similar results for the production frontier and for the rankings of inefficiency among households, consistent with previous studies (e.g. [14]).

Goodness of fit statistic for the inefficiency component,  $R_z^2$ , are reported at the bottom of table 3.4 for the alternative model specifications. For example, the value of  $R_z^2$  for the KGMHLBC model is 0.1035, indicating that 10.35% of the sample variation in inefficiency can be explained by the exogenous factors. Not surprisingly, the general model provides the best fit at 12.75%.

The coefficients of the exogenous factors reported in table 3.4 are not very interesting by themselves because they are the parameters of the pre-truncated distribution of the inefficiency term  $u_i$ . So these parameters do not tell us how the exogenous factors affect the distribution of  $u_i$ . In order to quantify the effects of exogenous factors, we compute

Table 3.4. Estimates for the inefficiency components in alternative models

LYIELD	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
Variables in function $\mu_i^2$					
$\mu$	-4.1(6.9)	-0.30(0.36)	-1.45(0.72)	-0.75(0.40)	0
EDUHIGH	0.034(0.049)	-0.018(0.0068)	0.053(0.024)	0	0
FEMHEAD	-5.3(41)	0.22(0.093)	-2.3(2.0)	0	0
DISTBUS	-0.36(0.16)	0.048(0.016)	-0.31(0.14)	0	0
DISTPHONE	0	0	0	0	0
DISTEXTN	0	0	0	0	0
OWNED	-1.4(1.0)	0.35(0.11)	-1.3(0.41)	0	0
CRDCSTR	0	0	0	0	0
RNFINC	0.82(1.2)	-0.36(0.19)	1.4(0.73)	0	0
TTACRES	0.0018(0.045)	-0.013(0.003)	0.024(0.012)	0	0
ACRES	0	0	0	0	0
Variables in function $\sigma_i^2$					
$\sigma_u^2$	2.7(5.9)	0.42(0.13)	0.59(0.14)	0.54(0.12)	0.34(0.11)
EDUHIGH	-0.0063(0.015)	-0.018(0.0068)	0	-0.014(0.0048)	-0.032(0.014)
FEMHEAD	-0.22(0.28)	0.22(0.093)	0	0.18(0.072)	0.41(0.17)
DISTBUS	-0.014(0.044)	0.048(0.016)	0	0.040(0.012)	0.087(0.030)
DISTPHONE	0	0	0	0	0
DISTEXTN	0	0	0	0	0
OWNED	-0.061(0.46)	0.35(0.11)	0	0.28(.073)	0.63(0.22)
CRDCSTR	0	0	0	0	0
RNFINC	-0.14(0.36)	-0.36(0.19)	0	-0.29(0.15)	-0.63(0.38)
TTACRES	-0.012(0.013)	-0.013(0.003)	0	-0.011(0.0015)	-0.020(0.014)
ACRES	0	0	0	0	0
# observations	815	815	815	815	815
Log-likelihood	-616.30	-623.63	-618.71	-623.42	-623.70
LR statistic	56.84	34.54	50.62	38.36	37.98
Wald statistic	26.80	18.28	29.74	77.69	27.17
1% critical value	26.22	16.81	16.81	16.81	16.81
$R_z^2$	0.1275	0.0848	0.1035	0.0936	0.0773

Note: Standard errors are in parentheses. The LR and Wald statistics test the null hypothesis that the exogenous factors have no joint influence on inefficiency.

Table 3.5. Correlation of efficiency estimates among alternative models

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
General	1				
Scaled Stevenson	0.9793	1			
KGMHLBC	0.9910	0.9848	1		
RSCFG- $\mu$	0.9839	0.9986	0.9843	1	
RSCFG	0.9700	0.9970	0.9833	0.9917	1

$\partial[E(-u_i|x_i, z_i)]/\partial z_i$  and  $\partial[V(u_i|x_i, z_i)]/\partial z_i$  for each observation. The formulas for computing these measures and their standard errors for the general model are provided in appendix D. To obtain the formulas for the nested models, we only need to impose the corresponding restrictions on the parameters.<sup>8</sup>

The partial effects of the exogenous factors evaluated at the sample mean are reported in

<sup>8</sup>Wang (2002) gives the expressions for these derivatives but not for the standard errors.

table 3.6 along with their standard errors. The signs of the partial effects are the same for all the models. However, different models give quantitatively different values for the partial effects. For example, the partial effects of TTACRES on the conditional mean of  $-u$  range from 0.0023 to 0.0072, and these differences are large relative to the standard errors of the estimates. So conclusions about the semi-elasticity of output with respect to farm size may differ by a factor of more than 100%, depending on which inefficiency model is used.

Table 3.6. Partial effects of exogenous factors, evaluated at the sample mean

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
Partial effects on $E(-u_i x_i, z_i)$					
EDUHIGH	.0080(.0044)	.0079(.0012)	.0052(.0044)	.0080(.00081)	.0081(.0029)
FEMHEAD	-.12(.11)	-.10(.051)	-.14(.058)	-.11(.049)	-.11(.052)
DISTBUS	-.037(.025)	-.021(.0038)	-.037(.016)	-.022(.0028)	-.022(.0083)
OWNED	-.19(.074)	-.14(.047)	-.17(.052)	-.14(.042)	-.14(.058)
RNFINC	.19(.12)	.16(.039)	.13(.11)	.17(.028)	.16(.090)
TTACRES	.0075(.0021)	.0058(.00067)	.0023(.0015)	.0061(.00040)	.0049(.0023)
Partial effects on $V(u_i x_i, z_i)$					
EDUHIGH	-.0042(.0020)	-.0045(.0015)	-.0024(.0020)	-.0044(.0012)	-.0045(.0016)
FEMHEAD	.035(.058)	.064(.037)	.066(.026)	.063(.034)	.065(.038)
DISTBUS	.016(.013)	.012(.0055)	.017(.0072)	.012(.0049)	.012(.0057)
OWNED	.083(.040)	.070(.029)	.078(.021)	.068(.026)	.071(.035)
RNFINC	-.097(.062)	-.091(.048)	-.061(.050)	-.091(.043)	-.088(.051)
TTACRES	-.0046(.0016)	-.0033(.0011)	-.0011(.00070)	-.0033(.00083)	-.0028(.0014)

Note: Standard errors are in parentheses.

Table 3.7 reports the average partial effects of EDUHIGH on  $E(-u_i|x_i, z_i)$  for alternative model specifications over observations within each of the four quartiles of the efficiency levels.<sup>9</sup> The KGMHLBC model shows an increasing trend of the partial effect of education on efficiency levels from low to high quartiles, while the scaled Stevenson model, RSCFG- $\mu$  model and RSCFG model suggest a decreasing trend. So using the KGMHLBC model we would conclude that the households with lower efficiency levels would not benefit as much from increased education as the ones with higher efficiency levels. However, an opposite conclusion would follow if we use the scaled Stevenson model, the RSCFG- $\mu$  model or the RSCFG model.<sup>10</sup>

Table 3.8 reports the correlations of partial effects of EDUHIGH on  $E(-u_i|x_i, z_i)$  among

<sup>9</sup>The quartiles were computed using the KGMHLBC model.

<sup>10</sup> Similar patterns are observed for the other exogenous factors but these results are not reported to conserve space.

Table 3.7. Average partial effects of EDUHIGH on  $E(-u_i|x_i, z_i)$ , for the observations within each of the four quartiles based on efficiency levels predicted in KGMHLBC model

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
0-25% percentile	0.0067	0.0092	0.0039	0.0092	0.0092
25-50% percentile	0.0074	0.0085	0.0052	0.0085	0.0085
50-75% percentile	0.0078	0.0080	0.0059	0.0081	0.0081
75-100% percentile	0.0079	0.0069	0.0072	0.0070	0.0071

alternative models. Most correlations are very low and some are even negative.<sup>11</sup> This further confirms that different models yield rather different partial effects. Therefore, if we are only interested in the signs of the yield semi-elasticities with respect to exogenous factors, model specification is not important. However, if we are interested in the magnitudes of the yield semi-elasticities, it is important to choose the appropriate model specification.

Table 3.8. Correlation of partial effects of EDUHIGH on  $E(-u_i|x_i, z_i)$  among alternative models

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
General	1				
Scaled Stevenson	-0.3910	1			
KGMHLBC	0.7811	-0.7899	1		
RSCFG- $\mu$	-0.3716	0.9991	-0.7861	1	
RSCFG	-0.4140	0.9882	-0.8047	0.9970	1

## 3.5 Model Selection

In this section, we apply the procedure proposed by AAOS to select an appropriate model for our empirical application. A bootstrap analysis then follows to evaluate the performance of the model selection procedure.

### 3.5.1 Empirical Model Selection

We start with the general model, and then use LR tests to find simpler models that the data do not reject. Estimation of the general model yields a log-likelihood value of -616.30.

<sup>11</sup> Similar patterns are observed for the other exogenous factors but these results are not reported to conserve space.

Table 3.9 reports the log-likelihood values for the six restricted models nested in the general model. Taking the general model as the unrestricted model, we then test the restrictions that would reduce the general model to simpler specifications. LR test statistics with Chi-squared critical values are listed in table 3.9 and provide the following results:

- We can reject the scaled Stevenson model ( $\delta = \gamma$ ), RSCFG- $\mu$  model ( $\delta = 0$ ), and RSCFG model ( $\mu = 0$ ) at the 5% significance level.
- We fail to reject the KGMHLBC model ( $\gamma = 0$ ) at any reasonable significance level.
- We can reject the Stevenson model ( $\delta = \gamma = 0$ ) and ALS model ( $\mu = \gamma = 0$ ) at any reasonable significance level.

Because both the Stevenson model and ALS model are rejected, we conclude that the exogenous factors do affect efficiency. Among RSCFG, RSCFG- $\mu$ , and scaled Stevenson models, the RSCFG model is preferred because we fail to reject the RSCFG model at any reasonable significance level using the RSCFG- $\mu$  model or the scaled Stevenson model as the unrestricted model. Moreover, among all the models, the KGMHLBC model is most preferred because it is the only one that we can accept at any reasonable significance level. Therefore, we select the KGMHLBC model as our final model.

Table 3.9. Results of specification tests for model selection

	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG	Stevenson	ALS
log-likelihood	-623.63	-618.71	-623.42	-623.70	-641.44	-642.04
LR statistics	14.66	4.82	14.24	14.80	50.28	51.48
# restrictions	6	6	6	7	12	13
1% c.v.	16.81	16.81	16.81	18.48	26.22	27.69
5% c.v.	12.59	12.59	12.59	14.07	21.03	22.36
10% c.v.	10.64	10.64	10.64	12.02	18.55	19.81

The value of log-likelihood for the general model is -616.30.

### 3.5.2 A Bootstrap Evaluation

The model selection procedure proposed by AAOS leads to one clearly preferred model, the KGMHLBC model, among the set of competing models. However, it is also relevant to ask

about the reliability of the model selection criterion, which is a question of the size and power properties of the LR tests. We investigate this question using the bootstrap. That is, we generate data via the bootstrap assuming that the KGMLBC model is correct, and then we see how reliably the model selection procedure picks the KGMLBC model. So far as we are aware this approach has not been used previously in the literature. It is useful because we are using the bootstrap to evaluate the probability with which the actual model selection procedure will pick the correct model.

The KGMLBC model is written as

$$y_i = x_i' \beta + v_i - u_i, \text{ where } u_i \sim N[\mu \cdot \exp(z_i' \delta), \sigma_u^2]^+ \text{ and } v_i \sim N(0, \sigma_v^2). \quad (3.10)$$

We take the following steps to conduct the parametric bootstrap:

1. Using the actual sample data  $\{(y_i, x_i, z_i)\}_{i=1}^n$ , estimate the KGMLBC model using MLE to get  $\hat{\theta} = \{\hat{\beta}, \hat{\delta}, \hat{\mu}, \hat{\sigma}_u^2, \hat{\sigma}_v^2\}$ . These results are provided in tables 3.3 and 3.4.
2. Next generate pseudo-data sets based on the parameter estimates from step 1. That is, for  $i = 1, \dots, n$ , draw  $u_i^*$  from  $N[\hat{\mu} \cdot \exp(z_i' \hat{\delta}), \hat{\sigma}_u^2]^+$ ,  $v_i^*$  from  $N(0, \hat{\sigma}_v^2)$ , and then compute  $y_i^* = x_i' \hat{\beta} + v_i^* - u_i^*$ .
3. Based on the pseudo-data  $\{y_i^*, x_i, z_i\}_{i=1}^n$  generated in step 2, estimate all seven inefficiency models using MLE. Take the log-likelihood value ( $ll^*$ ) and parameter estimates ( $\hat{\theta}^*$ ) in each of the models, denoted as  $\zeta^* = \{(ll_j^*, \hat{\theta}_j^*)\}_{j=1}^7$ , where  $j$  indexes the different models.
4. Repeat steps 2 and 3 1000 times to obtain  $\mathcal{B} = \{\zeta_b^*\}_{b=1}^{1000}$ .

We use the log-likelihood statistics in  $\mathcal{B}$  to conduct the AAOS specification tests for each pseudo-data set, taking the general model as the unrestricted model and conduct LR tests at the 5% significance level. The results are:



- We reject the true model in 5.7% of the pseudo-data sets, the scaled Stevenson model in 75% of the pseudo-data sets, the RSCFG- $\mu$  in 78% of the pseudo-data sets, and the RSCFG in 75% of the pseudo-data sets.
- We reject both the Stevenson model and the ALS model in 99.9% of the pseudo-data sets. That is, in only one of the 1000 data sets, we would wrongly conclude that the set of exogenous factors do not affect efficiency.
- We accept the true model and reject all of the other models in 66.0% of the pseudo-data sets. We reject the true model and accept an alternative one at the same time in only 0.4% of the data sets.
- In 28.4% of the pseudo-data sets, we simultaneously accept the true model and at least one of the alternative models. And we reject all of the models simultaneously in 5.3% of the data sets.

These results suggest that the AAOS model selection criteria do a good job of discriminating between models. If the KGMHLBC model is correct, the model selection procedure will reject it with small probability (6%), and will pick it unambiguously with relatively high probability (66%).

The bootstrap results also can be used to generate confidence intervals for any of our original estimates. These confidence intervals may be more accurate in finite samples than those generated by first order asymptotic approximations such as the delta method. For example, we can use the parameter estimates of the KGMHLBC model in  $\mathcal{B}$  to compute the partial effects for every observation in each pseudo-data set. Confidence intervals then follow directly from the set of  $\mathcal{B}$  estimates. For example, given 1000 pseudo-data sets a 90% confidence interval for a parameter ranges from the 50th to the 950th largest values of the bootstrap estimates of that parameter. This is called the “percentile bootstrap”. Table 3.10 reports 90% percentile bootstrap confidence intervals for the partial effects in the KGMHLBC model, evaluated at the sample mean. For purposes of comparison, it also

gives the 90% confidence intervals based on the delta method (i.e. using the standard errors computed as in appendix D and reported in table 3.6). The confidence intervals given by bootstrap and the delta method are not very different. This confirms the reliability of the delta method.

Table 3.10. Partial effects of the exogenous factors on  $E(-u_i|x_i, z_i)$  and their 90% confidence intervals based on bootstrap and the delta method in the KGMHLBC model, evaluated at the sample mean

	EDUHIGH	FEMHEAD	DISTBUS	OWNED	RNFINC	TTACRES
	.0052	-.14	-.037	-.19	.13	.0023
Bootstrap	(.00047, .011)	(-.22, -.048)	(-.058, -.0078)	(-.28, -.035)	(-.011, .30)	(.00011, .0053)
Delta M.	(-.0020, .012)	(-.24, -.045)	(-.063, -.011)	(-.26, -.084)	(-.051, .31)	(-.0017, .0048)

### 3.6 Post-Estimation Analysis

Post-estimation analysis is based on the results of our selected KGMHLBC model. Table 3.11 reports output elasticity estimates for local seed users and hybrid seed users calculated at their respective sample means with their standard errors in parentheses.<sup>12</sup> The sum of the output elasticities with respect to FERTILIZER, LABOR, and SEED is less than 1 (0.80 for local seed users and 0.74 for hybrid seed users). However, this is expected and does not mean the technology is decreasing returns to scale because we are holding land constant (production is measured as yield per acre). Results show that output elasticities with respect to FERTILIZER and SEED are higher for hybrid seed users than local seed users, but the output elasticity with respect to LABOR is higher for local seed users.

Table 3.11. Output elasticity with respect to inputs for local seed users and hybrid seed users, evaluated at the sample means

Inputs	Local seed users	Hybrid seed users
FERTILIZER	0.209 (.00076)	0.224 (.0011)
LABOR	0.300 (.0027)	0.177 (.0063)
SEED	0.293 (.0032)	0.336 (.0026)

Note: Standard errors are in parentheses.

<sup>12</sup> The means of FERTILIZER, LABOR, and SEED are computed after taking logarithms.

Figure 3.6 plots the density of the Battese and Coelli technical efficiency estimates. The minimum efficiency level is 18% and the maximum is 98%. The mean of technical efficiency is 71%, while the mode is around 80%. The distribution is left skewed.

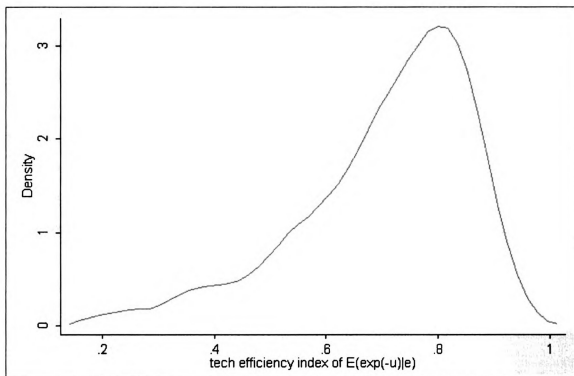


Figure 3.2. Kernel density estimate based on Battese and Coelli technical efficiency estimates

$R^2_z$  suggests that about 10% of the sample variation in inefficiency can be explained by the set of exogenous factors (see bottom of table 3.4). From table 3.6, EDUHIGH, RNFINC and TTACRES all have positive partial effects on the mean and negative effects on the variance of efficiency. FEMHEAD, DISTBUS, and OWNED all have negative effects on the mean and positive effects on the variance of efficiency. Therefore, an average household tends to have a higher efficiency level and a lower uncertainty on efficiency if it has a higher education level, more off-farm income, or larger farm size. Alternatively, it tends to have a lower efficiency level and higher uncertainty of efficiency if it has a female head, or is far from a bus-stop.

These results are mostly consistent with a priori reasoning and the previous literature. The effects of education, credit constraints, farm size and infrastructure on efficiency have been discussed extensively in the previous literature. The effect of female head could be due to the fact that females are subject to social discrimination in Kenya. There are generally two situations in which a female can become the head of a household. One is that she is a single mother, and the other is that her husband is dead. Females do not have the same inheritance rights as males in rural Kenya. A widow cannot obtain full rights to the land left by her husband and has to give away a certain proportion of the harvest to her husband's brothers. This may reduce the incentive to work intensively.

A surprising result is that farmers tend to be more efficient in rented fields than in their own fields. There are possible two reasons: 1) a fixed rent has to be paid at planting time, which provides more incentives for farmers who work in a rented field than in their own fields; 2) farmers rent fields that they know are productive. To the extent the second reason is a factor, the variable OWNED might capture the unobserved land quality not included as a covariate in the production frontier.

As explained earlier, not only the directions but the values of the partial effects on  $E(-u_i|x_i, z_i)$  are of economic interest. According to the KGMHLBC model (see table 3.10), one more school year would increase yield per acre by a little over half a percent for an average household, *ceteris paribus*. Being one kilometer closer to public transportation would increase yield per acre by 3.7 percent. An increase of one acre in farm size would raise yield per acre by less than one third of a percent. If the proportion of household members who receive off-farm income increases by 10 percent, yield per acre would increase by 1.3 percent. However, using the same amount and the same quality of inputs, a household with a female head tends to produce 14 percent less maize than a household with a male head, and farmers tend to produce 17 percent more maize working in rented fields than in their own fields.

### 3.7 Conclusion

This paper makes three contributions to the stochastic frontier literature. First, we provide formulas to compute the partial effects of exogenous farm characteristics on output levels and their standard errors for alternative model specifications. We also develop an  $R^2$ -type measure that shows the explanatory power of the exogenous factors that affect inefficiency.

Second, we examine the effects of model selection on inferences about firm inefficiency by applying several popular model specifications for the effects of firm characteristics on firm efficiency. The application is to Kenyan maize production data and we find that different specifications provide similar efficiency rankings of households and predict the same directions for partial effects of exogenous factors. However, the magnitudes of these estimated partial effects are rather different across model specifications. This finding calls for more attention to model selection in empirical stochastic frontier analysis.

Third, we apply the specification tests recently proposed by Alvarez, Amsler, Orea, and Schmidt ([4]) to choose between alternative model specifications for the Kenyan maize data. In our application these tests yield an unambiguous choice of best model, and an analysis of the model choice procedure using the bootstrap indicates that the model choice procedure is reliable. To our knowledge, bootstrapping has not been used previously to examine the size and power of these model selection criteria.

The empirical application uses the preferred model to identify factors that limit technical efficiency in maize production in Kenya, and quantify their partial effects on maize yields. We examine the effects of education, female head of household, distance from a bus stop, land owned or rented, extent of off-farm income, and farm size on the level of efficiency. Approximately 10% of the variation in efficiency levels is accounted for by these household characteristics, and while education, non-farm income, and farm size increase technical efficiency, female-headed households, distance from a bus stop, and land being owned rather than rented all decrease it.

# **APPENDICES**

# APPENDIX A

## Derivation of Equation (2.2)

By definition of lognormality,  $Z_j$  can be written as  $Z_j = e^{\mu_j + \sigma_j x_j}$  where  $x_j$  is standard normal,  $\mu_j = E(z_j)$ , and for  $z_j = \ln(Z_j)$ . Then equation (2.1) can be written as:

$$\begin{aligned} P_j(G_j) &= \beta \int_{-\infty}^{(g_j - \mu_j)/\sigma_j} (G_j - e^{\mu_j + \sigma_j x_j}) \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2} dx_j \\ &= \beta \left[ G_j \int_{-\infty}^{(g_j - \mu_j)/\sigma_j} \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2} dx_j - \int_{-\infty}^{(g_j - \mu_j)/\sigma_j} e^{\mu_j + \sigma_j x_j} \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2} dx_j \right] \\ &= \beta \left[ G_j N\left(\frac{g_j - \mu_j}{\sigma_j}\right) - e^{\mu_j + 0.5\sigma_j^2} \int_{-\infty}^{(g_j - \mu_j)/\sigma_j} \frac{1}{\sqrt{2\pi}} e^{-(x_j - \sigma_j)^2/2} dx_j \right] \\ &= \beta \left[ G_j N\left(\frac{g_j - \mu_j}{\sigma_j}\right) - e^{\mu_j + 0.5\sigma_j^2} N\left(\frac{g_j - \mu_j - \sigma_j^2}{\sigma_j}\right) \right], \end{aligned}$$

where  $N(\cdot)$  is the cumulative distribution function for the standard normal and  $g_j = \ln(G_j)$ .

# APPENDIX B

## Derivation of Equation (2.10)

Equation (2.8) can be written using the CRRA assumption as:

$$E \left\{ C_i^{-\alpha} [\max(G_j - Z_j, 0) - P_j] \right\} = 0 \quad (\text{B.1})$$

or

$$\int_0^\infty \int_0^{G_j} C_j^{-\alpha} (G_j - Z_j) f(Z_j, C_j) dZ_j dC_j = P_j E(C_j^{-\alpha}) \quad (\text{B.2})$$

Using the same notation as in A above and defining  $C_i = e^{\mu_{ci} + \sigma_{ci} x_{ci}}$  with  $\mu_{ci} = E[\ln(C_i)]$  and  $\sigma_{ci}^2 = \text{Var}[\ln(C_i)]$  where  $x_j$  and  $x_{ci}$  are bivariate standard normal with correlation coefficient  $\rho_{ij}$ , then (B.2) can be expressed:

$$\int_{-\infty}^\infty \int_{-\infty}^{\frac{g_j - \mu_j}{\sigma_j}} e^{-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})} (G_j - e^{\mu_j + \sigma_j x_j}) f(x_j, x_{ci}) dx_j dx_{ci} = P_j E[e^{-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})}] \quad (\text{B.3})$$

or

$$\int_{-\infty}^{\frac{g_j - \mu_j}{\sigma_j}} (G_j - e^{\mu_j + \sigma_j x_j}) f(x_j) dx_j \int_{-\infty}^\infty e^{-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})} f(x_{ci} | x_j) dx_{ci} = P_j E[e^{-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})}]. \quad (\text{B.4})$$

Also, from the properties of bivariate normal distributions, we have:

$$f(x_{ci} | x_j) = \frac{1}{\sqrt{2\pi(1 - \rho_{ij}^2)}} \exp\{(x_{ci} - \rho_{ij} x_j)^2 / [2(1 - \rho_{ij}^2)]\}. \quad (\text{B.5})$$



Substituting (B.5) into (B.4) and doing some algebra, we get

$$\int_{-\infty}^{\frac{g_j - \mu_j}{\sigma_j}} (G_j - e^{\mu_j + \sigma_j x_j}) e^{-\alpha \mu_{ci} - \rho_{ij} \alpha \sigma_{ci} x_j + 0.5(1 - \rho_{ij}^2) \alpha^2 \sigma_{ci}^2} f(x_j) dx_j = P_j E[e^{-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})}] \quad (\text{B.6})$$

Furthermore, we know that  $f(x_j) = \frac{1}{\sqrt{2\pi}} \exp(x_j^2/2)$  and  $E\{\exp[-\alpha(\mu_{ci} + \sigma_{ci} x_{ci})]\} = \exp(-\alpha \mu_{ci} + 0.5 \alpha^2 \sigma_{ci}^2)$ . Substituting these expressions into (B.6) gives

$$\int_{-\infty}^{\frac{g_j - \mu_j}{\sigma_j}} (G_j - e^{\mu_j + \sigma_j x_j}) e^{-(x_j + \rho_{ij} \alpha \sigma_{ci})^2/2} dx_j = P_j \quad (\text{B.7})$$

or

$$P_j = G_j N\left(\frac{g_j - \mu_j}{\sigma_j} - \alpha \rho_{ij} \sigma_{ci}\right) e^{-\mu_j + 0.5 \sigma_j^2 - \alpha \rho_{ij} \sigma_j \sigma_{ci}} N\left(\frac{g_j - \mu_j}{\sigma_j} + \alpha \rho_{ij} \sigma_{ci} - \sigma_j\right) \quad (\text{B.8})$$

Finally, notice that  $\rho_{ij} \sigma_i \sigma_{ci} = \text{Cov}(c_i, z_j) \equiv \sigma_{ij}$  so that (B.8) becomes (2.10).

# APPENDIX C

## Derivation of Equation (2.13)

Equation (2.13) can be easily derived from (2.12) if we show  $Cov(y, z_j) \approx \rho_{YZ_j} CV_Y CV_{Z_j}$  and  $Cov(w, z_j) \approx \rho_{WZ_j} CV_W CV_{Z_j}$ , where  $y = \ln(Y)$ ,  $z_j = \ln(Z_j)$  and  $w = \ln(W)$ . To begin, note that:

$$\rho_{YZ_j} CV_Y CV_{Z_j} = \frac{Cov(Y, Z_j)}{E(Y)E(Z_j)} = \frac{E(YZ_j) - E(Y)E(Z_j)}{E(Y)E(Z_j)} = \frac{E(YZ_j)}{E(Y)E(Z_j)} - 1. \quad (C.1)$$

Then from the properties of bivariate normal distributions, we have

$$\begin{aligned} E(YZ_j) &= E[\exp(y + z_j)] \\ &= \exp [E(y) + E(z_j) + 0.5Var(y) + 0.5Var(z_j) + Cov(y, z_j)] \\ &= \exp [E(y) + 0.5Var(y)] \times \exp [E(z_j) + 0.5Var(z_j)] \times \exp [Cov(y, z_j)] \\ &= E(Y)E(Z_j) \exp [Cov(y, z_j)] \end{aligned}$$

Substituting this result into (B.5), and rearranging gives:

$$Cov(y, z_j) = \ln \left( 1 + \rho_{YZ_j} CV_Y CV_{Z_j} \right) \approx \rho_{YZ_j} CV_Y CV_{Z_j}. \quad (C.2)$$

$Cov(w, z_j) \approx \rho_{WZ_j} CV_W CV_{Z_j}$  can be shown similarly.

# APPENDIX D

## Estimating Partial Effects of Exogenous Factors and their Standard Errors for the General Model

Assume there are  $K$  exogenous factors ( $K_1$  continuous variables and  $K_2 = K - K_1$  dummy variables). We deal with the continuous variables first. Let  $z_i^c$  be the  $K_1$  dimensional vector of the continuous variables. We derive the partial effects of  $z_i^c$  on the mean and variance of efficiency via differentiation as

$$\partial E(-u_i|x_i, z_i)/\partial z_i^c = \gamma^c \sigma_i (R_1 R_3 - R_2) - \delta^c \sigma_i R_1 (1 + R_3) \quad (\text{D.1})$$

$$\partial V(u_i|x_i, z_i)/\partial z_i^c = 2\gamma^c \sigma_i^2 (1 + R_3 + R_4) - \delta^c \sigma_i^2 R_4, \quad (\text{D.2})$$

where  $\mu_i = \mu \cdot \exp(z_i' \delta)$ ,  $\sigma_i = \sigma_u \cdot \exp(z_i' \gamma)$ ,  $\delta^c$  and  $\gamma^c$  are the coefficient vectors associated with  $z_i^c$ ,  $R_1$ ,  $R_2$ , and  $R_3$  are as defined in the text, and  $R_4 = R_1(R_2 + R_1 R_3 + 2R_2 R_3)$ .

Next we derive the variances of the partial effects of  $z_i^c$ . Let  $\theta' = (\delta' \ \gamma')$ , and  $g(\theta) = \partial[E(-u_i|x_i, z_i)]/\partial z_i^c$ , and  $h(\theta) = \partial[V(u_i|x_i, z_i)]/\partial z_i^c$ , where both  $g(\theta)$  and  $h(\theta)$  are  $K_1 \times 1$

dimensional vectors. Following the delta method,

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \longrightarrow N \left[ 0, \left( \frac{\partial g(\theta)}{\partial \theta'} \right) \Omega \left( \frac{\partial g(\theta)}{\partial \theta'} \right)' \right], \quad (\text{D.3})$$

$$\sqrt{n}[h(\hat{\theta}) - h(\theta)] \longrightarrow N \left[ 0, \left( \frac{\partial h(\theta)}{\partial \theta'} \right) \Omega \left( \frac{\partial h(\theta)}{\partial \theta'} \right)' \right]. \quad (\text{D.4})$$

We derive  $\partial g(\theta)/\partial \delta'$ ,  $\partial g(\theta)/\partial \gamma'$ ,  $\partial h(\theta)/\partial \delta'$  and  $\partial h(\theta)/\partial \gamma'$  as

$$\frac{\partial g(\theta)}{\partial \delta'} = -\sigma_i(\gamma^c z_i' + D)R_1(1 + R_3) - \sigma_i(\delta^c - \gamma^c)z_i'R_5, \quad (\text{D.5})$$

$$\frac{\partial g(\theta)}{\partial \gamma'} = \sigma_i \gamma^c z_i'(-R_1 - R_2 - R_1 R_4) + \sigma_i D(R_1 R_3 - R_2) + \sigma_i \delta^c z_i' R_5, \quad (\text{D.6})$$

$$\frac{\partial h(\theta)}{\partial \delta'} = \sigma_i^2 [\gamma^c z_i'(R_6 - 2R_4) - \delta^c z_i' R_6 - R_4 D], \quad (\text{D.7})$$

$$\frac{\partial h(\theta)}{\partial \gamma'} = \sigma_i^2 [\gamma^c z_i'(4 + 4R_3 + 4R_4 - R_6) + \delta^c z_i'(R_6 - 2R_4) + 2(1 + R_3 + R_4)D] \quad (\text{D.8})$$

where  $D = [I_{K_1} \ 0_{K_1 \times K_2}]$  is a  $K_1 \times K$  dimensional matrix, and

$$R_5 = R_1(1 + R_3) - R_1 R_4, \quad (\text{D.9})$$

$$R_6 = R_4 + R_1(2R_1 R_3 + 2R_1 R_3^2 - R_1 R_4 - 2R_2 R_4). \quad (\text{D.10})$$

$\frac{\partial g(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{g(\theta)}{\delta'} & \frac{g(\theta)}{\gamma'} \end{bmatrix}$  and  $\frac{\partial h(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{h(\theta)}{\delta'} & \frac{h(\theta)}{\gamma'} \end{bmatrix}$  are  $K_1 \times 2K$  dimensional matrices, which depend on the model parameters  $\delta$  and  $\gamma$ . We can get the estimates of  $\frac{\partial g(\theta)}{\partial \theta'}$  and  $\frac{\partial h(\theta)}{\partial \theta'}$  by substituting the estimates of  $\delta$  and  $\gamma$  into the above formulas. The variances of the partial effects can be estimated by substituting the estimate of  $\frac{\partial g(\theta)}{\partial \theta'}$  as well as the estimate of the variance-covariance matrix of  $\hat{\theta}$  into the formulas (D.3) and (D.4).

Next we compute partial effects of dummy variables. Let  $z_{ik}$  be the dummy of concern. The partial effect of  $z_{ik}$  on  $E(-u_i|x_i, z_i)$  and  $V(u_i|x_i, z_i)$  are

$$\begin{aligned} d(\theta) &= E(-u_i|x_i, z_i, z_{ik} = 1) - E(-u_i|x_i, z_i, z_{ik} = 0) \\ &= [-\sigma_i(R_1 + R_2)]|_{z_{ik}=1} - [-\sigma_i(R_1 + R_2)]|_{z_{ik}=0} \end{aligned} \quad (\text{D.11})$$

$$\begin{aligned} r(\theta) &= V(u_i|x_i, z_i, z_{ik} = 1) - V(u_i|x_i, z_i, z_{ik} = 0) \\ &= [\sigma_i^2(1 + R_3)]|_{z_{ik}=1} - [\sigma_i^2(1 + R_3)]|_{z_{ik}=0} \end{aligned} \quad (\text{D.12})$$

Similarly, following the delta method, we have

$$\sqrt{n}[d(\hat{\theta}) - d(\theta)] \longrightarrow N \left[ 0, \left( \frac{\partial d(\theta)}{\partial \theta'} \right) \Omega \left( \frac{\partial d(\theta)}{\partial \theta'} \right)' \right] \quad (\text{D.13})$$

$$\sqrt{n}[r(\hat{\theta}) - r(\theta)] \longrightarrow N \left[ 0, \left( \frac{\partial r(\theta)}{\partial \theta'} \right) \Omega \left( \frac{\partial r(\theta)}{\partial \theta'} \right)' \right] \quad (\text{D.14})$$

We then have  $\partial d(\theta)/\partial \delta'$ ,  $\partial d(\theta)/\partial \gamma'$ ,  $\partial r(\theta)/\partial \delta'$ , and  $\partial r(\theta)/\partial \gamma'$  as follows

$$\partial d(\theta)/\partial \delta' = [-\sigma_i R_1 (R_1 + R_3) z'_i]_{z_{ik}=1} - [-\sigma_i R_1 (R_1 + R_3) z'_i]_{z_{ik}=0} \quad (\text{D.15})$$

$$\partial d(\theta)/\partial \gamma' = [-\sigma_i (R_2 - R_1 R_3) z'_i]_{z_{ik}=1} - [-\sigma_i (R_2 - R_1 R_3) z'_i]_{z_{ik}=0} \quad (\text{D.16})$$

$$\partial r(\theta)/\partial \delta' = [-\sigma_i^2 R_4 z'_i]_{z_{ik}=1} - [-\sigma_i^2 R_4 z'_i]_{z_{ik}=0} \quad (\text{D.17})$$

$$\partial r(\theta)/\partial \gamma' = [(2 + 2R_3 + R_4) \sigma_i^2 z'_i]_{z_{ik}=1} - [(2 + 2R_3 + R_4) \sigma_i^2 z'_i]_{z_{ik}=0} \quad (\text{D.18})$$

$\frac{\partial d(\theta)}{\partial \theta'} = \left[ \frac{d(\theta)}{\delta'} \quad \frac{d(\theta)}{\gamma'} \right]$  and  $\frac{\partial r(\theta)}{\partial \theta'} = \left[ \frac{r(\theta)}{\delta'} \quad \frac{r(\theta)}{\gamma'} \right]$  are  $1 \times 2K$  dimensional matrices. The variances of the partial effects for  $z_{ik}$  can be estimated similarly as for the continuous variables described earlier.

# APPENDIX E

## Results of Specifying Relevant Explanatory Variables

Table E.1 reports the first-step OLS estimates for the original production frontier and the OLS estimates after we drop nine jointly insignificant variables. Standard errors are in parentheses. Robust standard errors using the Huber-White sandwich estimator of variance for households as clusters are used to model heteroscedasticity and autocorrelation among fields planted by the same households. Nine variables are dropped based on the F test [ $F(9,659)=0.25$ , P value=0.986]. Only STRESS, DRAINAGE and DRAINAGE<sup>2</sup> are individually insignificant at the 10% significance level in the remaining explanatory variables. However, STRESS is marginal [P value=0.112]. DRAINAGE and its squared term are jointly significant both in the original model [ $F(2,659)=4.57$ , P value=0.0107] and in the specified model [ $F(2,659)=4.72$ , P value=0.0092]. Therefore, we keep these variables in the production frontier component.

Table E.2 reports the second-step results of the LR and Wald tests for the general model, the scaled Stevenson model, the KGMHLBC model, the RSCFG- $\mu$  model, and the RSCFG model respectively. Our unrestricted models are the ones with the full set of exogenous factors. The restricted models have the reduced set of exogenous factors (EDUHIGH,

Table E.1. Specifying variables in the production frontier using OLS

Ln(output)	Original Model	Specified Model
LnN	0.15 (0.024)	0.15 (0.023)
LnLabor	0.36 (0.072)	.34 (0.060)
LnSeed	0.25 (0.11)	.32 (0.059)
LnN <sup>2</sup>	0.025 (0.0049)	0.025 (0.0047)
LnLabor <sup>2</sup>	-0.00002 (0.036)	0
LnSeed <sup>2</sup>	-0.065 (0.099)	0
LnN×LnLabor	-0.0023 (0.013)	0
LnN×LnSeed	-0.00031 (0.022)	0
LnLabor×LnSeed	-0.45 (0.079)	0
LnN×Hybrid	-0.058 (0.019)	-0.061 (0.020)
LnLabor×Hybrid	-0.19 (0.089)	-.17 (0.071)
LnSeed×Hybrid	0.094 (0.15)	0
LnN×Stress	0.026 (0.046)	0
LnLabor×Stress	-0.31 (0.16)	-0.25 (0.14)
LnSeed×Stress	-0.39 (0.25)	-0.43 (0.25)
Hybrid	0.24 (0.069)	0.23 (0.069)
Stress	-0.34 (0.30)	-0.33 (0.21)
Mono	-0.23 (0.070)	-0.23 (0.070)
Drainage	0.10 (0.068)	0.10 (0.066)
Drainage <sup>2</sup>	-0.007 (0.0065)	-0.007 (0.0063)
Tractor	0.19 (0.065)	0.19 (0.057)
Zones Dummies	not reported	
# Observations	815	815
R-squared	0.5236	0.5220

FEMHEAD, DISTBUS, OWNED, RNFINC, and TTACRES). Both the LR test and the Wald test fail to reject the null hypothesis that the four exogenous factors we dropped (DISTPHONE, DISTEXTN, CRDCSTR, and ACRES) all equal zero at any reasonable significance level for each of the models.

Table E.2. Tests results for specifying the exogenous factors in the efficiency component

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
Log-likelihood (unrestricted)	-613.02	-622.67	-616.13	-622.31	-623.05
Log-likelihood (restricted)	-616.30	-623.63	-618.71	-623.42	-623.70
LR statistics	6.56	1.92	5.16	2.22	1.30
Wald statistics	7.03	2.14	4.62	3.35	2.09
# restrictions	8	4	4	4	4
10% critical value	13.36	7.78	7.78	7.78	7.78

Table E.3 reports the third-step results of the LR and Wald tests for the alternative model specifications. The unrestricted models have the full set of explanatory variables in the production frontier while the restricted models have the reduced set of explanatory variables. Both the LR test and the Wald test fail to reject the null hypothesis that the

nine explanatory variables we dropped from the production frontier model in the first step all equal zero at any reasonable significance level for each of the models. In the end, we keep the reduced set of variables from the first step and the second step in our subsequent analysis.

Table E.3. Tests results for specifying explanatory variables in the production frontier

	General	Scaled Stevenson	KGMHLBC	RSCFG- $\mu$	RSCFG
Log-likelihood (unrestricted)	-614.89	-622.11	-617.21	-621.85	-622.16
Log-likelihood (restricted)	-616.30	-623.63	-618.71	-623.42	-623.70
LR statistics	2.82	3.04	3.00	3.14	3.08
Wald statistics	2.95	3.12	3.04	3.22	3.11



## BIBLIOGRAPHY

- [1] M. Ahmad and B Bravo-Ureta. An econometric decomposition of dairy output growth. *American Journal of Agricultural Economics*, 77:914–921, 1995.
- [2] D. Aigner, C.A.K. Lovell, and P. Schmidt. Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics*, 1977.
- [3] M. Ali and J. Flinn. Profit efficiency among basmati rice producers in pakistan punjab. *American Journal of Agricultural Economics*, 71:303–310, 1989.
- [4] A. Alvarez, C. Amsler, L. Orea, and P. Schmidt. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis*, 25:201–212, 2006.
- [5] A. Alvarez and C. Arias. Technical efficiency and farm size: a conditional analysis. *Agricultural Economics*, 30:241–250, 2004.
- [6] K. Arrow and R. Lind. Uncertainty and the evaluation of public investment decisions. *American Economic Review*, 60:364–378, 1970.
- [7] B. Babcock, C. Hart, and D. Hayes. Actuarial fairness of crop insurance rates with constant rate relativities. *AJAE*, 86.
- [8] C. Barrett. On price risk and the inverse farm size - productivity relationship. *Journal of Development Economics*, 51:193–216, 1996.
- [9] G.E. Battese and T.J. Coelli. Frontier production functions, technical efficiency and panel data: With applications to paddy farmers in india. *Journal of Productivity Analysis*, 1995.
- [10] D. Benjamin. Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, 46:51–84, 1995.
- [11] F. Black and M. Scholes. The pricing of options on corporate liabilities. *Journal of Political Economy*, 81, 1973.
- [12] M. Cao and J. Wei. Weather derivatives valuation and market price of weather risk. *Journal of Future Markets*, 24:1065–1089, 2004.

- [13] S.B. Caudill and J.M. Ford. 1993.
- [14] S.B. Caudill, J.M. Ford, and D.M. Gropper. Frontier estimation and firm specific inefficiency measures in the presence of heteroskedasticity. *Journal of Business and Economic Statistics*, 13:105–111, 1995.
- [15] R. Chambers. The valuation of agricultural insurance. University of Maryland, 2005.
- [16] G. Constantinides. Market risk adjustment in project valuation. *Journal of Finance*, 33:603–616, 1978.
- [17] J. Duncan and R. Myers. Crop insurance under catastrophic risk. *AJAE*, 82, 2000.
- [18] G. Hazarika G and J. Alwang. Access to credit, plot size and cost inefficiency among smallholder tobacco cultivators in malawi. *Agricultural Economics*, 29:99–109, 2003.
- [19] B. Goodwin and A. Ker. Nonparametric estimation of crop yield distributions: Implications for rating group-risk crop insurance contracts. *AJAE*, 71(1):139–153, 1998.
- [20] K. Hadri. Estimation of a doubly heteroscedastic stochastic frontier cost function. *Journal of Business and Economic Statistics*, 17:359–363, 1999.
- [21] C.J. Huang and J.T. Liu. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5:171–180, 1994.
- [22] H. Jacoby. Access to markets and the benefits of rural roads. *Economic Journal*, 110:713–737, 2000.
- [23] D. Jaffee and T. Russell. Catastrophe insurance, capital markets, and uninsurable risks. *Journal of Risk and Insurance*, 64:205–230, 1997.
- [24] D. Karanja, T. Jayne, and P. Strasberg. 1998.
- [25] A. Ker and B. Goodwin. Nonparametric estimation of crop insurance rates revisited. *AJAE*, 82:463–478, 2000.
- [26] S. Kumbhakar, B. Biswas, and D. Bailey. A study of economic efficiency of utah dairy farmers: A system approach. *Review of Economics and Statistics*, 1989.
- [27] S. Kumbhakar, S. Ghosh, and J. McGuckin. A generalized production frontier approach for estimating determinants of inefficiency in us dairy farms. *Journal of Business and Economic Statistics*, 9:279–286, 1991.
- [28] R. Lamb. Inverse productivity: Land quality, labor markets, and measurement error. *Journal of Development Economics*, 2003.

- [29] C. Lewis and K. Murdock. The role of government contracts in discretionary reinsurance markets for natural disasters. *Journal of Risk and Insurance*, 63:567–597, 1996.
- [30] R. Lucas. Asset prices in an exchange economy. *Econometrica*, 46:1429–1445, 1978.
- [31] W. Meeusen and J. van den Broeck. Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, 1977.
- [32] R. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4.
- [33] M. Miranda and J. Glauber. Systemic risks, reinsurance, and the failure of crop insurance markets. *AJAE*, 79:206–215, 1997.
- [34] C. Moss and J. Shonkwiler. Estimating yield distributions with a stochastic trend and nonnormal errors. *AJAE*, 75:1056–1062, 1993.
- [35] C. Nelson. The influence of distributional assumptions on the calculation of crop insurance premia. *North Central Journal of Agricultural Economics*, 12:71–78, 1990.
- [36] J. Nyoro, L. Kiriimi, and T. Jayne. Competitiveness of kenyan and ugandan maize production: Challenges for the future. Michigan State University International Development Working Paper, 2004.
- [37] A. Parikh, F. Ali, and M.K. Shah. Measurement of economic efficiency in pakistani agriculture. *American Journal of Agricultural Economics*, 77:675–685, 1995.
- [38] F. Place and P. Hazell. *Productivity Effects of Indigenous Land Tenure Systems in Sub-Saharan Africa*, 1993.
- [39] J. Puig-Junoy and J. Argiles. Measuring and explaining farm inefficiency in a panel data set of mixed farms. Pompeu Fabra University Working Paper, 2000.
- [40] D. Reifschneider and R. Stevenson. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review*, 32:715–723, 1991.
- [41] T. Richards, M. Manfredo, and D. Sanders. Pricing weather derivatives. *AJAE*, 86:1005–1017, 2004.
- [42] M. Rothschild and J. Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90:629–649, 1976.
- [43] M. Rubinstein. The valuation of uncertain income streams and the price of options. *The Bell Journal of Economics*, 7:487–500, 1976.

- [44] S. Sherlund, C. Barrett, and A. Adesina. Smallholder technical efficiency controlling for environmental production conditions. *Journal of Development Economics*, 69:85–101, 2002.
- [45] J. Skees and B. Barnett. Designing and rating an area yield crop insurance contract. *Review of Agricultural Economics*, 21:424–441, 1999.
- [46] J. Skees, R. Black, and B. Barnett. Designing and rating an area yield crop insurance contract. *AJAE*, 79:430–438, 1997.
- [47] R.E. Stevenson. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13:57–66, 1980.
- [48] J. Stokes and W. Nayda. The pricing of revenue assurance: Reply. 85:1066–1069, 2003.
- [49] J. Stokes, W. Nayda, and B. English. The pricing of revenue assurance. *AJAE*, 79:439–451, 1997.
- [50] T. Suri. Selection and comparative advantage in technology adoption. Yale University Job Market Paper, 2005.
- [51] L. Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. Cambridge and London: MIT Press, 1996.
- [52] C. Turvey. Contingent claim pricing models implied by agricultural stabilization and insurance policies. *Canadian Journal of Agricultural Economics*, 40:183–198, 1992.
- [53] C. Turvey and V. Amanor-Boadu. Evaluating premiums for a farm income insurance policy. *Canadian Journal of Agricultural Economics*, 37:233–247, 1989.
- [54] H.J. Wang. Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18:241–253, 2002.
- [55] H.J. Wang. A stochastic frontier analysis of financing constraints on investment: The case of financial liberalization in taiwan. *Journal of Business and Economic Statistics*, 2003.
- [56] H.J. Wang and P. Schmidt. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, 18:129–144, 2002.
- [57] S. Yin and C. Turvey. The pricing of revenue assurance: Comment. *AJAE*, 85:1072–1075, 2003.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02845 5297