THREE ESSAYS IN DEVELOPMENT ECONOMICS

By

Godwin Debrah

A DISSERTATION

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

Economics--Doctor of Philosophy
Agricultural, Food and Resource Economics--Dual Major

2017

ABSTRACT

THREE ESSAYS IN DEVELOPMENT ECONOMICS

By

Godwin Debrah

A large fraction of people in developing countries are engaged in some form of agricultural activity for a living, making it very important for any effort or program designed to address the issues with poverty or inclusiveness, to not abandon smallholder farmers. This dissertation, titled, *Three Essays in Development Economics* seeks to bring to the attention of policymakers on what to do and how to boost the incomes of rural dwellers in developing countries sustainably. I use data from Ghana for the first essay, the second essay is a theory paper and data from Tanzania is used for the third essay.

The first essay, which is titled, *Does the Inverse Farm Size-Productivity Hypothesis hold among Larger Farms. New Evidence from Ghana*, examines the relationship between the area planted in hectares and three measures of productivity, over a wide range of medium and large scale farm sizes that represent the fastest growing segment of farms in Africa and now account for a significant fraction of total area cultivated in the region. The results when Ordinary Least Squares estimation is used, point to a negative and significant relationship between farm scale and productivity. However, there is no statistically significant result in favor of the negative relationship once I instrument for the bias that may be due to measurement error in farm size variable. This can guide policymakers on how to redistribute land efficiently.

Inability of lenders to price for risk, limits the amount of loans financial institutions are able to give out. This is what the second essay titled, *Joint Liability Lending with Correlated Risks* explores. Because the rural-poor have no collateral to access loans, joint liability lending

has been a strategy microfinance institutions have used to price for risk and improve repayment rates. However, when project returns are correlated like those of farmers, joint liability lending may not help the lender to effectively price for risk to improve efficiency and repayment rates.

The second essay thus, theoretically, characterizes the optimal lending contracts in such a situation. I find that correlation reduces the efficiency of group-based joint liability lending relative to independent risks' case. Thus, correlation is bad for group lending. I also extend the model to show that, in some instances, it may be better for banks to separate whom they serve; serves either only borrowers with correlated project returns or borrowers with independent risk.

The third essay in this dissertation titled, *Predictors of the Choice of Rural Nonfarm Activity in Tanzania* investigates the predictors of participation in the rural nonfarm economy and the predictors of the choice between wage and self-employment conditional on participation. While some authors argue that households choose self-employment because they have no wage employment options, others argue self-employed people are self-selected entrepreneurs and should have some support. Understanding the predictors of the choice of employment would serve as a source of very relevant information to policymakers.

The results suggest that, assets are key predictors of households' participation in agricultural as well as in the rural nonfarm economy. For households that participate in the rural nonfarm economy, the choice between wage employment and self-employment is also related to the availability of assets. In sum, this dissertation uses both theoretical and empirical evidence to proffer solutions or suggestions that would help reduce poverty in developing countries.

I would like to dedicate this dissertation to my father Mr. Emmanuel Debrah and to the memory of my late mum Mrs. Selina Agbodza Debrah. I am eternally grateful for your love and extraordinary sacrifices that you made because of my sisters and me.

ACKNOWLEDGEMENTS

My profound gratitude goes to my major professors; Professor Christian Ahlin and Professor Thomas Jayne. Words, certainly, cannot describe how grateful I am for the financial support I received from Professor Thomas Jayne, without which I could not have finished my doctoral studies at Michigan State University. I am also particularly grateful to Professor Christian Ahlin for his support, patience and encouragement. You are forever part of my history and success story. I extend sincere thanks to Dr. Andrew Dillon and Dr. Leah Lakdawala. I really benefited from your valuable comments, directions and other support before and during my dissertation. I thank the PI and co-PIs of the GISAIA project.

I would also like to express gratitude to some of my friends who were of tremendous help to me during my doctoral studies at Michigan State University-- Dr. Benjamin Adu-Addai, Dr. Daniel Duah, Dr. Felix Yeboah, Dr. Serge Adjognon, Yi Li, Fei Jia, Chewe Nkonde and John Olwande just to mention a few. To my entire church family in Lansing, I say it was a great blessing meeting all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

RNFE	Rural Nonfarm Economy
IMF	International Monetary Fund
FAO	Food and Agricultural Organization
LSMS	Living Standards Measurement Study
ISA	Integrated Survey on Agriculture
RIGA	Rural Income Generating Activity
LFE	Linear Fixed Effects
CRE	Correlated Random Effects
MLE	Maximum Likelihood Estimation
TZS	Tanzanian Shillings
MFIs	Microfinance Institutions
IR	Inverse Relationship

1 INTRODUCTION

Although economic growth in many developing countries have led to some decline in poverty, inequality, unemployment and abject poverty still persist in many parts of the developing world (See recent World Bank 2016 report). Efforts by development agencies and the World Bank to eradicate extreme poverty and promote shared prosperity require that growth is inclusive. The poor and vulnerable would need good health, education and the needy given some safety nets including health insurance and access to modern technology. Majority of these poor households live in the rural areas and engage in some agriculture activity. Thus, efforts to lift these millions of people out of poverty would require that agricultural productivity be increased. Aggressive industrialization pursued, so that positive structural transformation can be achieved. The bulging youth and its attendant unemployment challenges in Sub-Saharan Africa, makes it imperative that policymakers aggressively work on expanding their economies, and provide jobs for the youth. Agricultural productivity while depending on technology and input used, also depends on farm management and whether the farmers are making the most of the land available to them amidst the challenges associated with acquiring land.

Despite the land issues that have bedeviled the Sub-Saharan Africa region, policymakers can maximize productivity by ensuring that land is available to those with high farm productivity. This can also be achieved by facilitating the existence of land markets. This dissertation that focuses on farm productivity, microcredit and the rural nonfarm economy, sheds light on how policymakers may be able to boost the income of rural dwellers most of whom participate in some form of agricultural activity, towards the eradication of extreme poverty and promoting shared prosperity. The first essay of this dissertation studies the relationship between productivity and farm size. We focus on the fastest growing segment of farms in Sub-Saharan Africa. Many previous studies focused on farm sizes that are less than 5 hectares.

The relationship has important ramifications for policy. If policymakers were to have a large parcel of land, there is the need for evidence-based decisions to be made regarding how that parcel of land could be reallocated. Would they allocate them in larger pieces to medium and large-scale farmers or subdivide them into smaller pieces for smallholder farmers. The latter could be motivated by the empirical regularity that small farms produce more per land area than larger scale farmers do. I use data from southern Ghana on medium scale farmers to investigate the inverse farm size and productivity hypothesis. This data is part of the Guiding Investments in Sustainable Agricultural Intensification in Africa (GISAIA) project. I also investigate whether labor market imperfection can drive the IR or not.

The results show a negative and significant relationship between scale/farm size and productivity when Ordinary Least Squares estimation is used. However, there is no statistically significant result in favor of the IR once I instrument for the bias that may be due to measurement error in farm size variable. This means that the empirical regularity may not hold over the range of medium scale farms considered in this study. This can guide policymakers on how to allocate land and how to formulate policies that would enable high productive farmers to get access to more land. A plethora of the literature on inverse farm size-productivity hypothesis, using data from Sub-Saharan Africa has focused almost exclusively on farm sizes that are less than 5 hectares. By considering larger scale farmers, i update the literature with newer evidence. I also find no evidence of labor market imperfection as the source of the IR.

The causes of low farm productivity are not limited to labor and land market imperfections or missing markets. Credit market imperfections also lead to inability of farmers to access loans to expand productivity. Even if labor or land markets work, many smallholder farmers are still constrained financially. Most farmers would need financial assistance to be able to purchase inputs including improved seeds. Aside the unavailability of collateral, farmers tend to have correlated risks, which makes them a very special category of borrowers. This motivates the second essay of the dissertation.

The second essay studies joint liability lending in the case where project returns are correlated, (Besley, 1994) identified three major things that make rural credit markets in developing countries less efficient compared to those in developed countries. These things were listed as collateral security unavailability, covariant risk and under developed states of related institutions. I approach the issue of correlated risks or correlated project returns theoretically. The problem is important both empirically and theoretically. Theoretically, it would be interesting to know how correlated risk affects the efficiency of the credit market. Empirically, correlated risk is a pervasive reality. It would be interesting to know whether lenders treat borrowers who are known to have correlated risks separately, or pools borrowers.

In this second essay, I derive and characterize the optimal lending contracts for a joint liability lending when risks are correlated. When correlation is introduced, I find that the parameter space for which lending can be fully efficient is smaller compared to the case where risks are not correlated. This is partly due to the monotonicity constraint that prevents the lender from raising the joint liability component of the contract above the interest rate, and partly because of affordability constraints. The monotonicity and affordability constraints work against a restoration of the implicit discount (premium) that joint liability offers (charges) safe (risky) borrowers- a discount/premium that varies negatively with correlation. Thus, correlation is bad for group lending. High correlation can lead to the exclusion of some potential borrowers such as the safe uncorrelated from the market there by reducing efficiency.

The results show that under certain conditions, such as situations where project returns are low and the fraction of correlated borrowers is high, fully efficient lending requires that the banks for correlated borrowers and those for non-correlated borrowers be separated. Governments or policymakers can set aside agricultural development banks that focus exclusively on farmers and develop strategies that can help improve efficiency in terms of outreach and repayment rates.

In the absence of credit or assets needed to go into entrepreneurial or non-agricultural self-employment, rural dwellers can engage in subsistence activities, or work in wage employment. The debate in the development economics literature on households' choice of employment motivates essay three. The third essay investigates the predictors of participation in the rural nonfarm economy and the predictors of the choice between wage and self-employment conditional on participation. While some authors have found that households choose self-employment because they have no option, others think self-employed people in rural areas are entrepreneurs and must be supported. That could mean wage workers are frustrated entrepreneurs. Further research is needed to ascertain the causes of the choice of employment by households. This is also necessary to accelerate governments' efforts geared towards eradicating extreme poverty.

The essay uses the World Bank LSMS data on Tanzania. In particular, the first three waves of this data set (2008/2009, 2010/2011 and 2012/2013). The FAO has data on rural income generating activities (RIGA) based on the LSMS data. I start by analyzing predictors of participation in rural non-agricultural wage and self-employment, and proceed further to investigate the predictors of the type of employment households choose once they participate in the rural nonfarm economy. Households are put into four groups. Group 1 has households that did not engage in any nonfarm activity. Group 2 comprises of households that engaged in some nonfarm activity consisting of only wage employment. Group 3 comprises of households that engaged in some nonfarm activity consisting of only self-employment and group 4 comprises of households that engaged in both wage and self-employment activities in the rural nonfarm economy.

The results suggest that, wealth, land or assets are key predictors of households' engagement in agricultural as well as the rural nonfarm economy. Whether households choose wage employment or self-employment upon participating is also related to the availability of such assets. The results seem to suggest that households are pulled into the rural nonfarm economy rather than being pushed into it as a survival strategy. Some of the results also suggest that reverse causality could be of less concern.

In sum, this dissertation uses both theoretical and empirical evidence to investigate ways policymakers can influence the poverty statuses of rural households or households that rely heavily on agriculture for their livelihoods. The essays guide African governments or governments of developing countries on interventions that can be made in terms of land allocation. Who gets credit and how credit can be made available to farmers to help them increase productivity and hence their incomes. The dissertation shows wealth or assets, or credit constraints can hinder the ability of rural dwellers to take advantage of the rural nonfarm economy as a way of boosting their incomes.

2 ESSAY 1: DOES THE INVERSE FARM SIZE-PRODUCTIVITY HYPOTHESIS HOLD AMONG LARGER FARMS? NEW EVIDENCE FROM GHANA

Introduction

Interest in the relationship between farm size and productivity in Africa, has been driven in recent years by rising doubts about the potential of smallholder-led agriculture growth in Africa e.g., (Collier & Dercon, 2014), and also by documented changes in farm size distributions observed in many African countries and in particular the rising share of cultivated land on medium-scale farms (Jayne et al., 2016; Jayne, Chapoto, Sitko, Nkonde, & Chamberlin, 2014). In light of such trends, major policy debates are arising over how the region's remaining prime agricultural land should be allocated, especially in light of rising land scarcity and land prices (Deininger, Savastano, & Xia, 2017; Holden & Bezu, 2016) and challenges associated with access to land for young people (Sezu & Holden, 2014).

The inverse farm size relationship (IR) refers to the observation that small plots or farms produce more output per unit area than larger plots or farms. The IR has been one of the enduring justifications for supporting smallholder farmers in developing countries. Recent studies upholding the inverse relationship between scale and productivity in Africa have generally been based on samples of farms almost entirely cultivating less than 10 hectares (Ali & Deininger, 2015; Barrett & Bellemare, 2010; Carletto, Savastano, & Zezza, 2013; Larson, Otsuka, Matsumoto, & Kilic, 2014). For example, less than one percent of the farms contained in the Living Standards Measurement Study-Integrated Survey on Agriculture (LSMS-ISA) analyzed in the studies cited above cultivate more than 10 hectares. Little is known about the relationship between output per unit area and farm size across the range of farms between 5 and

100 hectares, which have grown rapidly in recent years (Jayne et al., 2016).[1] Consequently, available evidence is unable to guide African governments' efforts to promote agricultural productivity through land policies that might encourage access to land to either small-scale or large farm units. Accurate information about the relationship between farm size and farm productivity for a wide range of farm scales can therefore provide valuable guidance into African governments' agricultural and land tenure strategies.

This paper examines the relationship between farm area planted and three measures of productivity, over a spectrum of farm sizes ranging from 5 to 100 [2] hectares in four districts of southern Ghana. These measures of productivity are: gross value of output per hectare planted, net value of output per hectare planted, and total factor productivity. Comparing the results of these alternative measures provide the means to evaluate the sensitivity of IR results to the use of partial vs. total factor productivity measures[3]. In another specification, I examine the robustness of my results to defining farm size in terms of the net value of output per hectare of potentially utilizable land (area planted plus fallow) in light of the possibility that, larger farms may utilize a smaller proportion of their total landholdings, which might be considered as foregone potential that could have been realized by others under alternative land distribution patterns.

The study makes several contributions to this literature. First, I investigate the IR hypothesis among farms sizes that are larger than typically studied in Africa, by using a sample of households that is statistically representative of agricultural households cultivating between 5

[1] In Ghana, farms cultivating between 5 and 100 hectares have grown rapidly since the early 1990s and now account for about half of all the farmland under cultivation (Table 2-1).
[2] The data has four observations greater than 100 hectares.
[3] The land productivity measure has been criticized as not being a true measure of productivity. Net profit has been argued as a better measure (Binswanger et al 1995). Total factor productivity looks at the productivity of all inputs and is able to reward (penalize) prudently (imprudently) managed farms. See table 2-20 for the various measures used in previous literature and their results.

and 100 hectares in the districts covered. Secondly, while a number of studies have conventionally measured productivity in terms of yield, I use both partial and total factor productivity measures as well as net values of production per hectare planted (the latter accounting for the costs of inputs and labor and therefore approximating profits per unit area). I then examine the consistency or sensitivity of results to the type of productivity measure used. The measures of productivity taken as a whole tend to point to a negative relationship when I use the Ordinary Least Squares method of estimation.

Third, I examine model robustness using the alternative measures of farm productivity and a different land area measures (area planted plus fallow land), and test for potential labor market imperfections as a possible cause of the IR.

Because family labor forms a significant component of farm labor, the valuation of family labor may affect the estimated relationship between scale and productivity. I investigate the role of family labor valuation in influencing the farm size/productivity relationship by computing shadow wages using the estimated production function and examining whether the relationship between productivity and land size differs depending on how family labor is valued. If labor market imperfections do not exist, I would expect the sign of the relationship between productivity and farm size to be unaffected whether family labor is valued using computed shadow wages or observed local wage rates. My estimation results suggest no evidence of such labor market imperfection.

As robustness checks, I examine the role of measurement error in farm size using an instrumental variable. I instrument for respondent reported area in the year of the survey (2014) using the respondent's subsequent report of 2013/14 area one year later. In summary, I find that, area planted is significantly inversely related to productivity among farms operating between 5 and 40 hectares in southern Ghana[4].

The rest of the paper is organized as follows. Section 2.2 reviews the literature. Section 2.3 describes the data. Section 2.4 presents the theoretical and empirical framework. Section 2.5 presents some robustness checks. Section 2.6 has the results and discussions from the regressions. The last section concludes.

Literature review

The inverse relationship between farm size and productivity was observed by Chayanov in Russia and later by (Sen, 1962) in India. Since then and with a few salient exceptions, their findings have been generally reinforced by subsequent research, at least within the range of relatively small farm sizes that tend to be examined in the literature. The works of (Lau & Yotopoulos, 1971), (Benjamin & Brandt, 2002), (Berry & Cline, 1979), (M.R. Carter, 1984), (Barrett & Bellemare, 2010), (Barrett, 1996), (Heltberg, 1998), (Barrett & Bellemare, 2010), (Larson, 2012), (Carletto et al., 2013), and (Ali & Deininger, 2015) found evidence of the existence of the IR. While (Kawasaki, 2010) used data from Japan and argued that a positive relationship between farm size and productivity could be obtained if there is little land fragmentation, (Dorward, 1999) using data from Malawi and (Kimhi, 2006) using maize plot data from Zambia, found a positive relationship. (Kevane, 1996), (Zaibet & Dunn, 1998), and (Binswanger, Deininger, & Feder, 1995), found no evidence to support the IR.

[4] This is when Ordinary Least Squares (OLS) estimation is used. There is no statistically significant result in favor of the IR once I instrument for the bias that may be due to measurement error in farm size variable using instrumental variable (IV) approach.

Many authors argue that the typical IR observed in the literature reflect either an omitted variable problem especially with respect to land quality and/or market imperfections (with labor, land and credit markets cited most frequently) or measurement error with respect to cultivated area. Measurement error can lead to the IR if smallholder farmers systematically understate their farm sizes and/or when larger-scale farmers systematically over-report their farm sizes. Recently, (Carletto et al., 2013) found that smallholders tend to overestimate plot sizes, while large scale farmers underestimate plot sizes.

In addition, they found that measurement error does not explain the IR, and that the IR is even more compelling after accounting for measurement error. (Gourlay, Dillon, McGee, & Oseni, 2016) also found evidence to support the IR. Their measurements of plot area were from compass and rope, GPS, and farmer self-reported. These results are different from (Lamb, 2003), who concludes that the IR in profits disappears after a dummy variable for share cropping and double cropping were used as instrument to overcome the bias due to measurement error. (Lamb, 2003) argues that, in the absence of any significant systematic measurement error in reported farm sizes, the IR stems from market imperfections and or inverse correlation between land quality and farm size. (Holden & Fisher, 2013) conclude that measurement error accounts for about 60% of the IR based on their study of farms smaller than 1 hectare in Malawi.

An example of how omitted variable bias can lead to the IR is given in (Chen et al., 2011). The argument is that, in rural China where local authorities divide land so that all local households get their minimum nutritional needs, very fertile lands are seen to be subdivided into smaller sizes making it seem that small farmlands are more productive if land quality is omitted in the analysis. (Assunção & Ghatak, 2003), (Bhalla & Roy, 1988), (Benjamin, 1995), (Chen et al., 2011), have all contributed to the IR literature from the omitted variables bias perspective.

While (Assunção & Ghatak, 2003) focused on individual heterogeneity among farmers, land quality measures are the main focus of the other studies. Recently, (Barrett and Bellemare., 2010) concluded that the IR cannot be entirely explained by unobserved land quality differences. Our study does not include soil quality measures. I include village dummies in all estimations to control for unobserved soil quality differences across villages and regions, recognizing that unobserved plot-level soil quality variation within villages will still be a concern.

The other major category of explanations for the IR is market imperfections. These are generally motivated by the fact that smallholders tend to use family labor more intensely than large farms in the presence of market imperfections. This makes labor-to-land, and output-to-land ratios higher on smaller farms, hence creating the observed IR (Carter, 1984; Carter & Wiebe, 1990). When land markets function imperfectly, larger landowners arguably have higher land-to-labor ratios. Because it is expensive to supervise labor, larger farms tend to have lower output to land ratios.

In place of an inverse relationship, (Kevane, 1996), used data from western Sudan and asserted that a positive relationship between productivity and farm size should be observed. In Kevane's work, if investment matters in production and market imperfections prevent smallholders from accessing credit to take advantage of a possible increase in productivity due to some initial investments, larger farms would be more productive.

Despite the plethora of studies in the literature, there is no clear consensus among agricultural economists as to what drives the IR or over which range of farm sizes it might hold. I follow (Ali & Deininger, 2015) and compute shadow wages from the production function to investigate the labor market imperfection argument. First, I test for differences in productivity between family and hired labor and secondly, I compare the relationship between productivity and farm size when family labor is valued using the district median wage versus using the estimated shadow wages. I am unable to reject similar productivity levels between hired and family labor, and neither do I find evidence of labor market imperfection. Section 3 has more discussions on this.

Very recently, (Bevis & Barrett, 2017) used a plot level data set from Uganda to argue that there is an edge effect (productivity of land higher around the edges of plots) which explains the puzzle. They attribute the source of this edge effect to behavioral mechanism. That is, farmers might take proper care of the more visible edge of plots than the interior. The data for this research however does not have the information that would allow me test this behavioral mechanism.

Choice of functional form may be especially important when testing the IR over a wide range of farm sizes. (Kimhi, 2006) found that when endogeneity of plot size is corrected using Heckman selectivity correction procedure, the relationship becomes U-shaped. This U shape is not unique to Kimhi's work. (Carter & Wiebe, 1990) found a positive relationship between net profits and farm size but a U-shape between output per acre and farm size. I thus run a regression that allows the slopes to differ by the range of farm size (0-5 ha, 5-20 ha and greater than 20 ha) to investigate this but found no significant results.

Results may also depend on assumptions about the nature of the production function. In the case of (Ali & Deininger, 2015), both Cobb-Douglas and Translog production functions were estimated for Rwandan farmers. They found that the negative relationship between profits and farm size disappeared once profits accounted for the cost of family labor valued at the local wage rate rather than at the lower shadow wage rate. Their work suggests that market imperfections may indeed be driving the IR at least in small farm settings where labor tends to be a major production input. Table 2-20 summarizes the methods and results of some selected literature.

Data

The data used in this study is from a survey conducted in southern Ghana between late July and first week of August in 2014 as part of the Guiding Investments in Sustainable Agricultural Intensification in Africa (GISAIA) project. Unlike most of northern Ghana which has recorded massive increases in the number of medium and large-scale farms in recent years (Chapoto, Mabiso, & Bonsu, 2013), most of southern Ghana is densely populated and faces acute land pressures. Four districts were purposively chosen from four regions of southern Ghana known to contain a relatively large number of medium and large-scale farms. The selected districts are Offinso North in the Ashanti region, Bibiani-Anhwiaso in the Western region, Afram Plains South in the Eastern region and Nkwanta North in the Volta region.

A simple random sampling method was used. Villages within districts were randomly selected and then farmers within villages randomly selected. Twenty villages were sampled at random from each selected district. Agricultural extension agents in those villages created a list of farmers operating more than 5 hectares in the 2013/2014 season in each of these 20 villages. Some of these households were subsequently found to have cultivated less than five hectares. The full population of listed farmers operating over 20 hectares was contained in the sample.

By sampling the entire population of such farmers, sampling bias issues are avoided. A farm or area planted or landholdings as used in this paper are all at the level of the household. In addition, for the purpose of this work, small farms are defined as those cultivating 5 hectares or less. Medium-scale farmers cultivate between 5 and 20 hectares. Large-scale farmers are those operating more than 20 hectares of land.[5] The sample of 503 farmers contains 57 farmers with operated land area less than 5 hectares, 385 farmers with operated area between 5 and 20 hectares, while 61 farmers have operated area above 20 hectares. Operated farm sizes at the 5th and 95th percentiles of the distribution are 5.3 and 41.9 hectares respectively. The results of this study can therefore be considered reasonably representative of farms operating up to roughly 40 hectares but not the entire population of medium and large-scale farms. Figure 2-5 has the distribution of the area planted or cropped area in hectares.

Where the household head of the sampled farm was not available at the time of the interview, he or she was replaced randomly with another farmer from the list. Most of the 38 replaced farmers had traveled out of the village at the time we visited those villages. Table 2-2 displays information on the number of farmers categorized by area planted. The survey collected information on all utilized fields in aggregates. Area measures are as reported by the farmers to enumerators; GPS or compass and rope methods were not used.

Table 2-3 presents descriptive statistics on the sampled farmers. The mean landholding is 18.32 hectares while the mean area cultivated was 12.85 hectares. For the price data, a section of the survey asked farmer respondents about the wages they paid on average for various agricultural activities per day. I use the median agricultural wage rate in the data to calculate the net value of production and also for testing the relative efficiency of hired versus family labor. Median village-level crop sales prices obtained from the survey data were used to value crop

[5] While farm size categorization is deemed reasonable for southern Ghana, I realize that it may not conform to those of other countries or regions with very different farm size distributions.

output. Prices of a few crops that were not available in the survey were replaced with prices from the regional markets, as reported on the website of the Ministry of Food and Agriculture in Ghana.

Theoretical and Empirical Framework

This section presents a simple theoretical framework that guides the empirical approach to estimating the relationship between land size and productivity. I consider a constant returns-to-scale Cobb-Douglas production function[6]. Farm households are assumed to face competitive prices and choose capital and labor to maximize profits given a fixed quantity of land.

Let output be given by $Y = AK^{\alpha}N^{\gamma}L^{1-\alpha-\gamma}$ (1)

where A is total factor productivity , K is amount of non-labor inputs used, N is labor input, L is fixed land quantity, r is rental rate of capital and w is the wage rate.

The problem is stated formally as:

$$Max\ \pi = AK^{\alpha}N^{\gamma}L^{1-\alpha-\gamma} - rK - wN \quad (2)$$

From the first order conditions with respect to labor and capital use, the following can be derived

$$N = \left[\frac{w}{\gamma AK^{\alpha}L^{1-\alpha-\gamma}}\right]^{\frac{1}{\gamma-1}} \text{ and } K = \left[\frac{r}{\alpha AN^{\gamma}L^{1-\alpha-\gamma}}\right]^{\frac{1}{\alpha-1}}$$

[6] Decreasing returns to scale by itself can lead to the IR. Results are not different from (Ali & Deininger, 2015) who used a translog production function in one specification of the empirical analysis. Many studies have not rejected the appropriateness of the Cobb Douglas production function

After some algebra, the optimal levels of capital and labor use in terms of fixed land can be written as

$$N^* = \left[\frac{w}{\gamma A\left[\left(\frac{r}{\alpha A}\right)^{\alpha}\right]^{\frac{1}{\alpha-1}}}\right]^{\frac{\alpha-1}{1-\alpha-\gamma}} \mathrm{L} \quad \text{and} \quad K^* = \left[\frac{r}{\gamma A\left[\left(\frac{w}{\gamma A}\right)^{\gamma}\right]^{\frac{1}{\gamma-1}}}\right]^{\frac{\gamma-1}{1-\alpha-\gamma}} \mathrm{L}$$

So that, the optimal output level can be written as

$$Y^* = DL \tag{3}$$

Where $D = A\left\{\left[\frac{r}{\gamma A\left[\left(\frac{w}{\gamma A}\right)^{\gamma}\right]^{\frac{1}{\gamma-1}}}\right]^{\frac{\gamma-1}{1-\alpha-\gamma}}\right\}^{\alpha}\left\{\left[\frac{w}{\gamma A\left[\left(\frac{r}{\alpha A}\right)^{\alpha}\right]^{\frac{1}{\alpha-1}}}\right]^{\frac{\alpha-1}{1-\alpha-\gamma}}\right\}^{\gamma}$

From (3), it can be seen that, under constant returns to scale, the coefficient of the land variable in the regression should be equal to one when we take logs of both sides. Empirically, most papers have used the log-linear specification in estimating the relationship. This is seen by taking the natural log of both sides of (3) to get

$$lnY = lnD + lnL \tag{4}$$

Therefore, a regression equation could be specified as;

$$lnY = \beta_0 + \beta_1 lnL + \epsilon \tag{5}$$

In which case, the null hypothesis is that $\beta_1 = 1$ while $\beta_1 < 1$ denotes the IR. This means that, output rises less quickly as land size rises. Alternatively, one can go back to (3) and divide through by L before taking the natural log of both sides. That gives us

$$ln\frac{Y}{L} = lnD + 0 \cdot lnL$$

Thus a regression specification from this, which is actually the most common specification in the literature, will be of the form;

$$lny = \beta_0 + \beta_1 lnL + \varepsilon \quad , \text{where } y = \frac{Y}{L} \tag{6}$$

In this case, the null is that $\beta_1 = 0$, while $\beta_1 < 0$ indicates a negative relationship between output per unit area and land area. I use specification in (6) in this research. This model is similar to that of (Assunção & Braido, 2007), except that in their model, farmers maximized expected profits. The theoretical framework above can also be used to motivate a regression model in which the dependent variable is the net value of production or profit per hectare. This is straightforward because, the profit function can be written as a linear function in the land size variable. As mentioned earlier, I present estimation results for both gross and net value of output per hectare planted. I also show results for total factor productivity measure for comparison purposes.

2.1.1 Estimation

To test the IR hypothesis, I run the regression of the form

$$lny_i = \beta_0 + \beta_1 lnL_i + \boldsymbol{X_i}\boldsymbol{\Gamma} + \varepsilon_i \tag{7}$$

where y_i is either total factor productivity or the gross or net value of output per hectare planted for household i, L_i is area planted and X is a vector of covariates which includes dummies for villages and whether major staple crops are grown (rice, maize and yam, groundnut and soy), and labor demand variables. Tables 2-3 and 2-4 have descriptive statistics of the variables used in the analysis.

Labor Market Imperfections

I test whether labor market imperfections drive the IR. To do this, I first test whether the efficiency of family labor is the same as that of hired labor. This enables me to see whether it costs more to use hired labor or not, and to determine how that affects larger scale farmers or gives advantage to smallholder farmers. (Feder, 1985) shows, "a model with no supervision effects on labor productivity would predict no relationship between farm size and productivity".

Data on labor used by the farmers was self-reported. Information was collected on all plots that were used by the farmer but the information is not at the plot level. Farmers stated lump sums of how much they expended on all utilized plots, both cash and in-kind expenditures for hired and exchange labor. Following (Benjamin, 1992), I test for the differences in efficiency using the equation $lnN = \alpha + \beta L + \delta lnw + (1-\theta)\frac{N^H}{N}$ (8)

Where N is total labor demand, w is the price of labor, N^H is hired labor thus $\frac{N^H}{N}$ is the fraction of hired labor used. i test under the null hypothesis that $\theta = 1$ or $(1-\theta) = 0$. A negative sign would suggest that hired labor is more efficient. By contrast, if hired labor were less efficient than family labor, this would make large-scale farms that use hired labor relatively intensively face a higher effective labor cost per worker than small farms. This can result in large farms having a lower labor to land ratio compared to small-scale farms and consequently, have a lower output to land ratio. Alternatively, the higher effective cost of labor can cause large-scale farmers to substitute capital for labor. If this were to be the case, the large-scale farms would hire less labor, so family members of small farms, who cannot work as wage laborers apply their excess labor intensively on their own farms, hence driving down the shadow wage of family labor. This is the classical explanation for how labor market imperfections may drive the IR.

The variable $\frac{N^H}{N}$ may have to be treated as endogenous because of a possible division bias or simultaneity. Results using ordinary least squares and instrumental variable estimations, fails to reject similar productivity levels, although estimates become quite noisy in the instrumental variable estimation. In a recent work by (Lafave & Thomas, 2014), separation or recursion in the agricultural household model is rejected but this is not due to differential cost of hired versus family labor. Suggesting that, markets are not complete and hired and family labor may not be of the same efficiency.

Various measures of net production value are calculated in this paper in order to observe any sensitivity of results to alternative labor valuation rules. I first compute the net value of production valuing family and exchanged labor at the hired labor median wage rate in the district. Estimation of the production function gives us the output elasticity of family labor.[7] The value of marginal product of family labor is estimated from the production function, which is used as the shadow wage. I then examine the robustness[8] of the relationship between net output per hectare and farm size when family labor is valued at shadow wages or the median agricultural wage rate in the district. This allows us to test the labor market imperfection, which says that, when households have surplus labor due to labor market imperfections, they apply

[7] MPL=$\hat{\beta}(\frac{\hat{Y}}{L})$, following (Jacoby, 1993) and also (Ali & Deininger, 2015).

[8] Getting shadow wages from the production function suggests bootstrapping for robust standard errors. I do this but do not report. The bootstrapped standard errors are slightly larger than those from the OLS estimation.

them on their small farms intensely thereby making the marginal product of labor consistently lower on small farms than the outside wage rate. Results derived from using the average agricultural wage rate and the shadow wage should not be significantly different if labor market imperfection does not exist.

For both sets of net value per hectare categories (based on observed local wage rates vs. shadow prices), I derive three different measures of family labor costs. First, I value child labor just as an adult laborer. In the second case, child labor is dropped from the analysis entirely so that, only family adult labor is accounted for. In the third case, hired labor is the only type of labor accounted for. These alternative approaches of valuing family labor provide a good assessment of the potential role of labor market imperfections in influencing the nature of the relationship between farm size and farm productivity.

Robustness Checks

Measurement error in area planted

Because recent studies have found that respondent reported farm size can be prone to measurement error, studies examining the relationship between farm productivity and farm size need to consider the extent to which their results are affected by potential measurement error bias. To do this, and as robustness check on my main results, I use a subsequent measure of area planted obtained from farmers a year after the main survey, as an instrument for the area planted variable. I show below how this strategy works. The basic idea is that in a true model of the form,

$$y = \alpha + \beta L + \varepsilon \ \ , L \perp \varepsilon,$$

if the regressor is measured inaccurately but there are two different inaccurate measures of it, say,

$L_1 = L + \varepsilon_1$ and $L_2 = L + \varepsilon_2$ where $L \perp \varepsilon_1, \varepsilon_2$ and $\varepsilon_1, \varepsilon_2 \perp \varepsilon$, $\varepsilon_1 \perp \varepsilon_2$

Without loss of generality, L_2 can be used as an instrument for L_1. Then in the limit,

$$\hat{\beta}_{IV} = \frac{cov(y, L_2)}{Cov(L_1, L_2)} = \frac{cov(\alpha+\beta L+\varepsilon, L_2)}{Cov(L_1, L_2)} = \frac{cov(\alpha+\beta L+\varepsilon, L+\varepsilon_2)}{Cov(L+\varepsilon_1, L+\varepsilon_2)} = \frac{cov(\alpha+\beta L+\varepsilon, L+\varepsilon_2)}{var(L)}$$

$$plim(\hat{\beta}_{IV}) \rightarrow \beta + \frac{cov(y, \varepsilon_2)}{var(L)} = \beta$$

Because the last condition $\varepsilon_1 \perp \varepsilon_2$ may not necessarily hold[9], results may not be consistent. However, Table 2-19 shows evidence of mean reversion from the two measures of area planted reported by the farmers. This is expected of random measurement errors; high reported values tend to be followed by low reported values. This may also suggest that the instrument could be plausible. I report the results as an attempt to address the potential measurement error bias.

Total Factor Productivity

I examine the sensitivity of our farm size/productivity relationship to whether partial or total factor productivity measures are used. While the net value of output measures may most closely approximate farmer profits per unit land, policy makers may be greatly interested in which scale of farming provides the greater return to all factors of production. Total factor productivity is difficult to measure but I calculate a representation of it. See (Li, Feng, & Fan, 2013) for details.

Assume the production function for simplicity is given by

[9] A typical case is when there is reporter fixed effect of mis-reporting.

$$Y_i = A_0 K_i^{\alpha_k} L_i^{\alpha_L} N_i^{\alpha_N} \exp(\varepsilon_i)$$

where Y_i is the gross output level of farmer i. K_i, L_i, N_i denote physical capital inputs, land and labor inputs respectively whereas α_j for j=K, L, N denote the output elasticities of capital, land and labor respectively. Taking logs and simplifying yields

$$lnY_i = lnA_0 + \alpha_K lnK_i + \alpha_L lnL_i + \alpha_N lnN_i + \varepsilon_i = B + \alpha_K lnK_i + \alpha_L lnL_i + \alpha_N lnN_i + \varepsilon_i$$

The coefficient of return to scale, CRTS= $\alpha_K + \alpha_L + \alpha_N$

Define $\alpha_j^* = \alpha_j / CRTS$ for i= K, L, N and obtain total factor productivity as

$$\text{TFP}_\text{i} = \frac{Y_i}{{K_i}^{\alpha_K^*} L^{\alpha_L^*} {N_i}^{\alpha_N^*}}$$

Results

I first estimated equation (8) to examine the relative productivity of family versus hired labor. From the results in Table 2-5, the coefficient of interest is the fraction of labor that is hired. This coefficient is positive but insignificant, suggesting that the fraction of hired labor has no significant impact on the total amount of labor demanded or used. This also means that, we cannot reject similar productivity of hired and family labor based on this test.

The results indicate no clear evidence that labor market imperfection adversely affects large farms due to additional costs of supervision,[10] associated with using hired labor. This type

[10] If large differences existed, it would mean that, cost of supervising hired labor would be higher and this corroborates the arguments that, supervision costs make larger scale farmers have a lower labor to land and consequently, lower output to land ratio. See (Feder, 1985) for a detailed discussion of how supervision cost can lead to a systematic relationship between productivity and land size.

of labor market imperfection could drive the IR if transaction costs associated with hired labor were large, as discussed in detail in section 2.2. Although there could be bias in the estimation of coefficients of interest in Table 2-5 due to simultaneity, or division bias, the average agricultural wage variable is significant at 5% and carries the expected sign. The log of area planted variable is also highly significant and has a positive sign. Gender and educational attainment of the household head are used as instruments for the fraction of hired labor used (reported in table 2-6). Similar levels of productivity cannot be rejected between the two types of labor. Both OLS and IV estimation results are similar to those obtained for rural Java by (Benjamin, 1992). However, the IV estimation is quite noisy so I cannot talk of strong identification. I treat findings in this paper as more of correlations.

After testing for possible differences in efficiency between family and hired labor, I estimate the value of marginal product of family labor (shadow wages) from a Cobb Douglas production function drawing on the works of (Ali & Deininger, 2015), and (Jacoby, 1993). Table 2-17 shows how the estimated shadow wages compare to the district median wages. Table 2-7 has the results from the production function estimation. The returns to scale estimate is reported in Table 2-7. I cannot reject constant returns to scale (CRS) production function at 1% significance level. It can also be seen that the coefficients of family labor and exchange labor are not significant while that of hired labor is highly significant. The remaining variables have the expected sign.

To understand the basic nature of the relationships between farm size and farm productivity, I run bivariate LOWESS regressions in both levels and logs and for gross and net measures of output per hectare. These bivariate relationships are presented in Figures 2-1 through 2-4. All four figures suggest an inverse relationship between planted area and the various measures of farm productivity for the range of farm sizes we concentrate on (5 to 40 hectares).

Table 2-8 contains the OLS results in which the dependent variables are the various gross and net output per hectare measures discussed in Section 2.4.1. Table 2-8 results are considered the baseline results to be compared to results from other estimations. NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor. NVP4 and NVP5 correspond to NVP1 and NVP2 respectively, except that the former uses the shadow price of family labor to value family labor. Table 2-9 contains estimation results from using the calculated shadow wages to compute net value of production instead of the median district wages used earlier.

Both sets of OLS results from Tables 2-8 and 2-9 uphold the IR. The estimated coefficients on log of land variable from the baseline results (table 2-8) are -0.31 and -0.53 for gross value per hectare and net value per hectare measures of productivity. These estimates are not much different from those obtained in recent studies such as (Ali & Deininger, 2015), (Carletto et al., 2013) and (Gourlay et al., 2016). Regardless of whether family labor is valued at shadow prices derived from the production function or from observed district median agricultural wage rates, and regardless of whether family or family plus hired or just hired labor is counted in the valuation of labor, the consistent finding is a significant inverse relationship between farm size and productivity between 5 to 40 hectares of cultivated area.

Table 2-10 presents the first-stage estimation when possible measurement error in area planted is taking into consideration. The results are significant and show that the instrument performs well in explaining the variations in the area planted variable. The relevance assumption required of a valid instrument appears to be well satisfied. Table 2-11 and Table 2-12 vary only in that labor is valued according to district median wage rates in Table 2-11 and according to shadow wages in Table 2-12. The negative relationship between farm size and productivity continues to hold in these models, and the point estimate on the farm size variable remains highly negative but now only weakly significant or in some cases insignificant relative to analogous results in Tables 2-8 and 2-9. The instrumental variable regression point estimates become noisy but do not change very much.

From Table 2-12, it can be seen that the estimates are smaller in magnitude than those in Table 2-9 and they also remain negative but are now insignificant. In addition, compared to baseline results in Table 2-8, the estimates are of similar magnitude but imprecisely measured. Taken altogether, the results in Tables 2-8 through 2-12 indicate that labor market imperfection do not appear to drive the IR. When Ordinary Least Squares estimation is used, the relationship is negative but there is no statistically significant result in favor of the IR once I instrument for the bias that may be due to measurement error in farm size variable.

Table 2-19 suggests that the farmers in the sample on relatively small holdings tended to overstate their area planted in the subsequent (2015) year compared to the original survey year (2014). As a result, this may have artificially inflated farm productivity at the lower part of the farm size distribution if the 2014 figure is understated. However, the results shown provide similar point estimates of the relationship albeit with less precision in the estimates. This result may be in contrast to findings in (Carletto et al., 2013; Gourlay et al., 2016) who found that failure to correct for respondent-reported area measurement error rather works against the IR.

Tables 2-13 and 2-14 present IR test results using operated land area (area planted plus fallowed land) instead of just area planted. When available land is to be redistributed to smallholders or medium or large scale farmers, policymakers may be interested in how farmers use (or not) the entire land under their control and not simply the amount cultivated. This is an important indicator of social efficiency in land utilization. Results show that the IR still exists but is less statistically significant for the net value per hectare measures in the IV estimation, just as was found in Table 2-11. It can also be seen from the last column in Table 2-13 that, if only hired labor was included in computing costs of labor, one could have concluded that there is no significant relationship, which could be misleading.

The fact that the IR is stronger and more precisely estimated in models of gross farm output per hectare rather than net farm output per hectare may mean that relatively large farms may be more efficient users of key inputs such as fertilizer, seed and hired labor, or may substitute more efficient mechanization for manual land preparation.

Table 2-15 presents results from IR tests based on a production function approach, in which other inputs to the production process are added to equation (7). Because these variables are endogenous, we favor the earlier reported models but report Table 2-15 results as a robustness check. These production function models produce statistically significant IR results consistent with those reported earlier from Ordinary Least Squares estimations.

In Table 2-16, we compare the robustness of the partial productivity models reported so far with those based on total factor productivity. Table 2-16 shows that both the OLS and the IV estimation show a highly significant inverse relationship between farm size and total factor productivity. The area elasticities of productivity from the OLS TFP results (-0.34) are near the mid-point of those from the partial productivity models. The area elasticities from the IV results (-1.16) are higher than the other estimates obtained and highly significant.

The weight of the evidence suggests that, regardless of whether total or partial productivity measures are used, regardless of plausible alternative ways of valuing family labor, regardless of whether farm size measures are based on area planted, area planted plus fallowed land, or total landholding size, I observe a strong inverse relationship between scale and productivity on farms between 5 and 40 hectares in southern Ghana when Ordinary Least Squares estimation is used. However, there is no statistically significant result in favor of the IR once I instrument for the bias that may be due to measurement error in farm size variable. I also did some sensitivity analysis by dropping data points that appear to be outliers based on Figure 2-6 and in another instance, restrict the sample to area sizes between the 1st and 99th percentile. The results are in table 2-21 and are not different from our main results.

Conclusion

This study examines the relationship between farm size and farm productivity, on farms cultivating between 5 and 100 hectares in southern Ghana but with emphasis on 5 to 40 hectares range. The study is unique in that, it covers the rapidly growing segment of “medium-scale” farms in Africa, which now accounts for a significant fraction of total area cultivated in the region (Jayne et al., 2016). Most available studies examining the relationship between farm size and productivity in Africa are based on small-scale farm samples with very few observations over 10 hectares. This study therefore can inform contemporary policy discussions about the pros and cons of promoting larger-farm scales in Africa e.g., (Collier & Dercon, 2014). I examine the relationship between farm size and farm productivity using (i) both partial and total factor productivity measures; (ii) alternative measures of farm size (area planted and area planted plus fallowed); (iii) valuing family labor at local farm wage rates and at shadow wages to account for the possibility of labor market imperfections driving the results; and (iv) using respondent based area measures as an instrumental variable in an attempt to correct for measurement error bias.

I find that the inverse relationship between farm size and farm productivity is consistently upheld through all permutations of these models. However, the instrumental variable estimation shows widely varying area elasticities of productivity, which, while all consistently negative, are often imprecisely measured. Thus, there is no evidence in favor of the IR when I use instrumental variable estimation to control for the bias that may be due to measurement error and account for labor use.

What do these results mean for policy? An important emerging policy debate centers on how unutilized land in Africa should be allocated to competing users. One school of thought argues that small is still beautiful while another argues for favoring larger-scale farmers who are often asserted to make more productive use of available land. My results can at least partially guide these discussions but have several limitations. First, while the findings from Ghana uphold the productivity advantages of relatively small farms, the largest portion of the data covers farms only up to 40 hectares. Therefore, I am not in a position to assess the relative productivity of small farms of, say 5 hectares, with large farms of 5000, 1000, or even 100 hectares.

Second, different farm scales may produce different general equilibrium effects that are not examined here. (Mellor, 1976), (Johnston & Kilby, 1975) and others have observed that "unimodal" farm distribution patterns, such as those found in much of green revolution Asia, have resulted in very different patterns of expenditures in local rural economies and hence produce multiplier effects of different magnitudes to those generated under bi-modal farm distribution systems such as those in much of Latin America. Given the potential importance of these general equilibrium effects of alternative agrarian structures, comparisons of the relative productivity of "small" versus "large" farms can provide highly important but incomplete information to guide governments toward comprehensive land and agricultural policy strategies.

APPENDICES

Appendix A: Tables for Essay 1

Table 2-1: Changes in Farm Structure in Ghana (1992 to 2013)

Farm Size Category	Number of Farms		% Growth in Number of Farms 1992 to 2013	% of Total Operated Land	
	1992	2013		1992	2013
0-2 ha	1,458,540	1,582,034	8.5	25.1	14.2
2-5 ha	578,890	998,651	72.5	35.6	31.3
5-10 ha	116,800	320,411	174.3	17.2	22.8
10-20 ha	38,690	117,722	204.3	11.0	16.1
20 -100 ha	18,980	37,421	97.2	11.1	12.2
Over 100 ha	-	1,740	-	-	3.5
Total	2,211,900	3,057,978		100	100

Source: (Jayne et al., 2016).

Table 2-2: Sample Size by Area Planted in Hectares

District	Region	<=5 ha	5-10 ha	10-20 ha	>20 ha	Total
Bibiani-Anhwiaso	Western	14	47	24	13	98
Nkwanta North	Volta	19	88	43	10	160
Afram Plains South	Eastern	12	49	35	19	115
Offinso North	Ashanti	12	53	46	19	130
Total (n)		57	237	148	61	503

Source: Authors' compilation from survey data.

Table 2-3: Household Demographics and Input Use by Operated Farm Area

Variable	Full Sample (n=503)	<=5 ha (n=22)	5 to 20 ha (n=383)	Above 20 ha (n=98)
	Means and Percentages			
Age of household head	45.51	47.27	44.91	47.51
Household head years in current settlement	38.18	39.32	38.04	38.49
Experience in farming	20.84	23.23	20.61	21.18
Male headed household (%)	95.00	91.00	94.00	99.00
Education of household head (%)				
No formal education	46.67	31.82	50.68	34.74
Basic education	22.08	45.45	20.94	21.05
Secondary education	18.13	13.64	17.90	20.00
Tertiary education	13.13	9.09	10.47	24.21
Household head previously employed (%)	10.93	9.09	10.70	12.24
Household head attracted to farming:				
Because parents were farmers (%)	49.50	40.91	51.96	41.84
Because farming is a business (%)	14.91	9.09	12.01	27.55
Household head applied for loan (%)	14.31	0.14	0.12	0.22
Operated area size (ha)	18.32	4.30	10.47	52.17
Area planted (ha)	12.85	4.30	8.90	30.23
Used fertilizer (%)	52.63	30.00	50.00	67.71
Fertilizer (kg/ha)	48.72	17.98	47.22	61.71
Number of crops grown	2.97	2.71	2.96	3.05
Number of fields	3.35	3.23	3.31	3.55
Used weedicide (%)	86.03	90.00	86.24	84.38
Used pesticide (%)	9.11	15.00	8.73	9.38
Used manure (%)	3.24	5.00	2.91	4.17
Used hired labor (%)	94.83	90.91	95.04	94.90
Used family labor (%)	74.55	68.18	78.33	61.22
Used communal labor (%)	16.90	9.09	18.02	14.29
Used mechanization (%)	74.55	45.45	74.15	82.65
Hired labor days per ha	36.62	71.02	35.41	33.64
Family labor days per ha	13.84	16.34	15.39	7.22
Communal labor days per ha	4.38	2.14	4.56	4.18
Household planted maize (=1)	85.49	81.82	84.86	88.78
Household planted yam (=1)	58.45	36.36	59.53	59.18
Household planted rice (=1)	17.50	22.73	17.75	15.31
Household planted groundnut (=1)	26.24	31.82	28.20	17.35
Household planted soy (=1)	0.60	4.55	0.26	1.02
Household planted cocoa (=1)	19.28	31.82	17.49	23.47

Source: Authors' computation from survey data. Operated farm area is area planted plus fallow area..

Table 2-4: Descriptive Statistics

Variable Label	Units of Measurement	Mean	Standard Deviation	Sample Size
Net value of production (NVP1) per hectare (valuing all labor, including family adult and child, and hired)	Ghana Cedis	5146.76	32399	503
Net value of production (NVP2) per hectare (values family adult and hired labor)	Ghana Cedis	5175.80	32413	503
Net value of production (NVP3) per hectare (values only hired labor)	Ghana Cedis	5363.01	32446	503
Gross Value of production per hectare	Ghana Cedis	6309.55	32496	503
Area planted	hectares	12.85	15.70	503
Area Operated	hectares	18.32	59.60	503
Hired labor	person-days	409.58	823.36	503
Family labor	person-days	121.96	200.94	503
Communal labor	person-days	37.72	98.70	503
Fertilizer cost	Ghana Cedis	591.24	1071.6	503
Chemical cost	Ghana Cedis	392.61	386.7	503
Fertilizer use	kilogram	450.99	881.7	498
Fraction of households planting maize	percentage	85.49	N/A	503
Fraction of household planting yam	percentage	58.45	N/A	503
Fraction of household planting rice	percentage	17.50	N/A	503
Fraction of household planting groundnut	percentage	26.24	N/A	503
Fraction of households planting soy	percentage	0.60	N/A	503
Fraction of households planting cocoa	percentage	19.28	N/A	503

Source: Authors’ computation from survey data.

Table 2-5: Test for Efficiency of Family versus Hired Labor (OLS)

	Log (Total Labor Demand)
Log (Area planted)	0.47 **
	(0.072)
Log(Average agric wage)	-0.14**
	(0.069)
Fraction of hired labor	0.17
	(0.140)
Constant term	5.42***
	(0.340)
Village dummies	Yes
N	493
Adjusted R^2	0.2351

Standard errors are in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2-6: Test for Efficiency of Family versus Hired Labor (IV)

	Log (Total Labor Demand)
Log (Area planted)	0.485***
	(0.106)
Log (Average agric wage)	-0.136*
	(0.077)
	-0.052
Fraction of hired labor	(1.249)
Village dummies	Yes
Constant term	5.535***
	(0.709)
N	492

Standard errors are in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Gender and education were used as instruments for the fraction of hired labor used.

Table 2-7: Cobb-Douglas Production Function Estimation

	Log (Gross Value of Production)
Log (Area planted)	0.576***
	(0.09)
Log (Hired labor days)	0.093***
	(0.03)
Log (Family labor days)	0.031
	(0.02)
Log (Communal labor days)	0.027
	(0.03)
Log (Chemical cost)	0.035
	(0.03)
Log (Fertilizer cost)	0.057***
	(0.02)
Shock in past 5 years	-0.206*
	(0.12)
Crop dummies	Yes
Village dummies	Yes
Returns to scale (RTS)	0.82
P-value (Null: RTS=1)	0.9421
Constant term	7.633***
	(0.46)
N	502
Adjusted R^2	0.3544

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-8: Estimates of the IR Valuing Family Labor at District Median Wages (OLS)

	Log (Gross Value Output Per Hectare)	Log (Net Value Output Per Hectare)		
	(1)	(NVP1)	(NVP2)	(NVP3)
Log (Area planted)	-0.31***	-0.53***	-0.53***	-0.25*
	(0.087)	(0.157)	(0.156)	(0.138)
Crop dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	8.04***	8.17***	8.30***	7.99***
	(0.401)	(0.696)	(0.691)	(0.637)
N	502	385	388	428
Adjusted R^2	0.2323	0.1704	0.1726	0.1910

Standard errors are in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Crop dummies are for maize, rice, yam, soy and groundnut. I also as robustness checks use area under the main crops (maize, rice, yam, soy, and groundnut) in place of crop dummies in all models. The results were not different. NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor.

Table 2-9: Valuing Family Labor at Shadow Wages (OLS)

	Log (Net Value Per Hectare)	
	(NVP4)	(NVP5)
Log (Area planted)	-0.437***	-0.442***
	(0.152)	(0.150)
Village dummies	Yes	Yes
Crop dummies	Yes	Yes
Constant term	8.318***	8.331***
	(0.680)	(0.672)
N	412	412
Adjusted R^2	0.1702	0.1642

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Crop dummies are for maize, rice, yam, soy and groundnut.
NVP4 is the net value of production that values family, communal and child labor at family shadow wages. NVP5 is the same as NVP4 but does not include child labor in its calculation.

Table 2-10: First Stage of IV Estimation

	Log (Area Planted Per Hectare)
Instrument[1]	0.136***
	(0.033)
Crop dummies	Yes
Village dummies	Yes
Constant term	1.67***
	(0.214)
Constant term	1.67***
	(0.214)
N	475
Adjusted R^2	0.2219

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[1] The instrument is the second measure of area planted collected in 2015 (see Section 3.4).
Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-11: Correcting for Measurement Error Bias in Area Planted (IV)

Log (Gross Value Per Hectare Planted)		Log (Net Value Per Hectare Planted)		
	(1)	NVP1	NVP2	NVP3
Log (Area planted)	-0.81**	-0.48	-0.68	-0.43
	(0.408)	(0.533)	(0.537)	(0.538)
Crop dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	9.08***	8.17***	8.57***	8.53***
	(0.858)	(1.260)	(1.267)	(1.278)
N	475	366	369	406
R^2	0.308	0.348	0.345	0.354

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
NVP1 is the net value of production that values family, communal and child labor using Median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor. The instrument is the second measure of area planted collected in 2015 (see Section 3.4).
Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-12: Correcting for Measurement Error Using Shadow Wages

	Log (Net value Per Hectare)	
	(NVP4)	(NVP5)
Log Area planted)	-0.13	-0.15
	(0.603)	(0.596)
Crop dummies	Yes	Yes
Village dummies	Yes	Yes
Constant term	7.79***	7.83***
	(1.439)	(1.420)
N	391	391
R^2	0.3282	0.3334

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The instrument is the second measure of area planted collected in 2015. (see Section 3.4). NVP4 is the net value of production that values family, communal and child labor at family shadow wages. NVP5 is the same as NVP4 but does not include child labor in its calculation. Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-13: Estimates of the IR Using Operated Farm Size (OLS)

	Gross Value Per Hectare	Net Value of Production		
	(1)	NVP1	NVP2	NVP3
Log (Area operated)	-0.406***	-0.536***	-0.551***	-0.079
	(0.083)	(0.139)	(0.138)	(0.125)
Crop dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	7.927***	7.987***	8.158***	7.638***
	(0.421)	(0.703)	(0.698)	(0.650)
N	502	385	388	428
R^2	0.361	0.354	0.353	0.350

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor. The instrument is the second measure of area planted collected in 2015 (see Section 3.4). Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-14: Estimates of the IR Using Operated Farm Size (IV)

	Gross Value Per Hectare	Net Value Per Hectare		
	(1)	NVP1	NVP2	NVP3
Log (Area operated)	-0.861*	-0.465	-0.679	-0.420
	(0.454)	(0.549)	(0.553)	(0.547)
Crop dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	9.621***	7.923***	8.599***	8.267***
	(1.167)	(1.495)	(1.503)	(1.509)
N	475	366	369	406
R^2	0.236	0.345	0.342	0.345

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor. The instrument is the second measure of area planted collected in 2015 (see Section 3.4).
Crop dummies are for maize, rice, yam, soy and groundnut

Table 2-15: Alternative Estimate of the IR (OLS)

	Log (Gross value of production/ha)
Log (Area planted)	-0.238**
	(0.09)
Gender	0.173
	(0.26)
Education	0.019*
	(0.01)
Experience in farming	0.018
	(0.02)
Experience squared	-0.000
	(0.00)
Hired labor days/ha	0.002**
	(0.00)
Family labor days/ha	0.003
	(0.00)
Communal labor days/ha	0.005
	(0.00)
Fertilizer kg/ha	0.003***
	(0.00)
Shock in past 5 year	-0.204*
	(0.12)
Weedicide Dummy	0.293*
	(0.16)
Mechanization Dummy	0.230*
	(0.14)
Constant term	7.263***
	(0.52)
N	480
Adjusted R^2	0.271

This is sometimes called the production function approach.

Table 2-16: Estimates of the IR Using Total Factor Productivity

	OLS	IV
Log(Area planted)	-0.341***	-1.156**
	(0.11)	(0.48)
Crop dummies	Yes	Yes
Village dummies	Yes	Yes
Constant term	3.878***	5.489***
	(0.45)	(1.00)
N	502	475
R^2	0.337	0.214

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The instrument is the second measure of area planted collected in 2015 (see Section 3.4).
Crop dummies are for maize, rice, yam, soy and groundnut.

Table 2-17: Computed Shadow Wages Versus District Median Wages

District	A Median Marginal Value Product of Family Labor (Cedis per day)	B District Median Wage (Cedis per day)
Bibiani	6.74	15
Afram Plains South	19.98	15
Offinso North	17.37	12
Nkwanta North	9.11	10

Source: Authors' compilation from survey data

Table 2-18: OLS Estimation of IR in Levels

	Gross Value Per Hectare	Net Value Per Hectare		
	(1)	NVP1	NVP2	NVP3
Area planted	-80.314	-61.693	-62.406	-65.252
	(119.73)	(119.45)	(119.50)	(119.59)
Crop dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	5177.758	3568.199	3620.320	3800.741
	(11692.55)	(11665.48)	(11670.14)	(11679.03)
N	502	502	502	502
R^2	0.139	0.138	0.138	0.138

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2-19: Comparing Self-reported Farm Sizes

Area Planted Category	Proportion that Over-reported in 2015 Compared to 2014	Proportion that Under-reported in 2015 Compared to 2014
5 hectares and less	0.267	0.733
5-20 ha	0.483	0.460
Above 20ha	0.723	0.214
N	285	200

Source: Author’s compilation from survey data.

Table 2-20: Summary of Previous Studies

Author	Study Area	Range of Farm Sizes Analyzed	Measure of Productivity or Efficiency-Numerator	Measure of Productivity or Efficiency--Denominator	Family Labor Costed? Y/N	Found IR
Sen (1962)	India	Not available	Gross value/Value added	Holdings	Yes	Yes
Lau and Yotopoulos (1971)	India	Greater than and less than 10 acres	Net Profit	Cultivable land in acres	Yes	Yes
Bardhan (1973)	India	Not available	Gross value	Cropped area	Yes	Yes
Carter (1984)	India	Less than 1 acre to above 12 acres	Gross value	Farm size	Yes	Yes
Bhalla & Roy (1988)	India	Not available but from 0 to above 12 acre	Gross value	Net cultivated area, land available for cultivation	No	IR weakly found
Kevane (1996)	Western Sudan	from small to above 12 Mukhammas	Gross value	Endowment of land own	No	No

Table 2-20 (cont’d)

Barrett (1996)	Madagascar	<25 ares and >500ares	Volume produced minus consumption—rice	Area cultivated	No	No
Heltberg (1998)	Pakistan	Mean is 1 hectare	Farm value added	Operated holding size	No	Yes
Doward (1999)	Malawi	<1 and >2 hectares	Output per hectare	Mean holding size	No	No
Benjamin and Brandt (2002)	China	Mean:10.5 mu	Gross Output	Cultivated land	No	No
Lamb (2003)	India	Mean 2.18 acres	Profit	Total cropped area	Yes	Yes but goes away
Kimhi (2006)	Zambia	Saddle point 3 ha Mean: 1.78 hectares	Maize output	Area allocated to maize	No	Yes and No
Assuncao and Braido (2007)	India	0.08-83.87 acres	Gross value	Cropped area	No	Yes
Kawasaki (2010)	Japan	Mean:0.90 ha	Rice output kg/ha	Total planting area	No	No

Table 2-20 (cont'd)

Barrett et al (2010)	Madagascar	Mean: 16.24 ares	Yield per unit area	Cultivated area	No	Yes
Chen et al (2011)	China	Mean: 0.07 hectare	Total crop output	Farm land cultivated	No	No
Larson et al (2012)	10 African countries	0.41-7.28 hectares	Yield kg/ha	Area under crop	No	Yes
Carletto et al (2013)	Uganda	0.01 to 600 acres	Net agricultural revenue	Area operated	Yes	Yes
Holden & Fisher (2013)	Malawi	0.1-0.8 ha	Net agricultural revenue	Farm size	Yes	Yes
Li et al (2013)	China	0.3-100 mu	Gross value/net profit	Farmland area	Yes	Yes
Ali & Deininger 2015	Rwanda	0.05-2 hectare	Gross value/profit	Cultivated area	Yes	No
Dillon et al (2016)	Nigeria	Not available	Gross output value	Plot size	No	Yes

Source: Author's compilation.

Table 2-21: OLS Results Using Sample between 1st and 99th Percentile

	Log Gross Value Per Hectare	Log Net Value Per Hectare		
	(1)	NVP1	NVP2	NVP3
Log (Area planted)	-0.275***	-0.478***	-0.478***	-0.209
	(0.08)	(0.15)	(0.15)	(0.13)
Main crops dummies	Yes	Yes	Yes	Yes
Village dummies	Yes	Yes	Yes	Yes
Constant term	7.953***	8.042***	8.175***	7.854***
	(0.37)	(0.67)	(0.66)	(0.61)
N	498	381	384	424
R^2	0.370	0.352	0.351	0.356

Standard errors are in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B: Figures for Essay 1

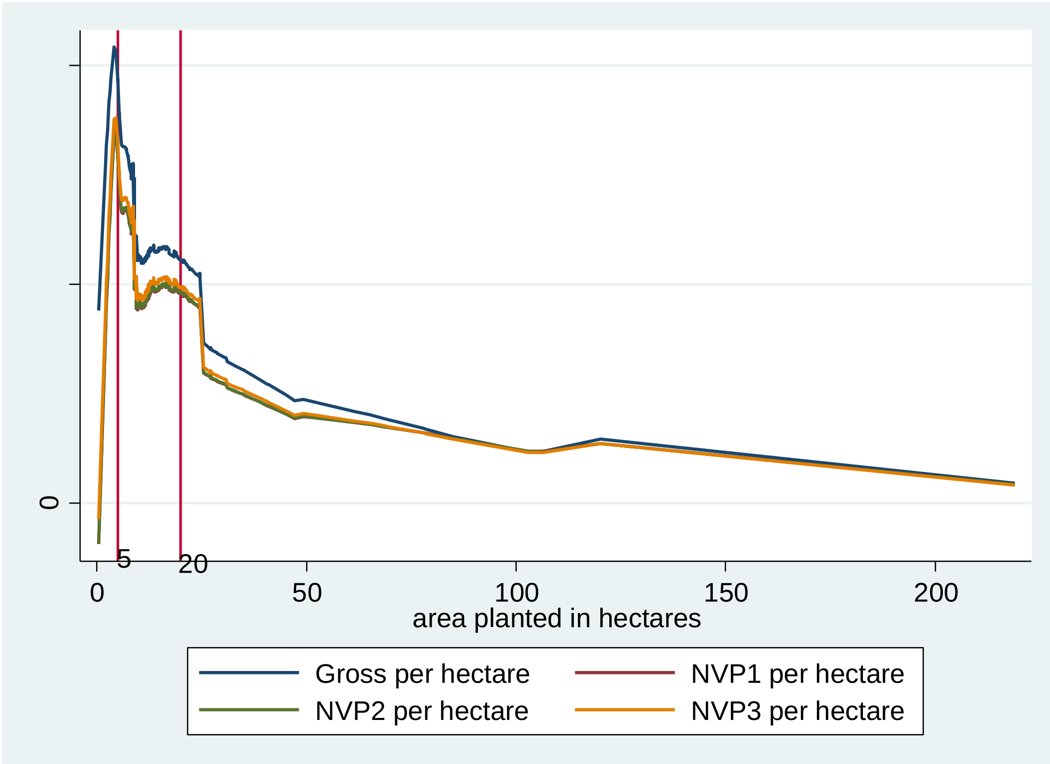

Figure 2-1: Plot of Measures of Productivity and Area Planted in levels

NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor.

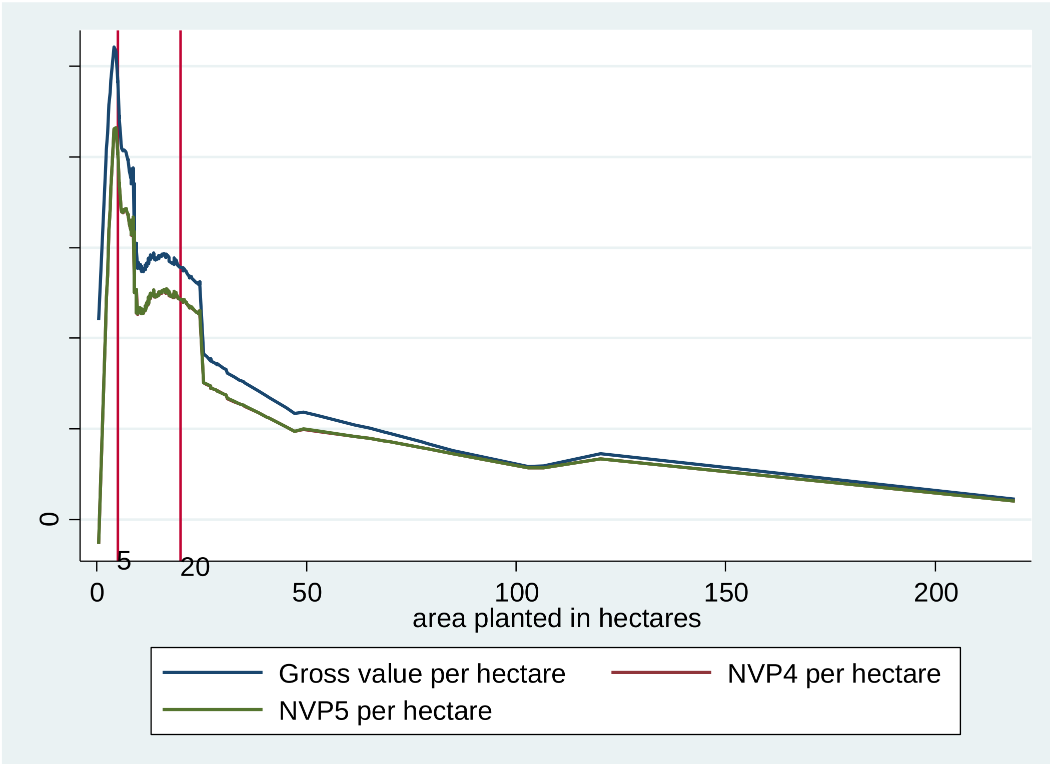

Figure 2-2: Plot of Measures of Productivity and Area Planted in levels using shadow wages
NVP4 is the net value of production that values family, communal and child labor at family shadow wages. NVP5 is the same as NVP4 but does not include child labor in its calculation.

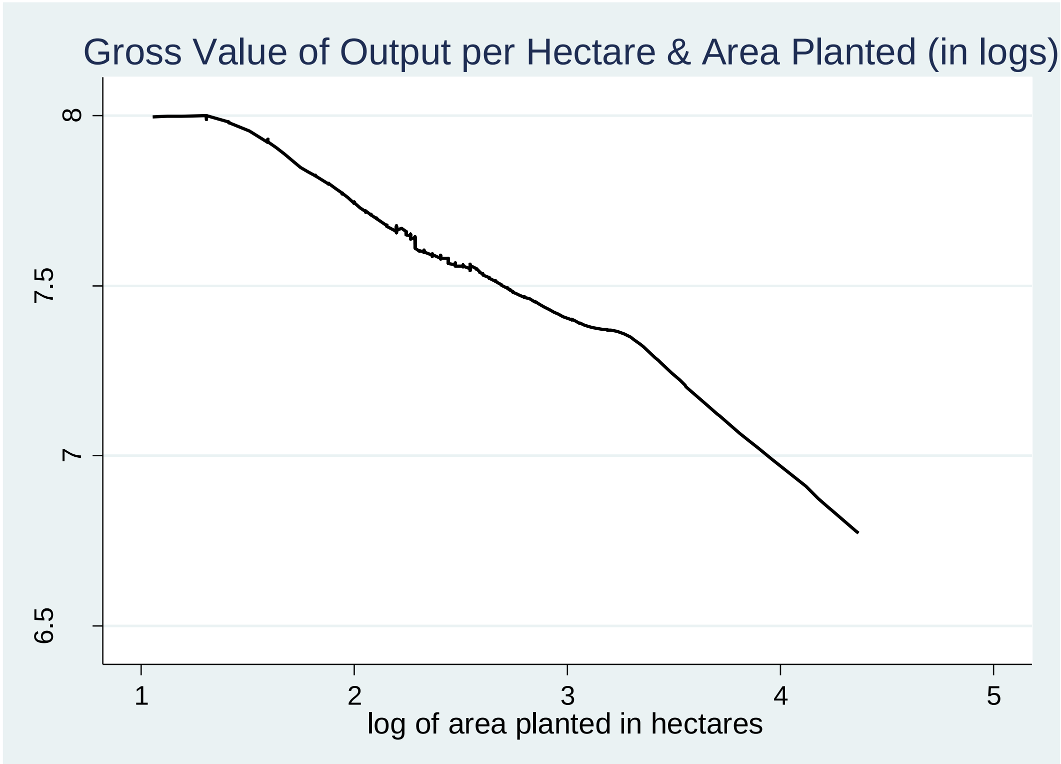

Figure 2-3: Gross Output per Hectare Against Area Planted

Family labor is value at the district median wage rate

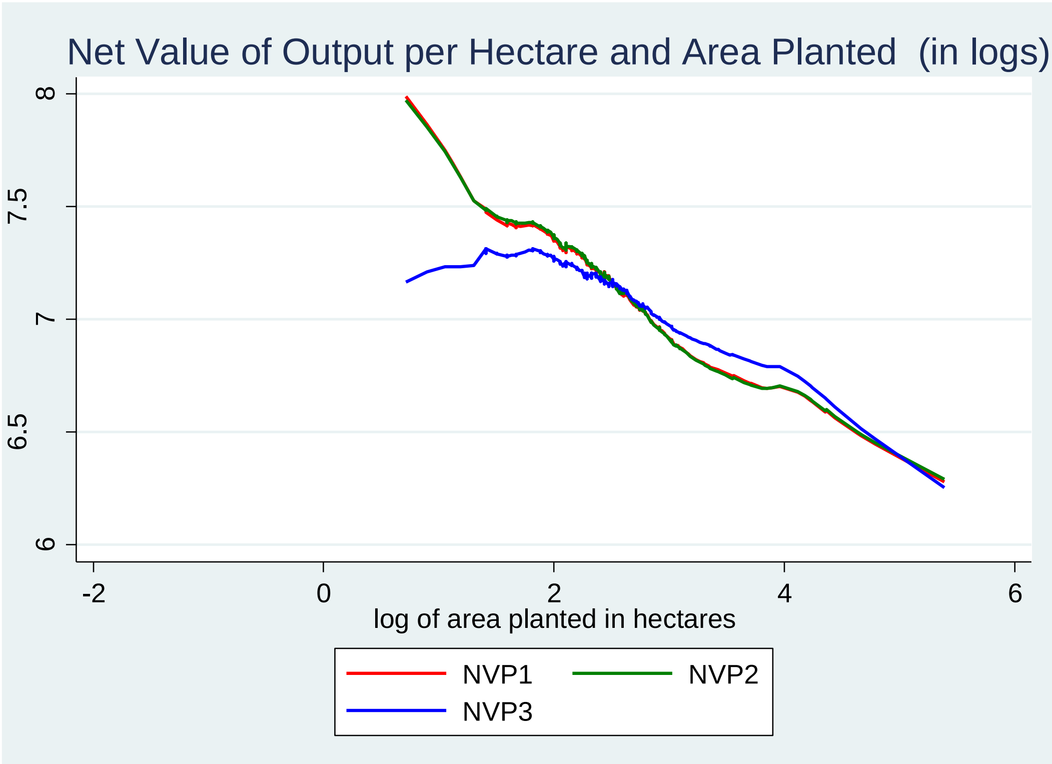

Figure 2-4: Plot of Measures of Productivity and Area Planted in logs

NVP1 is the net value of production that values family, communal and child labor using median wage of agricultural activities in the respective districts. NVP2 is the same as NVP1 but does not include child labor in its calculation. NVP3 uses only hired labor and does not include the cost of using other types of labor.

.

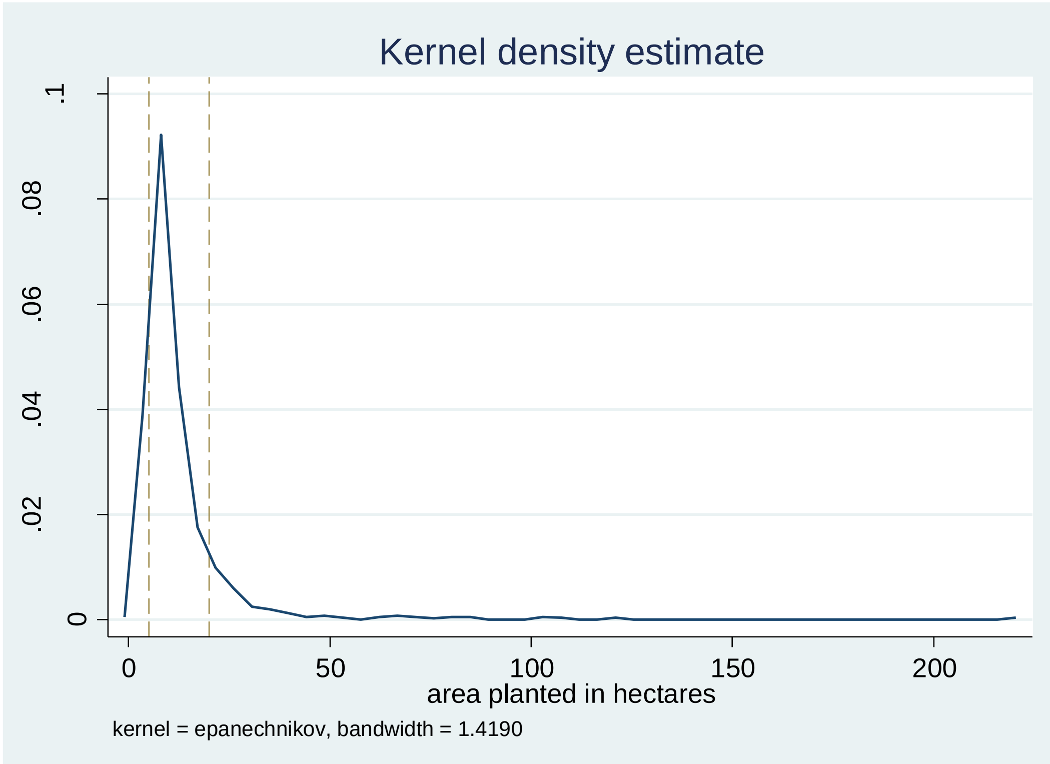

Figure 2-5: Distribution of Area Planted Variable

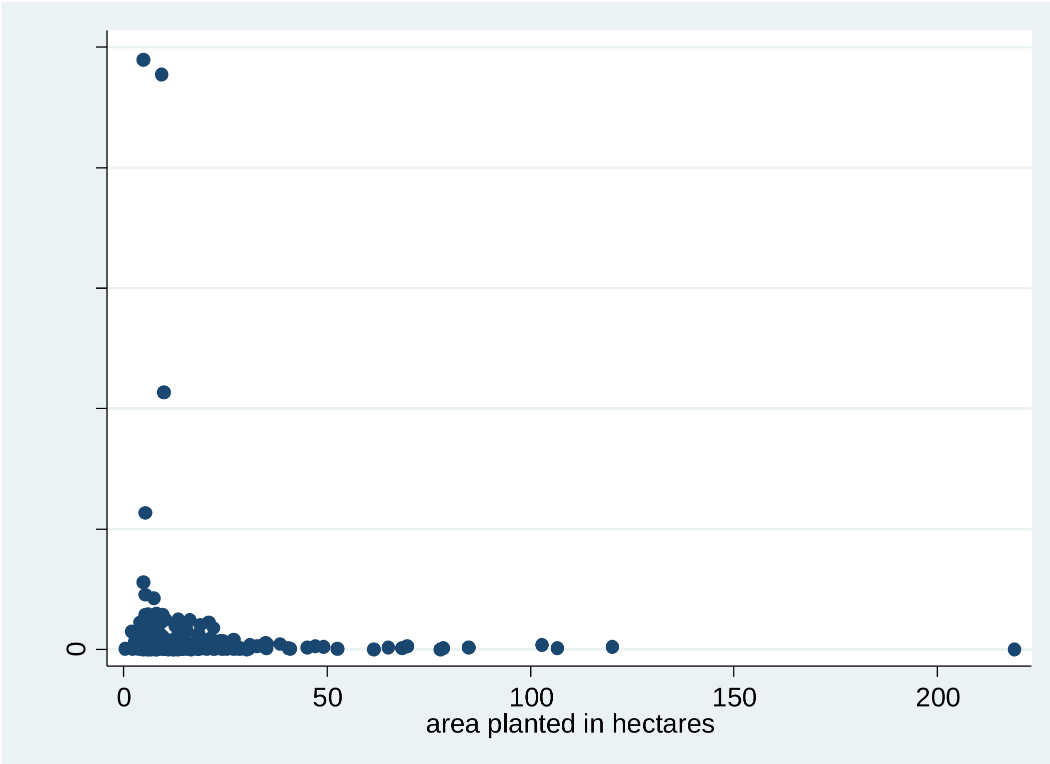

Figure 2-6: Sensitivity Analysis

3 ESSAY 2: JOINT LIABILITY LENDING WITH CORRELATED RISK

Introduction

The advent of micro-lending undoubtedly brought relief to many small businesses especially in developing countries, where collateral security cannot be provided by prospective borrowers because poverty rates are very high. It is estimated that as at 31st December 2010, more than 205 million people had been reached with loans by microfinance institutions (MFIs), (Maes & Reed, 2012). Despite backlashes received by MFIs for benefiting off the poor and despite findings that microfinance is probably not as miraculous as we would be made to believe (Banerjee, Duflo, Glennerster, & Kinnan, 2015), the facility remains prevalent in developing countries even in recent times.

In addition to serving as a source of funds for existing and new businesses, microfinance theoretically has been shown can aid some economies escape poverty traps in an occupational choice framework (Ahlin & Jiang, 2008). Reports from the mix market[11] indicate that about 3,652 MFIs are registered, and are involved in the financial intermediation, and financial services subsector of the economies of developing countries. With free entry and free exit that exists in the micro-lending market, successes chalked by existing MFIs have attracted potential firms both for-profit and the initial[12] not-for-profit firms that focus mainly on outreach and maximization of borrower welfare. Particularly impressive about the performance of MFIs is the repayment rates that have gone up considerably given that, the loans are advanced to poor people perceived to have good projects but have neither capital nor collateral to secure loans from traditional banks.

[11] This is a platform where registered microfinance institutions record their data and exchange information.
[12] Originally, MFIs were mainly operated by not-for profit organizations.

Microfinance contracts began as a contract between a borrower and a lender; where the individual borrower was solely responsible for his debt (individual liability). Other strategies such as dynamic incentive; where borrowers are promised a lower future interest rate or a higher loan amount in the future are also being used. (Karlan & Zinman, 2009) study a large MFI in South Africa that uses the dynamic incentive strategy. Group-based joint liability lending methods have also been used and remain common in the micro-lending markets, (de Quidt, Fetzer, & Ghatak, 2016). However, for-profit institutions are found to use more of individual lending while not-for-profit institutions use group-based lending methods (Cull, Demirgüç-Kunt, & Morduch, 2009; de Quidt et al., 2016).

Joint liability lending involves borrowers forming groups to access loans. The requirements are that, the liability is jointly held so that, successful group members are liable for part or the entire debt of an unsuccessful group member. The success of joint liability lending in the adverse selection setting, rests at least theoretically, (Ghatak, 1999, 2000) on the facts that borrowers are allowed to choose their own partners or group members, and secondly on the fact that, the loan is given on joint liability terms. The idea is that, in a village setting for instance, where his or her neighbor knows almost everyone, borrowers have local information which is not available to the lender and which makes it possible for safe borrowers to partner with other safe borrowers. Thus, a joint liability lending contract offers an implicit or hidden discount to safe borrowers and charges risky borrowers an implicit or hidden premium. Conditional on success, a safe borrower pays less in expectation than a risky borrower does in a joint liability contract due to having safer partners.

Although operations of MFIs are seen to be with small non-agricultural businesses, some loans advanced to agricultural workers are also granted on joint liability basis by rural and agricultural development banks. This raises concerns about the ability of joint liability lending to help in pricing for risk, or help in improving repayment rates if project returns of borrowers are correlated. This paper explores this scenario. As (Besley, 1994) puts it "A special feature of agriculture which provides the income of most rural residents is the risk of income shocks. These include weather fluctuations that affect all the producers of a particular commodity. Such shocks affect the operation of credit markets if they create the potential for group of farmers to default at the same time". To be able to price for risk and improve repayments rates, lenders should be able learn from the repayment behavior of borrowers. But as stated in (Ahlin & Waters, 2016; Ghatak, 2000), correlated risk influences how information is revealed to the Bank. For example, consider the case of perfect spatial correlation, then a joint liability contract for a group of two people will provides the lender with only one independent information about a borrower. This makes the group contract not superior to an individual liability contract.

This paper derives and characterizes the optimal lending contracts of joint liability lending when risks are correlated under adverse selection. This problem is very important both empirically and theoretically. Empirically, correlated risk is a pervasive reality and thus worth exploring. Theoretically, it would be interesting to see how the correlated risk affects group lending's ability to positively impact credit markets. (Ahlin, 2009) delves into correlated risks. It puts a simple structure on the correlation to understand how groups are formed and finds that borrowers anti-diversify risk in group formation so as to lower the occurrence of having to carry the debt burden of a group member. This anti-diversification strategy by borrowers, lays credence to how important it is to understand further, how the optimal lending contract and parameter space for efficiency are altered in the presence of correlation. I compare results from this work to the efficiency outcomes under independent risks. I additionally consider a common instance where the lender is faced with a pool of borrowers with a fraction having correlated risks and the other fraction having independent risks.

I find that correlation reduces the parameter space for fully efficient lending using joint liability contracts relative to independent risks case. The lender cannot effectively rely on the joint liability to improve risk-pricing under correlation. This is partly because of the monotonicity constraint, which prevents the lender from increasing the joint liability beyond the gross interest rate, and partly because of affordability constraints. I also find that, it may be better in some cases for lenders to serve borrowers separately, if the pool of borrowers consists of a fraction with correlated risks and a fraction with independent risks. This may help explain existence of specialized microfinance institutions such as agricultural banks separately from standard microfinance institutions. Separation of these banks could stem from fear on the part of lenders that they may have loans advanced to correlated borrowers becoming bad debts. This has left the financing of correlated risk borrowers' projects in the hands of governments and non-governmental organizations in cases where collateral is not available on the part of borrowers.

(Besley, 1994; IFC, 2012; Ramana, 2004) are a few of the papers that acknowledge the fact that covariant risk make agricultural financing unattractive to lenders. Ramana (2004) report why MFIs do not lend to farmers citing the Gramen Bank and Unit Desa system of the Bank of Rakyat Indonesia as banks that have focused exclusively on rural areas but not on agricultural lending in contrast with the Bank of Agriculture and Agricultural Cooperatives (BAAC) of Thailand which focuses exclusively on lending to agricultural workers and none to nonfarm activities.

The rest of the paper is organized as follows; the next section discusses some related literature. Section 3.3, outlines the basic models. In sections 3.4 and 3.5, I discuss independent risks and how I introduce correlation into the model. Optimal group lending contracts with correlated risks are derived in sections 3.6 and 3.7. Section 3.8 expands on this theory by dividing the pool of borrowers into correlated borrowers and uncorrelated borrowers. Section 3.9 concludes while the appendix contains proofs of propositions.

Literature Review

In the absence of collateral security from borrowers, micro-credit lenders have resorted to strategies such as joint liability lending. They seek to price for risk and induce quick repayment of loans. (Ghatak, 1999; Tassel, 1999), show group lending can improve efficiency of the credit market compared to traditional individual loans. Both papers were under adverse selection framework. A plethora of the literature on joint liability lending in the framework of adverse selection has focused on how the method can perform better than the individual loan contracts. The works of (Ghatak, 1999, 2000; Tassel, 1999) have all been in that direction. (Bhole & Ogden, 2010) as well as (de Quidt et al., 2016) study individual lending versus group lending but in a strategic default setting. They find welfare of borrowers to be higher with group lending than individual lending under certain conditions. One of such conditions is when penalty is allowed to be different among members of a group. Under the context of moral hazard (Chowdhury, 2007) shows that, if loans are not advanced to group members sequentially, group lending is not any better than individual lending[13].

[13] (Chowdhury, 2007; de Quidt et al., 2016) papers look at dynamic lending

The very early works by (Varian, 1990) and (Stiglitz, 1990) have investigated the potency of joint liability in harnessing local information to induce high repayment rates. (Varian, 1990) proposes a model in which the bank does its own screening, and does not rely on local information among borrowers. Banks interviewed group members and the eligibility or otherwise of a group member determined the fate of the other partners. (Besley & Coate, 1995) have also looked at how joint liability affects the willingness to pay on the part of borrowers. Even in the situation where borrowers have imperfect knowledge about the project types of their group members, (Armenda, Aghion, & Gollier, 2000) show that joint liability can lead to lower interest rate and help overcome some credit market inefficiencies.

Much earlier, even before the seminal contributions of (Ghatak, 1999, 2000), (Besley, 1994), identifies three major things that make rural credit markets in developing countries different from those in developed countries. These he highlights as collateral security unavailability, covariant risk and under developed states of related institutions. Not many papers to the best of my knowledge have looked at the optimal lending contract of Joint liability lending when there are correlated risks. (Ahlin & Waters, 2016; Ghatak, 2000) talked about the adverse effects of correlated risks on group lending and dynamic lending but only at the level of conjecture.

(Ahlin & Townsend, 2007b) tested for repayment implications, using data from Thai borrowing groups. They found that "a higher correlation of output can raise or lower repayment, depending on the model". The papers which discuss correlated risks in some detail as this work seeks to do are (Ahlin, 2009; Ahlin & Townsend, 2007a; Katzur & Lensink, 2012). (Ahlin, 2009) focused on matching[14], which distinguishes it from this paper. (Katzur & Lensink, 2012) study group lending with correlated project returns, which is the closest to this paper. They show that positive correlation can improve efficiency of group lending contracts. Their results require that the correlation between safe borrowers is sufficiently higher relative to the correlation between risky borrowers in a two-person group. I introduce and study correlation more generally and uniformly across risk types in this paper.

Baseline Model

3.1.1 Economic environment

The environment and baseline models follow a simple credit market and group lending model in (Ahlin & Waters, 2016; Ghatak, 1999). Assume there is a continuum of agents that are risk neutral and has measure one. Each agent is endowed with a unit of labor. Agents are endowed with a project that requires one unit of capital and one unit of labor. However, agents have no endowment of capital. Agents' projects are known to differ in risk type $p \epsilon [\underline{p}, 1)$ although there is an outside option that yields exogenously given net return of $\overline{u} \geq 0$. Agents would need to borrow a unit of capital to start their projects since they have no initial wealth.

The project of a type p agent pays R_p with probability p and pays 0 otherwise. Assume the riskiness of the project is private information known to the agent but not the lender. As in (Stiglitz & Weiss, 1981), I assume that all projects have the same expected return $\overline{R} = R_p \cdot p$, $\forall\, p \epsilon [\underline{p}, 1)$. This means that risky projects pay more when an agent succeeds. I also assume

[14] Ahlin (2009) found evidence of homogenous sorting by risk, and risk anti-diversifying strategy among group members

limited liability, which means agents who are unsuccessful owe nothing to the lender. Output can be verified as being either successful or failed but the lender cannot verify various shades of success. The assumption of limited liability and costly verification of output makes debt contracts the only feasible contracts. Borrowers who are able to repay their debts do, and so, there are no enforcement problems. There is a single lender who is risk neutral and would be willing to lend provided it earns an expected return of ρ where ρ, is the opportunity cost of capital per loan. The lender or bank knows the distribution of borrowers but not their probabilities of success.

I assume $\overline{R} > \rho + \overline{u}$, which makes all projects have higher expected return than costs of capital and labor invested. With social surplus strictly increasing in the number of projects funded, fully efficient[15] market would lend to all agents.

Suppose there two[16] types of agents $p \in \{p_r, p_s\}$ where $0 < p_r < p_s < 1$ and $R_s < R_r$.

Let $\theta \in (0, 1)$ be the population of risky borrowers and let the population average of any function $g(p)$ be denoted by $\overline{g(p)} = \theta g(p_r) + (1 - \theta) g(p_s)$. Similarly, $\overline{p}$ denotes the mean risk-type and $\overline{p^2}$ the mean squared-type.

[15] Fully efficient lending goes hand-in-hand with maximal outreach because, from our assumptions, social surplus is strictly increasing in the number of projects funded.
[16] The model can be generalized to more than two types.

3.1.2 Individual Lending in a Static Environment

In this section, I review the results under individual lending and also group lending before introducing correlation so that results can be compared. Under full information where agents' types are known, the lender can price for risk by charging specific interest rates such that the following equation holds;

$p_\tau r_\tau = \rho \implies r_\tau = \frac{\rho}{p_\tau}$ for each risk type τ. $r_\tau(r)$ denotes the loan amount plus interest. This is the first best outcome which can be seen to be efficient and equitable with all surpluses accruing to borrowers. Now, for the case where agent's risk types are unknown to the lender,

Let $\overline{p} = \theta p_r + (1 - \theta)p_s$ be the average success probability and r the repayment amount. An agent of type $\tau \in \{r, s\}$, will borrow to carry out the project if and only if,

$\overline{R} - p_\tau r \geq \bar{u} \Leftrightarrow r \leq \hat{r}_\tau \equiv \frac{\bar{R}-\bar{u}}{p_\tau}$. The first inequality says that the expected returns from the project less the expected repayment amount should exceed the outside option's return. The second inequality follows from a rearrangement of the first and gives us a reservation interest rate $\hat{r}_\tau$, above which an agent of type τ will choose the outside option instead. The reservation interest rate is lower for the safe borrower, and thus the safe borrower is harder to attract. If safe borrowers borrow, then so will the risky ones since safe borrowers succeed more often and repay with a higher probability. Thus $r \leq \hat{r}_s = \frac{\bar{R}-\bar{u}}{p_s}$ is a necessary condition for fully efficient lending.

A sufficient condition for both types to find it affordable to borrow is $r \leq R_s$.

Now, the lender would be willing to give out a loan if $\overline{p}r = \rho$ thus the break-even interest rate is $r = \frac{\rho}{\bar{p}}$

To attract all borrowers, we need $\frac{\rho}{\bar{p}} \leq \frac{\bar{R}-\bar{u}}{p_s}$ ie $N \geq \frac{p_s}{\bar{p}}$ $where\ N = \frac{\bar{R}-\bar{u}}{\rho}$ and for affordability as before,

$r \leq R_s$ or $\frac{\rho}{\bar{p}} \leq \frac{\bar{R}}{p_s}$ ie $\frac{p_s}{\bar{p}} \leq G$ where $G = \frac{\bar{R}}{\rho}$

As described in Ahlin and Waters (2014), $N = \frac{\bar{R}-\bar{u}}{\rho}$ is the net excess return to capital in this market and $G = \frac{\bar{R}}{\rho}$ is the gross excess return to capital in this market. Thus, efficient lending is achieved if the net excess return to capital in this market is larger than the extent of asymmetric information, represented by $\frac{p_s}{\bar{p}}$. There is however, a second best option, which involves lending to only risky agents. In which case there is inefficiency (in the sense that maximal outreach is not attained) as the lender is unable to price for risk to attract safe borrowers.

Group Lending with independent risks

Here also, I review very quickly the model of joint liability lending presented in (Ghatak, 2000) and results from (Gangopadhyay, Ghatak, & Lensink, 2005). Relying on the environment discussed earlier, a joint liability contract requires a borrower to pay a joint liability say c, in addition to the repayment amount r on her own loan if the group[17] member fails while she

[17]We assume a group is of size two throughout this work.

succeeds. Here, it is assumed that agents know each other's type but this is unknown to the lender. (Ghatak, 2000) shows that borrowers would form groups homogenously based on risk type in equilibrium. For a borrower of type $\tau \in \{r, s\}$, the expected payoff under homogenous matching is given as

$\bar{R} - p_\tau r - p_\tau(1 - p_\tau)c = \bar{R} - p_\tau[r + (1 - p_\tau)c\,]$. The effective interest rate $r + (1 - p_\tau)c$ is seen to vary positively with risk of the borrower and penalizes risky borrowers as full information similarly does, although the lender has no information on risk types. It must be noted that, a contract may attract only risky borrowers, both risky and safe, or none. In addition, under any contract, the risky borrower earns more than the safe borrower (it can be seen by examining the payoffs). Thus for group lending to achieve full efficiency, I maximize the safe borrower's payoff subject to lender's zero profit, affordability and monotonicity constraints, and assuming homogenous matching obtains.

Maximize, $\bar{R} - p_s r - p_s(1 - p_s)c$ subject to

$c \leq r$, $R_s \geq r + c$ and $\bar{p}r + \overline{p(1 - p)}c \geq \rho$

Before solving for the contracts, some key observations can be made. It can be seen that, raising c and lowering r along the lender's isoprofit curve, would raise the safe borrowers payoff. This is because the safer borrower's indifference curve in the (r, c) space has a larger slope in magnitude relative to the bank's iso-profit curve.[18] What it means is that, a higher c relative to r would put more burden on the states of the world where there is a failure and thus on risky borrowers. But since c cannot exceed r, the best contract to attract safe borrowers is that with

[18] -1/$(1 - p_s)$ versus $-\bar{p}/\overline{p(1 - p)}$

$c = r$, where full affordability is affordable and the maximum affordable level of c if full liability is not affordable.

Solving the maximization problem, we have the best-for-safe contract as

$c = r = \frac{\rho}{p(2-p)}$ assuming affordability is not an issue.

On the other hand, when $r = c$ is not affordable, the optimal contract is derived in as;

$(r, c) = \left\{\rho \frac{p_s - \overline{p(1-p)}G}{p_s\overline{p^2}}, \rho \frac{\bar{p}G - p_s}{p_s\overline{p^2}}\right\}$. The results are thus summarized in the proposition below

Proposition 0: Under the assumptions of the model, a group contract that maximizes borrower surplus subject to homogeneous matching, borrower limited liability, lender break even and monotonicity achieves full efficiency, if and only if

$$N \geq \begin{cases} B_1 - \frac{B_1 - B_2^*}{C_2^* - C_1}[G - C_1] & G \in [C_1, C_2^*] \\ B_2^* & G \geq C_2^* \end{cases}$$

Where $C_1 = B_1 = \frac{p_s}{\bar{p}}$, $B_2^* = \frac{p_s(2-p_s)}{\overline{p(2-p)}}$ and $C_2^* = \frac{2p_s}{\overline{p(2-p)}}$

Otherwise, only risky agents borrow. This is the independent risk case results and we shall compare it to the correlation case in the sections to follow. Essentially, we can see from the above results that, as G increases away from C_1, the net returns that is required for efficient lending decreases also linearly in B_1 and gets to a minimum at B_2^*. The dependence on G illustrates the affordability consideration in deriving the contracts. We cannot go below the floor of B_2^* because the contract cannot have c greater than r, and indication of the limitation group lending has in improving risk-pricing. See (Ahlin & Waters, 2016) for a full discussion and comparison with dynamic individual lending.

Group Lending with Correlated Risk

I introduce correlation of risk types into the group lending model in a manner that preserves the individual probabilities of success. Suppose there are two borrowers i and j, with probabilities of success $p_i \ and \ p_j$ respectively. The joint distribution is given in the Table 3-1.

The table presents the unique way to introduce correlation while preserving success probabilities given $p_i \ and \ p_j$. It can be verified that the columns and rows add up to equal the individual probabilities of success. $\varepsilon = 0$ corresponds to the no-correlation case. I focus on positive correlation only. (Katzur & Lensink, 2012), show that negative correlation between risky borrowers may lead to a breakdown of positive assortative matching for some of the first-best contracts. I maintain the assumption that borrower types are unknown to the lender. In addition, I assume here that the lender knows project returns are correlated and knows the joint distribution of the probabilities of success. A positive ε adds to the probabilities that both either succeed together or fail together and subtracts from the probabilities that an agent succeeds and the group member fails.

Although the table depicts a constant mass ε being added to or subtracted from the cells, ε may depend on the probabilities of success. There are two ways one can symmetrically introduce correlation across all types of projects. In the first case, herein after called the 'constant-mass case' a constant amount of mass is added to same-outcome group events. The probability of group members having the same outcome is increased by a constant mass relative to the independent risk case. This also now means that the probability of group members realizing different outcomes is lowered by the amount of the constant mass as compared to when correlation is zero. Thus in this constant mass case, $\varepsilon(p_i, p_j) = \varepsilon > 0$ for all $p_i \ and \ p_j$. The second case involves not adding a constant mass but scaling a mass so that, homogeneous projects have the same correlation coefficient. I call this case the 'constant-correlation' case. In this second case therefore, $\varepsilon(p_i, p_j)$>0 is not constant but the correlation is constant for all $p_i\ and\ p_j$. Essentially both cases are doing the same thing by introducing correlation symmetrically although the approaches differ as described above.

The result in (Katzur & Lensink, 2012) that group lending contracts may have better efficiency outcomes, hinges on the assumption that the correlation between safe projects is sufficiently larger than the correlation between risky projects. Thus correlation is not symmetrically introduced in that paper. Whereas homogeneous projects have the same correlation coefficient in our second case, it is higher for safe projects under (Katzur & Lensink, 2012).

Constant Mass Case (Case 1)

Suppose $\varepsilon(p_i, p_j) = \varepsilon > 0$ for all $p_i\ and\ p_j$

Given the requirement that elements in the cells in the chart above, must be less than one and greater than zero, it follows that, assuming $\varepsilon \leq p_r(1 - p_s)$ is sufficient to ensure this. The results on the independent risk types hinges on the homogeneous matching results derived in (Ghatak, 1999). With the introduction of correlation in project returns, I show that the homogenous matching result still holds under case 1.

Lemma 1

Under the assumptions in the model, and in the presence of correlated risks, homogeneous matching is the only equilibrium.

By examining the joint distribution, it can be seen that

Observation 1

Both types of borrowers in the presence of correlated risks pay the joint liability less often than under independent risks. This is because they either both succeed or both fail with a higher probability compared to independent risks case.

As mentioned already, given the assumptions in the model, a joint liability contract that attracts safe borrowers would also attract risky ones simply because the safe borrowers succeed more often and repay the joint liability with a higher probability. As such, to investigate the feasibility of fully efficient lending, I zero in on what can attract the safe borrower into accepting the contract. It follows that, if full liability is affordable, the best contract should have $c = r$ and extract the maximum possible if $c = r$ is not affordable. More formally, I rely on the following lemma to derive the condition under which fully efficient lending with borrowers having all surpluses is attainable.

Lemma 2

The risky borrower earns more than the safe borrower in any joint liability contract with $c \leq r$. This can be shown by comparing payoffs of the safe and risky borrowers.

This also means that a contract that attracts the safe borrower also attracts the risky borrower.

By Lemma 2, the contract attracts either both safe and risky borrowers or only risky borrowers or none. To attract only risky borrowers, the lender can just set $r = \frac{\rho}{p_r}$. To derive a necessary and sufficient condition under which fully efficient lending is attainable, I maximize a safe borrower's payoff

$\bar{R} - p_s r - [p_s(1-p_s) - \varepsilon\ ]c$ subject to the following constraints;

1. $0 \leq c \leq r$ Monotonicity
2. $\bar{p}r + \overline{p(1-p)}c - \varepsilon c \geq \rho$ Zero profit constraint
3. $r + c \leq R_s$ Limited Liability

Using the first two constraints to solve for the case where affordability is not an issue, we get

$r = c = \frac{\rho}{\overline{p(2-p)} - \varepsilon}$, granted $p_s p_r > \varepsilon$.[19] The remaining limited liability constraint holds if

$G \geq \frac{2p_s}{\overline{p(2-p)} - \varepsilon}$

When affordability is an issue (when $\frac{p_s}{\bar{p}} \leq G < \frac{2p_s}{\overline{p(2-p)} - \varepsilon}$), we maximize the safe borrower's payoff subject to the following constraints

$R_s \geq r + c$ and $\bar{p}r + \overline{p(1-p)}c - \varepsilon c \geq \rho.$ The contract can be derived as

$(r, c) = \left\{\rho\frac{p_s - [\overline{p(1-p)} - \varepsilon]G}{p_s(\overline{p^2} + \varepsilon)}, \rho\frac{\bar{p}G - p_s}{p_s(\overline{p^2} + \varepsilon)}\right\}$. Clearly, monotonicity constraint is satisfied for $\frac{p_s}{\bar{p}} \leq G < \frac{2p_s}{\overline{p(2-p)} - \varepsilon}$

[19] $p_s p_r > \varepsilon$ is guaranteed by a standard assumption of $p_s + p_r > 1$ made in the literature and our earlier assumption that $\varepsilon < p_r(1-p_s)$

Proposition 1

Under the assumptions in the model, a group contract with joint liability that maximizes borrower surplus subject to the following conditions; homogenous matching, limited liability on borrower, lender break-even and monotonicity achieves full efficiency if and only if

$$N \geq \begin{cases} B_1 - \frac{B_1 - B_2}{C_2 - C_1}[G - C_1] & G \in [C_1, C_2] \\ B_2 & G \geq C_2 \end{cases}$$

Where C_1[20] $= B_1 = \frac{p_s}{\bar{p}}$, $B_2 = \frac{p_s(2-p_s)-\varepsilon}{\overline{p(2-p)}-\varepsilon}$ and $C_2 = \frac{2p_s}{\overline{p(2-p)}-\varepsilon}$

Otherwise, only risky agents borrow.

Comparing the results here to those under independent risk, we can see that, the level of r and c charged when full liability is affordable are higher under correlation.

It is important to note that, under correlation, c plays lesser role in risk-pricing as group members are more likely to succeed or fail together. Bailouts occur less frequently under correlation. So the extent to which the lender can rely on the joint liability component of the contract to price for risk is curtailed in the correlation case.

The effective interest rate under this contract can be written as

$\hat{r}_i = r + \left[(1-p_i) - \frac{\varepsilon}{p_i}\right]c$. It can be shown that this decreases in p_i by taking the derivative. Thus, the effective interest rate under correlation also offers a discount to safe borrowers. I reserve the discussion of the intuition of these results until proposition 3.

[20] The subscript (1) means contract is for 1 borrower.

Altogether, the net returns that is required for efficient lending is higher than under independent risk case. This essentially means that correlation reduces the parameter space for fully efficient lending and makes joint liability less efficient relative to independent risk case.

Proposition 2

The parameter space over which joint liability lending achieves full efficiency, reduces as the correlation, represented by a constant mass of ε between homogenous projects increases.

Constant-correlation Case (Case 2)

In this section, I put some structure on ε so that it is not constant but depends on the probabilities of success. I demonstrate that results obtained under the constant-mass case can be replicated using the constant correlation method of symmetrically introducing correlation.

Define $\varepsilon(p_i, p_j) = \tilde{v} * \min\{p_i(1-p_j), p_j(1-p_i)\}$. This enables the distribution of projects of homogeneous groups to have the same correlation coefficient. Assume $\tilde{v} \in [0,1]$. $\tilde{v}$ is a fraction of the maximum possible correlation between two projects. Thus with homogenous group, $\tilde{v}$ is just the correlation coefficient. See Appendix for details.

Lemma 3

Under the constant correlation case, homogeneous matching is the only equilibrium.

Lemma 4

Under the constant correlation coefficient case, the risky borrower earns more than the safe borrower does. The proof follows similarly as Lemma 1 and by noting that $0 \leq \tilde{v} \leq 1$.

Using Lemma 4, the contracts are similarly derived from the maximization problem but replacing constant ε with its constant correlation form in the objective function and constraints.

When affordability is not an issue, the contract can be derived as

$r = c = \frac{\rho}{[\bar{p}+(1-\tilde{v})\overline{p(1-p)}]}$ and when affordability is an issue, we have the contract as

$$(r, c) = \left\{\rho \frac{p_s - [(1-\tilde{v})\overline{p(1-p)}]G}{p_s[\bar{p}-(1-\tilde{v})\overline{p(1-p)}]}, \rho \frac{\bar{p}G - p_s}{p_s[\bar{p}-(1-\tilde{v})\overline{p(1-p)}]}\right\}$$

Proposition 3

Under the assumptions in the model, a joint liability contract that maximizes borrower surplus subject to the following conditions; homogenous matching, limited liability, lender breaking even and monotonicity achieves full efficiency if and only if

$$N \geq \begin{cases} B_1 - \frac{B_1 - B_2'}{C_2' - C_1}[G - C_1] & G \in [C_1, C'_2] \\ B'_2 & G \geq C'_2 \end{cases}$$

Where

$C_1 = B_1 = \frac{p_s}{\bar{p}}$, $B'_2 = \frac{[p_s + p_s(1-p_s)(1-\tilde{v})]}{[\bar{p}+(1-\tilde{v})\overline{p(1-p)}]}$ and $C'_2 = \frac{2p_s}{[\bar{p}+(1-\tilde{v})\overline{p(1-p)}]}$ and $\tilde{v}$ is as defined before.

Otherwise, only risky agents borrow.

The prediction from the constant-mass correlation parameter ε is not different from the constant correlation case. In subsequent derivations, I focus on constant-correlation case. It can be seen that the region for fully efficient lending narrowing as correlation increases. To get an insight on this result, I interrogate the effective interest rate in the presence of correlation. In the face of correlation, the effective interest rate can be written as $\hat{r}_i = r + (1 - p_i)(1 - \tilde{v})c$. Setting $\tilde{v} = 0$ yields that of the independent risk case. When correlation is very high ($\tilde{v}$ approaching 1), the joint liability component becomes very small. The relationship between the effective interest and the correlation parameter $\tilde{v}$ as well as p_i is the same as obtained in proposition 1.

The effective interest rate differential between safe and risky borrowers is given by $(p_s - p_r)(1 - \tilde{v})c$. Recall that the effective interest rate differential is the implicit discount to safe borrowers and implicit premium charged to risky borrowers that makes the lender able to improve risk-pricing and the efficiency of credit markets using group lending. Let $c' = (1 - \tilde{v})c$ be the effective liability. c' decreases as correlation $\tilde{v}$ increases, if c does not change. As $\tilde{v}$ goes up, we would like to raise c to compensate for the reduction in c'so that c'remains the same or unchanged. Nevertheless, c' cannot be restored to its initial value because c does not rise enough. To show this,

First, consider the case where affordability is not an issue. The binding constraints are

$c \leq r$ (Monotonicity) and $\bar{p}r + \overline{p(1-p)}(1 - \tilde{v})c \geq \rho$ (zero profit constraint).

As correlation $\tilde{v}$ increases, the effective liability c' and the $\overline{p(1-p)}(1-\tilde{v})c$ component of the zero profit constraint reduce if c does not change. The lender would need to raise revenue to restore the effective liability and hence the zero profit constraint to their initial levels. This could be done by increasing c sufficiently to counter the reduction in c' but this cannot be achieved without violating the binding monotonicity constraint which says c cannot exceed r. This suggests that r and c would need to be raised simultaneously. We can think about increasing r to make way for c to increase sufficiently by making strictly positive profits[21] but this only makes it harder to attract safe borrowers. Therefore, raising r and c simultaneously, without violating both the monotonicity and zero profit constraints means c cannot rise enough to restore the original effective interest rate differential (or the effective liability). Unchanged profits but a reduced implicit discount for safe borrowers makes it harder to attract safe borrowers-- thus reducing the parameter space for fully efficient lending under joint liability.

Second, consider the case where affordability is an issue. The binding constraints are the zero profit constraint and the affordability constraint, that is, $\bar{p}r + \overline{p(1-p)}(1-\tilde{v})c \geq \rho$ and $r + c \leq R_s$ respectively. Again, as correlation $\tilde{v}$ increases, $\overline{p(1-p)}(1-\tilde{v})c$ component of the zero profit constraint decreases if c does not change and the lender would need to raise revenue. This can be done by increasing c or r or both to keep the lender's zero profit constraint binding. However, the binding affordability constraint ($r + c \leq R_s$), means c and r cannot increase or decrease at the same time to achieve the intended purpose without violating the binding affordability constraint. c and r must move in opposite directions (since their sum is a constant).

[21] An increase in c sufficiently would be able to restore the decline in revenue as a result of the increase in correlation but this comes with an increase in r which would mean there is excess revenue than needed for a binding zero profit constraint. This then means the zero profit constraint holds with strict inequality. The lender makes strictly positive profit from lending since every project when successful faces a higher r and c than those needed for a zero profit.

Thus, the lender can raise revenue only by raising r and lowering c [22]. The drop in c further reduces the effective interest rate differential and hence the implicit discount to safe borrowers. Again, unchanged profit but a reduced implicit discount for safe borrowers makes it harder to attract safe borrowers—thus reducing the parameter space for fully efficient lending under joint liability.

Figures 3-5 in the Appendix shows the relationship between the joint liability c, and correlation $\tilde{v}$. It can be seen that, c increases initially up to a point and then begins to fall. To summarize the results, we can show the following proposition.

Proposition 4

Under the constant correlation case, the parameter space over which joint liability lending achieves full efficiency reduces as the correlation ($\tilde{v}$) increases.

Figure 3-1 shows the graph of the correlated risks case against independent risks case. The correlated risks case is the broken lines. It can be seen that it lies above the independent risk case and thus achieves full efficiency under a smaller parameter space. This is easy to see because the efficiency parameter space for the independent risk case is larger compared to the correlated risk case.

[22] The lender cannot rather increase c and lower r because r is repaid more often than c and it requires a larger raise in c to make-up for a small reduction in r. This can also be seen by noting that the slope of the affordability constraint in the (r, c) space has a smaller slope in magnitude relative to that of the bank's iso-profit curve. ie -1 versus $-\bar{p}/\overline{p(1-p)}(1-\tilde{v})$. Therefore, moving to a higher profit curve while staying on the affordability constraint requires raising r and lowering c.

Correlation in a Market of Mixed Borrowers

An interesting case with the introduction of correlation in the group lending strategy is a situation, where the lender is faced with a mixed pool of borrowers. An example would be a typical situation of a lender having to serve borrowers from a pool of farmers and a pool of small business owners. In this section, I look at a lender who is faced with these two categories of borrowers. It would be interesting to know whether the lender would serve these categories separately or would pool the borrowers together in offering the contract.

I assume borrowers come from either of two pools; A and B. All borrowers from pool A are correlated with all other borrowers in pool A. All borrowers from pool B are independent of all other borrowers. We can think of the lender being faced with a fraction k, of borrowers having correlated risks and a fraction (1-k) having independent risks. The correlation type (belonging to pool A or pool B), is independent of risk type (p_i). I study how this impacts the optimal contracts, efficiency and whether or not it is better for the lender to separate borrowers instead of pooling them.

In this section also, the bank or lender doesn't know the risk (p_i) types. The lender is assumed to know correlation type of the borrower (A or B) but the lender still allows voluntary group formation so that borrowers will feel buy-in.[23] I use the constant-correlation case throughout this section because it enables me plot graphs easily. In this new scenario, there can be four types of groups in equilibrium;

[23] For example, borrowers could have a disutility of having their free association restricted; sometimes people would like to form groups on religious or social lines to enable monitoring and enforce repayment. The lender may not want to meddle with group formation so that, borrowers would not blame their default on their inability to choose their group members. The lender could also be interested mainly in homogeneous matching on risk type not correlation type. We could also assume k>1/2 so that there is still some max {0, 2k-1} population of correlated borrowers left, even if lender were to dictate matching (by diversification). The analysis then proceeds similarly from there.

A^{safe} : Safe and Safe correlated group

A^{risky}: Risky and Risky correlated group

B^{safe}: Safe and Safe uncorrelated group

B^{risky} : Risky and Risky uncorrelated group

A formal proof of the matching pattern is in proposition 1 of Ahlin (2016) which I rephrase as Lemma 5 below.

Lemma 5

Under the assumptions in the model, in equilibrium, almost every group is homogeneous in both risk and correlation type.

It is important to note that, all two-person groups formed to apply for loans are in one of the groups listed above. Lemma 5 rules out a case where a group A member partners with a group B member. All such groups can be shown to have a measure zero in equilibrium.

3.1.3 Optimal Contracts

Lemma 6

The safe uncorrelated borrowers earn less than any other borrower in any joint liability contract with $c \leq r$.

I show this by comparing payoffs of the groups A^{safe}, A^{risky}, B^{safe} and B^{risky}above. This also means that a contract that attracts the B^{safe} borrower also attracts borrowers in other groups.

Safe borrowers cross-subsidize risky borrowers and in the presence of correlation, uncorrelated borrowers also cross-subsidize correlated borrowers. To derive the optimal contract therefore, I rely on Lemmas 5 and 6 and maximize the expected payoff of B^{safe} group subject to our usual constraints.

B^{safe} Group member's payoff is $\bar{R} - p_s r - [p_s(1-p_s)\,]c$

Subject to

$0 \leq c \leq r$, and $\bar{p}r + \overline{p(1-p)}(1-k\tilde{v})c \geq \rho$

Assume $\tilde{v} < \frac{\theta p_r(p_s - p_r)}{k\overline{p(1-p)}}$. Then, when affordability is not an issue, the contract can be derived as follows

$$r = c = \frac{\rho}{[\bar{p} + \overline{p(1-p)}(1-k\tilde{v})]}$$

Now, when affordability is an issue, (when $\frac{p_s}{\bar{p}} \leq G < \frac{2p_s}{[\bar{p} + \overline{p(1-p)}(1-k\tilde{v})]}$)

the following constraints are the relevant ones and are similar to section 3.6's derivations.

$$R_s \geq r + c$$

$$\bar{p}r + \overline{p(1-p)}(1-k\tilde{v})c \geq \rho$$

The contract is derived as $(r,c) = \left\{\rho \frac{p_s-[(1-\tilde{v}k)\overline{p(1-p)}]G}{p_s[\overline{p^2}+(\tilde{v}k)\overline{p(1-p)}]}, \rho \frac{\bar{p}G-p_s}{p_s[\overline{p^2}+(\tilde{v}k)\overline{p(1-p)}]}\right\}$

See appendix for details.

Using Lemma 6 and results from the maximization problem, the following can be shown

Proposition 5

Under the assumptions in section 5, a joint liability contract that maximizes borrower surplus subject to the following conditions; homogenous matching, limited liability on borrower, lender breaking even and monotonicity achieves full efficiency for $\tilde{v} < \frac{\theta p_r(p_s-p_r)}{k\overline{p(1-p)}}$ if and only if

$$N \geq \begin{cases} B_1 - \frac{B_1-B_2''}{C_2''-C_1}[G-C_1] & G \in [C_1, C''_2] \\ B''_2 & G \geq C''_2 \end{cases}$$

Where $C_1 = B_1 = \frac{p_s}{\bar{p}}$, $B"_2 = \frac{p_s(2-p_s)}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$ and $C"_2 = \frac{2p_s}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$

Otherwise, safe uncorrelated borrowers are excluded from the credit market.

An interesting observation from the analyses in mixed pool of borrowers' case in section 3.8.1 is that there exists non-monotonicity in the ability to reach a larger pool of borrowers in the presence of correlation. When correlation is zero, fully efficient lending is achieved over a larger parameter space than when correlation is positive. One would have thought that, for a fixed correlation between project returns, as the fraction of correlated borrowers increases from 0 towards 1, it would be easier to reach a mixed pool of borrowers than an all-correlated group (monotonicity). It turns out that, in some instances, it is easier to reach an all-correlated pool than a mixed pool of borrowers.

The non-monotonicity in the conditions for fully efficient lending with respect to k under this section, stems from the fact that, as k increases, there are fewer borrowers from the uncorrelated pool of borrowers, and it is harder to reach the few uncorrelated borrowers left. In addition, the uncorrelated borrowers become increasingly worse off as k increases since they cross subsidize the correlated borrowers. In such a case, separation would do better than pooling of the mixed pool of borrowers. When all borrowers are correlated, there are no safe, uncorrelated borrowers that need to be reached. Consider a case where $k = 1$ versus when $k = 1 - \delta$ for some small $\delta > 0$.

The aggregate cross-subsidy from safe borrowers to risky borrower, and from uncorrelated borrowers to correlated borrowers would be very similar under $k = 1$ and $k = 1 - \delta$ for small δ. The problem is that while for $k = 1$, there are no uncorrelated borrowers to reach, under $= 1 - \delta$ one worries about how to reach the small group of uncorrelated borrowers (δ) who are harder to attract because the safe-uncorrelated borrowers are the least well-off. Although this smaller group can be left unreached so that we have just a δ loss of efficiency, the difficulty in reaching them illustrates how fully efficient lending discontinuously becomes easier to attain as k reaches 1.

Proposition 6

Define $\bar{k}_a = \frac{\theta p_r(p_s - p_r)}{\overline{p(1-p)}[1+(1-p_s)(1-\tilde{v})]}$ and $\bar{k}_b = \frac{\theta p_r(p_s - p_r)}{\overline{p(1-p)}[1-(1-p_s)(1-\tilde{v})]}$. Clearly, $0 < \bar{k}_a < \bar{k}_b < 1$

Under assumptions in section 3.8, fully efficient lending can be achieved over a larger parameter space (G, N), by lender pooling borrowers if $k < \bar{k}_a$ and by lender separating borrowers if $k > \bar{k}_b$

3.1.4 Graphical presentation of propositions 6

We have seen in section 3.8.1 that when the conditions for proposition 5 do not hold, or when project net excess returns are low, the safe uncorrelated are not able to borrow and efficiency of the lending strategy is reduced. Following this, I explore in this section, whether we can have full efficiency by separating the banks (ie. separate into a lender that would lend to only the borrowers in Pool A and a lender focusing exclusively on Pool B borrowers). Using the values we use in plotting the figures, we can calculate $\bar{k}_a = 0.519$ and $\bar{k}_b = 0.574$

In Figure 3-2, we add the graph of the mixed pool of borrowers to Figure 3-1 it. For clarity, "B2" corresponds to the independent risk case. "B2p" corresponds to the correlated risk case, while "B2pp" corresponds to the mixed pool of borrowers' case. The mixed pool of borrowers' line lies in between the other two. As we can see from the figure, when net excess return to capital embedded in the project is low and lies between B2p and B2pp, it is better to pool borrowers and give one contract. This would be the way to achieve full efficiency. In this case, there is no point separating, otherwise we lose the safe correlated group.

In figure 4-3 where the fraction of correlated borrowers is 0.75, we see a situation that calls for separation of banks into those that would serve borrowers with correlated risk and those that would serve borrowers with independent risks for full efficiency. When net excess return to capital embedded in the project is low and lies between B2p and B2pp, it is better to separate borrowers and serve them separately. This would be the way to achieve full efficiency. In this case, there is no point pooling borrowers, otherwise we lose the safe uncorrelated group.

The need for separation of the pool of borrowers into those with correlated risk and those with independent risk for fully efficient lending may help explain and justify the existence of agricultural development banks separately from banks serving self-employed borrowers. (Besley, 1994; IFC, 2012; Ramana, 2004) all acknowledge the fact that covariant risk makes agricultural financing unattractive to microfinance firms. (Ramana, 2004) document how MFIs do not lend to farmers citing the Gramen Bank and Unit Desa system of the Bank of Rakyat Indonesia as banks that have focused exclusively on rural areas but not on agricultural lending as against the Bank of Agriculture and Agricultural Cooperatives (BAAC) of Thailand which focuses exclusively on lending to agricultural workers and none to nonfarm activities.

Conclusion

We derive and characterize the optimal lending contracts for group-based joint liability lending with correlated risks. With the introduction of correlation, we find that the parameter space for fully efficient lending under correlated risks is smaller compared to independent risks case. When correlation increases, the differential effective interest rate that offers implicit discount (charges implicit premium) to safe (risky) borrowers is reduced. This reduces the ability of the lender to use joint liability lending to price for risk and improve efficiency of credit markets.

When full affordability is feasible, the monotonicity constraint coupled with a binding lender zero-profit constraint, prevents the lender from increasing the joint liability component enough to maintain effective interest rate differential that could have helped in risk-pricing. On the other hand, when affordability is an issue, the affordability constraint requires that the interest rate be increased and the joint liability component be reduced in response to greater correlation. Reduction in the joint liability further reduces the effective differential interest rate. unchanged profits but a reduced discount makes it difficult to attract safe borrowers. Correlation thus reduces the effectiveness of group lending and especially when affordability is low. It can lead to the exclusion of some potential borrowers such as the safe uncorrelated from the market thereby reducing efficiency.

An interesting observation from the analyses is the existence of non-monotonicity in the ability to reach a larger pool of borrowers in the presence of correlation. One would have thought that, for a fixed correlation between project returns, as the fraction of borrowers with correlated risk increases from 0 towards 1, it would be easier to reach a mixed pool of borrowers than an all-correlated group. It turns out that, in some instances, it is easier to reach an all-correlated group than a mixed group of borrowers.

The non-monotonicity stems from the fact that, as k increases, there are fewer borrowers from the uncorrelated pool of borrowers and it becomes harder to reach the few uncorrelated borrowers. In such a case, separation would do better than pooling of the mixed pool of borrowers. In addition, the uncorrelated borrowers become increasingly worse off as k increases since they cross subsidize the correlated borrowers. When all borrowers are correlated, there are no safe, uncorrelated borrowers that need to be reached.

The results show that under certain conditions, such as situations where project returns are low and the fraction of correlated borrowers is high, full efficiency requires that the banks for correlated borrowers and those for non-correlated borrowers be separated. This may help explain the existence of agricultural development banks separately from banks serving self-employed non-agricultural borrower.

APPENDICES

Appendix A: Tables and Figures for Essay 2

Table 3-1: Joint Output Distribution in the Presence of Correlation

	j Succeeds (p_j)	j Fails (1-p_j)
i succeeds (p_i)	$p_i p_j + \varepsilon$	$p_i(1-p_j) - \varepsilon$
i Fails (1- p_i)	$(1-p_i)p_j - \varepsilon$	(1-p_i)$(1-p_j) + \varepsilon$

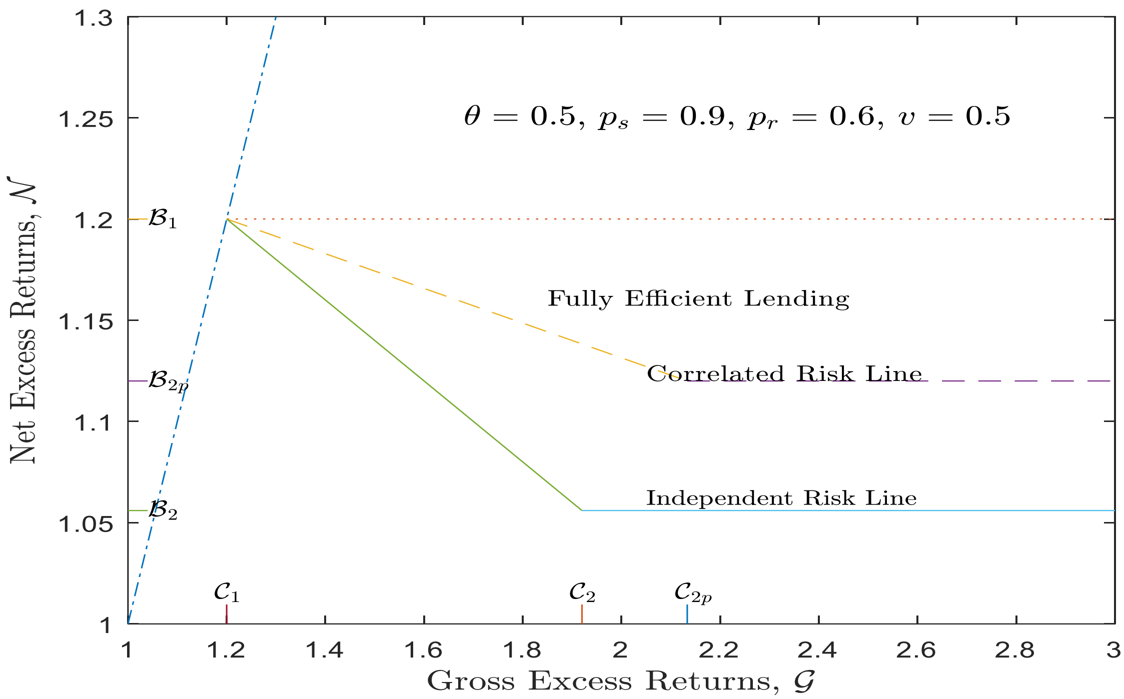

Figure 3-1: Correlated Borrowers

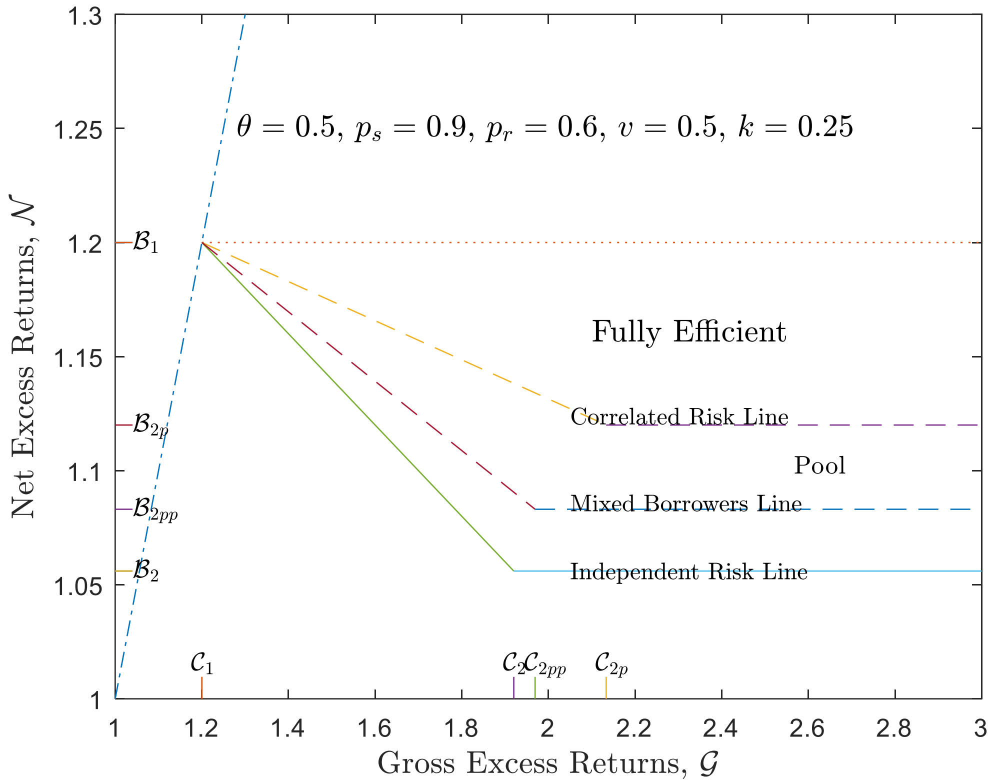

Figure 3-2: Mixed Pool of Borrowers for Lower k

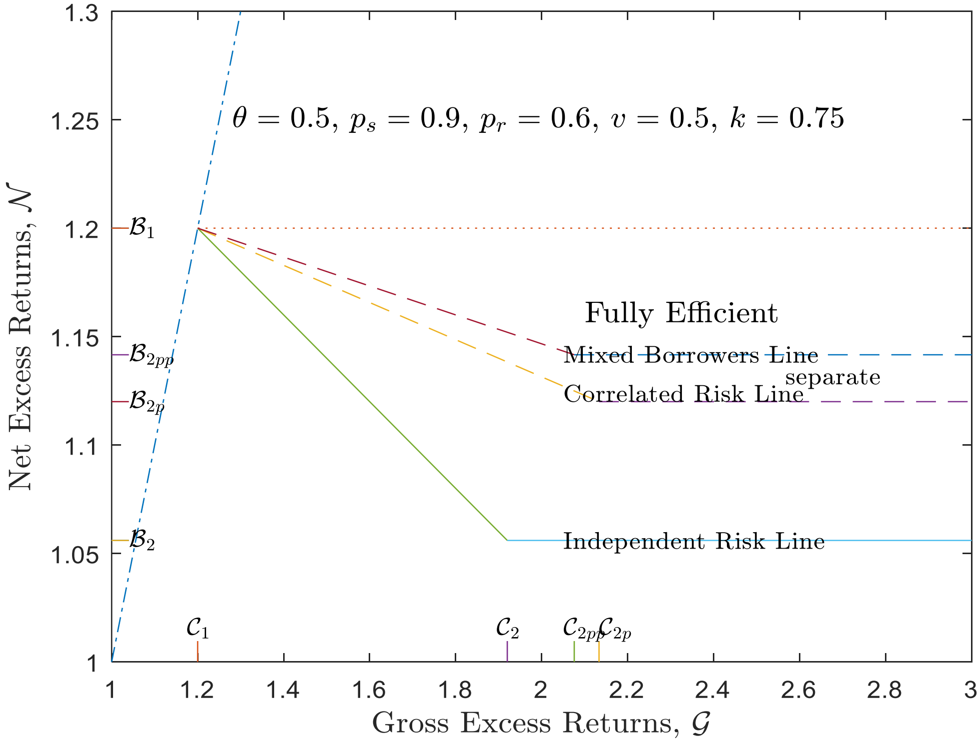

Figure 3-3: Mixed Pool of Borrowers for Higher k

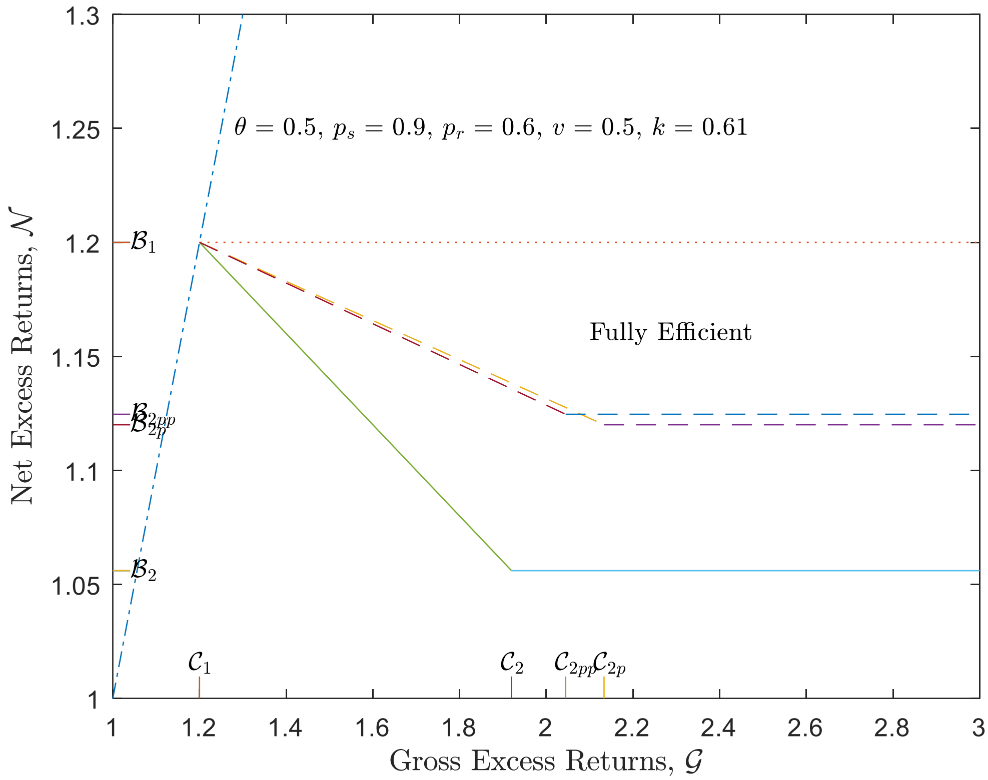

Figure 3-4: K=0.61 Where Mixed Line Crosses Correlated Line to Lie Above it)

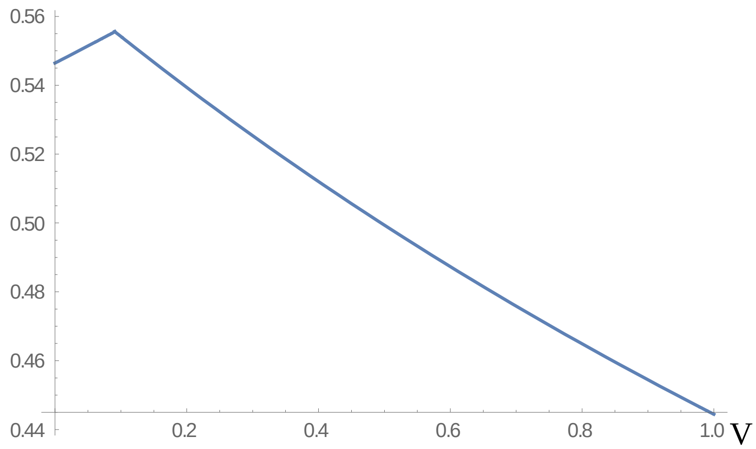

Figure 3-5: Plot of Joint Liability against Correlation among borrowers (C against V)
This is plotted by setting G=2, $\rho = 1$, $p_s = 0.9$, $p_s = 0.6$, $\theta = 0.5$
v<0.09(When Full Liability is Affordable) v>0.09(When Full Liability is Not Affordable)

Appendix B: Proofs of Lemmas and Propositions in Essay 2

Proof of Lemma 1 and Lemma 3

To show that homogeneous matching obtains, is efficient and is the only equilibrium, I compare the sum of group payoff from homogenous matching to group payoff from heterogeneous matching. This condition ensures that homogeneous matching is stable even in the presence of side contracting.

Consider two groups each of size two. I begin with the proof of Lemma 3. Lemma 1 follows similarly.

The group payoff of a safe borrower for partnering with another safe borrower,

$$\bar{R} - p_s r - [p_s(1-p_s) - \varepsilon_{ss}]c + \bar{R} - p_s r - [p_s(1-p_s) - \varepsilon_{ss}]c \quad (1)$$

[24]

The group payoff of a risky borrower pairing with another risky borrower,

$$\bar{R} - p_r r - [p_r(1-p_r) - \varepsilon_{rr}]c + \bar{R} - p_r r - [p_r(1-p_r) - \varepsilon_{rr}]c \quad (2)$$

a non-homogenous group's payoff (1st agent safe , second risky) is

$$\bar{R} - p_s r - [p_s(1-p_r) - \varepsilon_{sr}]c + \bar{R} - p_r r - [p_r(1-p_s) - \varepsilon_{sr}]c \quad (3)$$

(1st agent risky and second agent safe)

$$\bar{R} - p_r r - [p_r(1-p_s) - \varepsilon_{rs}]c + \bar{R} - p_s r - [p_s(1-p_r) - \varepsilon_{rs}]c \quad (4)$$

(1) + (2) Yields,

[24] ε_{ss} represents the ε when a safe agent partners with another safe agent

$$4\bar{R} - 2p_s r - 2p_r r - 2p_s(1-p_s)c - 2p_r(1-p_r)c + 2\varepsilon_{ss}c + 2\varepsilon_{rr}c \qquad (5)$$

And (3) + (4) yields

$$4\bar{R} - 2p_s r - 2p_r r - 2p_s(1-p_r)c - 2p_r(1-p_s)c + 2\varepsilon_{sr}c + 2\varepsilon_{rs}c$$

(6). this implies that we need

$-2p_s(1-p_s)c - 2p_r(1-p_r)c + 2\varepsilon_{ss}c + 2\varepsilon_{rr}c \geq -2p_s(1-p_r)c - 2p_r(1-p_s)c + 2\varepsilon_{sr}c + 2\varepsilon_{rs}c$

After some algebra and using the assumption about the form of the correlation parameter ε in this second case, we can have that, $(p_s - p_r)^2 - \Sigma \geq 0$ where $\Sigma = \varepsilon_{sr} + \varepsilon_{rs} - \varepsilon_{ss} - \varepsilon_{rr}$.

$\Leftrightarrow \quad \Sigma \leq \tilde{v}p_s(1-p_r) + \tilde{v}p_r(1-p_s) - \tilde{v}p_s(1-p_s) - \tilde{v}p_r(1-p_r) = \tilde{v}(p_s - p_r)^2$.Thus, homogenous matching obtains. Proof of Lemma 1 follows similarly by substituting ε in place of $\varepsilon_{sr}, \varepsilon_{rs}, \varepsilon_{ss}, \varepsilon_{rr}$.

Proof of Lemma 2

This is true because $\bar{R} - p_s r - [p_s(1-p_s) - \varepsilon]c < \bar{R} - p_r r - [p_r(1-p_r) - \varepsilon]c$

$\Leftrightarrow (p_s - p_r)(p_s + p_r - 1)c < (p_s - p_r)r$ Which is true since $(p_s + p_r - 1) < 1$ and $c \leq r$

Proof of Proposition 1

Max $\bar{Y} = \bar{R} - p_s r - [p_s(1-p_s) - \varepsilon]c$ subject to

$0 \leq c \leq r$, Monotonicity $[m]$

$\bar{p}r + \overline{p(1-p)}c - \varepsilon c \geq \rho$ Lender break even or zero profit constraint $[\mu]$

Where m and μ are the Lagrange multipliers respectively

The first order conditions are

$\lfloor r \rfloor$: $-p_s + \mu\bar{p} + m = 0$

$\lfloor c \rfloor$: $-p_s(1-p_s) + \varepsilon + \mu\overline{p(1-p)} - \varepsilon\mu - m = 0$

Solve to get

$\mu = \frac{p_s(2-p_s)-\varepsilon}{\overline{p(2-p)}-\varepsilon} > 0$ and $m = \frac{\theta(p_s-p_r)(p_s p_r-\varepsilon)}{\overline{p(2-p)}-\varepsilon} > 0$ if we assume $p_s p_r > \varepsilon$ in which case we have the solutions as $r = c = \frac{\rho}{\overline{p(2-p)}-\varepsilon}$

The remaining affordability constraint is satisfied if

$$R_s \geq r + c \Leftrightarrow R_s \geq 2r = 2\frac{\rho}{\overline{p(2-p)}-\varepsilon} \Leftrightarrow R_s p_s \geq 2p_s \frac{\rho}{\overline{p(2-p)}-\varepsilon} \Leftrightarrow \frac{\bar{R}}{\rho} \geq \frac{2p_s}{\overline{p(2-p)}-\varepsilon}$$

$$\Longrightarrow G \geq \frac{2p_s}{\overline{p(2-p)}-\varepsilon}$$

To attract the safe type, we need the expected payoff from the project to exceed the outside option. The condition is the inequality below

$$\bar{R} - p_s r - [p_s(1-p_s) - \varepsilon\,]r \geq \bar{u}$$

$$\Longrightarrow N \geq \frac{p_s(2-p_s)-\varepsilon}{\overline{p(2-p)}-\varepsilon}$$

Now, when affordability is an issue, (when $\frac{p_s}{\bar{p}} \leq G < \frac{2p_s}{\overline{p(2-p)}-\varepsilon}$)

The problem is

Max $\bar{Y} = \bar{R} - p_s r - [p_s(1-p_s) - \varepsilon\,]c$ subject to the following constraints

$R_s \geq r + c$ $\lfloor \lambda \rfloor$

$\bar{p}r + \overline{p(1-p)}c - \varepsilon c \geq \rho \qquad\qquad [\mu]$

$[r]$: $-p_s - \lambda + \mu\bar{p} = 0$

$[c]$: $-[p_s(1-p_s) - \varepsilon] - \lambda + \mu\overline{p(1-p)} - \varepsilon\mu = 0$

Solve to obtain the Lagrange multipliers as

$\mu = \frac{p_s^2}{(\overline{p^2}+\varepsilon)} > 0$ And $\lambda = \frac{\theta(p_s-p_r)(p_s p_r-\varepsilon)}{(\overline{p^2}+\varepsilon)} > 0$ as before, if we assume $p_s p_r > \varepsilon$,

The contract can be derived as

$(r,c) = \left\{\rho\frac{p_s-[\overline{p(1-p)}-\varepsilon]G}{p_s(\overline{p^2}+\varepsilon)}, \rho\frac{\bar{p}G-p_s}{p_s(\overline{p^2}+\varepsilon)}\right\}$

To attract safe borrowers, we can show that we need the following condition,

$N \geq \frac{p_s^2+\varepsilon}{(\overline{p^2}+\varepsilon)} - \left[\frac{\theta(p_s-p_r)(p_s p_r)}{p_s(\overline{p^2}+\varepsilon)}\right]G$

Constant correlation coefficient

$\varepsilon(p_i, p_j) = \tilde{v} * \min\{p_i(1-p_j), p_j(1-p_i)\}$

Let $\tilde{t}$ be the correlation coefficient between two projects with success probabilities p_i and p_j.

By definition, $\tilde{t} = \frac{Cov\,(Y_i,Y_j)}{\sigma_i\sigma_j}$ where Y_i, Y_j are the Bernoulli random variables associated with the project returns and σ_i and σ_j are the standard deviations.

$$\tilde{t} = \frac{Cov\,(Y_i, Y_j)}{\sigma_i \sigma_j} = \frac{E(Y_i\,Y_j) - E(Y_i)E(\,Y_j)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} = \frac{(p_i p_j + \varepsilon - p_i p_j) Y_i Y_j}{Y_i \sqrt{p_i(1-p_i)}\, Y_j \sqrt{p_j(1-p_j)}} = \frac{\varepsilon}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} =$$

$$\frac{\tilde{v} p_i(1-p_j)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}$$

The last equality comes from using $\varepsilon = \tilde{v} p_i(1-p_j)$ (assuming $p_i \leq p_j$ without loss of generality.

Hence for homogeneous projects ($p_i = p_j$) we have

$\tilde{t} = \tilde{v}$. Hence $\tilde{v}$ is just the correlation coefficient.

For non-homogeneous projects,

$$\Leftrightarrow \tilde{t} = \frac{\tilde{v} p_i(1-p_j)}{\sqrt{p_i(1-p_i)}\sqrt{p_i(1-p_j)}} = \frac{\tilde{v} p_i(1-p_j)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}$$

$$\Leftrightarrow \tilde{t} \; = \frac{\tilde{v} p_i(1-p_j)}{\sqrt{p_j(1-p_i)}\sqrt{p_i(1-p_j)}} = \frac{\tilde{v}\sqrt{p_i(1-p_j)}}{\sqrt{p_j(1-p_i)}}$$ so that,

$\tilde{t}$ is written generally as $\tilde{t} = \tilde{v}\bar{v}$ where $\bar{v} = min\left\{\frac{\sqrt{p_i(1-p_j)}}{\sqrt{p_j(1-p_i)}}, \frac{\sqrt{p_j(1-p_i)}}{\sqrt{p_i(1-p_j)}}\right\}$ is the maximum

correlation possible between two projects.

Thus $\tilde{v}$, is a fraction of the maximum correlation possible between the two project returns. In particular, for homogenous groups for which $p_i = p_j$, the maximum correlation possible is 1 hence $\tilde{v}$ is the correlation coefficient as a straight forward calculation shows.

Proof of Proposition 2

Consider for instance the condition $N \geq \frac{p_s(2-p_s)-\varepsilon}{\overline{p(2-p)}-\varepsilon}$ which is obtained under full affordability.

Let $M = \frac{p_s(2-p_s)-\varepsilon}{\overline{p(2-p)}-\varepsilon}$. Differentiating M, the right hand side of the above inequality with respect to ε yields, $\frac{dM}{d\varepsilon} = \frac{\theta(p_s-p_r)(2-p_s-p_r)}{[\overline{p(1-p)}-\varepsilon]^2} > 0$ which means that, M increases in ε and

Secondly, $\frac{dC_2}{d\varepsilon} = \frac{2p_s}{[\overline{p(1-p)}-\varepsilon]^2} > 0$

Thus, the parameter space for fully efficient lending reduces as correlation increases. The boundary of the parameter space comprises of a horizontal part and a negatively sloped part which starts from (C_1, B_1) and gets to its floor at (C_2, M). The entire boundary of the efficiency parameter space rotates upward with the introduction of correlation.

Proof of Proposition 3

Max $\bar{R} - p_s r - [p_s(1-p_s) - \tilde{v}p_s(1-p_s)]c$ subject to

$0 \leq c \leq r$, $R_s \geq r + c$, and $\bar{p}r + \overline{p(1-p)}(1-\tilde{v})c \geq \rho$. The rest follows as in proof of proposition 1.

Proposition 4

Consider for instance the condition $N \geq \frac{[p_s + p_s(1-p_s)(1-\tilde{v})]}{[\bar{p} + (1-\tilde{v})\overline{p(1-p)}]}$, which is obtained under full affordability.

Let $M' = \frac{[p_s + p_s(1-p_s)(1-\tilde{v})]}{[\bar{p} + (1-\tilde{v})\overline{p(1-p)}]}$. Differentiating M' the right hand side of the above inequality with respect to $\tilde{v}$ yields, $\frac{dM\prime}{d\tilde{v}} = \frac{\theta p_r p_s(p_s - p_r)}{\left[\bar{p} + (1-\tilde{v})\overline{p(1-p)}\right]^2} > 0$ which means that, M' increases in $\tilde{v}$

Secondly, $\frac{dC\prime_2}{d\tilde{v}} = \frac{2p_s\overline{p(1-p)}}{\left[\bar{p} + (1-\tilde{v})\overline{p(1-p)}\right]^2} > 0$

Thus, the parameter space for fully efficient lending reduces as correlation increases.

As in proof of proposition 2, the boundary of the parameter space comprises of a horizontal part and a negatively sloped part which starts from $(C_1\ , B_1\ )$ and gets to its floor at $(C_2'\ , M')$. The entire boundary of the efficiency parameter space rotates upward with the introduction of correlation.

Lemma 6

We show this by comparing payoffs of the groups A^{safe}, A^{risky}, B^{safe} and B^{risky}. The payoffs are as follows

A^{safe} : $\bar{R} - p_s r - [p_s(1-p_s)(1-\tilde{v})\ ]c$

A^{risky} : $\bar{R} - p_r r - [p_r(1-p_r)(1-\tilde{v})\ ]c$

B^{safe} : $\bar{R} - p_s r - [p_s(1-p_s)\ ]c$

B^{risky} : $\bar{R} - p_r r - [p_r(1-p_r)\ ]c$

By inspection, $A^{safe} > B^{safe}$ and $A^{risky} > B^{risky}$ since $\tilde{v} < 1$. Also, $B^{risky} > B^{safe}$ hence B^{safe} is the least well off in the group.

Proposition 5

Maximize $\bar{R} - p_s r - [p_s(1-p_s)\,]c$

Subject to

$0 \leq c \leq r$, and $\bar{p}r + \overline{p(1-p)}(1-k\bar{v})c \geq \rho$ (When full affordability is feasible)

$\lfloor r \rfloor$: $-p_s + \mu\bar{p} + \lambda = 0$

$\lfloor c \rfloor$: $-p_s(1-p_s) + \mu\overline{p(1-p)}(1-k\bar{v}) - \lambda = 0$

Solve to get; $\mu = \frac{p_s(2-p_s)}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]} > 0$ and $\lambda = p_s - \mu\bar{p} > 0$ if $\tilde{v} < \frac{\theta p_r(p_s-p_r)}{k\overline{p(1-p)}}$

The contract is then derived as

$r = c = \frac{\rho}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$ Assuming $\tilde{v} < \frac{\theta p_r(p_s-p_r)}{k\overline{p(1-p)}}$

The remaining affordability constraint is satisfied as before if

$G \geq \frac{2p_s}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$. To attract safe uncorrelated borrowers, we need

$\bar{R} - p_s r - [p_s(1-p_s)\,]c > \bar{u}$

That is $N \geq \frac{p_s(2-p_s)}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$

When affordability is an issue, (when $\frac{p_s}{\bar{p}} \leq G < \frac{2p_s}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}$)

We have the following constraints similar to section 3.6’s derivations.

$R_s \geq r + c$

$\bar{p}r + \overline{p(1-p)}(1-k\tilde{v})c \geq \rho$

Solve first order conditions to get $\mu = \frac{p_s^2}{[\bar{p}-\overline{p(1-p)}(1-k\tilde{v})]} > 0$ and $\lambda = \mu\bar{p} - p_s > 0$

The contract is derived as; $(r,c) = \left\{\rho\frac{p_s-[(1-\tilde{v}k)\overline{p(1-p)}]G}{p_s[\overline{p^2}+(\tilde{v}k)\overline{p(1-p)}]}, \rho\frac{\bar{p}G-p_s}{p_s[\overline{p^2}+(\tilde{v}k)\overline{p(1-p)}]}\right\}$ and to attract the safe uncorrelated borrowers, we need

$$N \geq \frac{p_s^2}{(\overline{p^2}+(\tilde{v}k)\overline{p(1-p)})} - \left[\frac{\theta(p_s-p_r)p_r-\tilde{v}kp(1-p)}{(\overline{p^2}+(\tilde{v}k)\overline{p(1-p)})}\right]G$$

Propositions 6

Define $\bar{k}_a = \frac{\theta p_r(p_s-p_r)}{\overline{p(1-p)}[1+(1-p_s)(1-\tilde{v})]}$ and $\bar{k}_b = \frac{\theta p_r(p_s-p_r)}{\overline{p(1-p)}[1-(1-p_s)(1-\tilde{v})]}$. Clearly, $0 < \bar{k}_a < \bar{k}_b$.

To show $\bar{k}_a < \bar{k}_b < 1$, it suffices to show $\bar{k}_b < 1$.

$$\bar{k}_b < 1 \Leftrightarrow \frac{\theta p_r(p_s-p_r)}{\overline{p(1-p)}[1-(1-p_s)(1-\tilde{v})]} < 1 \Leftrightarrow \theta p_r(p_s-p_r) < \overline{p(1-p)}[1-(1-p_s)(1-\tilde{v})]$$

$$\Leftrightarrow -\tilde{v}\bar{p} < (1-\tilde{v})\overline{p^2}$$

The cutoff values $\bar{k}_a$ and $\bar{k}_b$ that we have in propositions 6 are obtained by comparing the B_2 levels (for $\bar{k}_a$) and the slopes of the negatively sloped portion of the efficient lending parameter space boundary of the all-correlated, and that of the mixed-borrowers contracts (for $\bar{k}_b$). Thus for the cutoff value of k (for $\bar{k}_a$), we solve $\frac{p_s(2-p_s)}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]} < \frac{[p_s+p_s(1-p_s)(1-\tilde{v})]}{[\bar{p}+(1-\tilde{v})\overline{p(1-p)}]}$, and for the cutoff value of k (for $\bar{k}_b$), we solve $\frac{\frac{p_s}{\bar{p}}-\frac{p_s(2-p_s)}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}}{\frac{2p_s}{[\bar{p}+\overline{p(1-p)}(1-k\tilde{v})]}-\frac{p_s}{\bar{p}}} < \frac{\frac{p_s}{\bar{p}}-\frac{[p_s+p_s(1-p_s)(1-\tilde{v})]}{[\bar{p}+(1-\tilde{v})\overline{p(1-p)}]}}{\frac{2p_s}{[\bar{p}+\overline{p(1-p)}(1-\tilde{v})]}-\frac{p_s}{\bar{p}}}$. For $k\epsilon(\bar{k}_a, \bar{k}_b)$, it is clear that the functions intersect once.

We use the figures in section 3.9 to illustrate how these values are derived using the graphs in Figures 3-2, 3-3 and 3-4.

The displayed graphs in Figures 3-2, 3-3 and 3-4 are for k =0.25, 0.75 , 0.61 respectively, it can be seen from the graphs that, first of all, the lower bound of the gross excess return (G) line for the Mixed borrowers line (M), is smaller than that of the "All correlated" line (C). This means, we can have only 3 scenarios as k increases. And the three scenarios are depicted in the graphs. In figure 3-2, the efficient lending parameter space boundary of C lies completely above that of M and the reverse of this is the case in figure 3-3. In figure 3-4, the line for M crosses that of C and later lies above of it. We can see that, at low values of k, C lies above and as k increases, M rises, crosses initially and is steeper than C. Then with further increases in k, M lies above of C and is less steep. Thus, C lying above M is sufficient condition for pooling of borrowers. M lying above is not sufficient to call for separation of borrowers because, we can have the case where k=0.61 for instance, and separation is not fully justified. However, as k rises and M becomes less steep, we see M falling above C and remains there. Thus the condition that the slope of M is less steep compared to the slope of C is sufficient condition to call for separation. This ends the proof.

4 ESSAY 3: PREDICTORS OF THE CHOICE OF RURAL NONFARM ACTIVITY IN TANZANIA

Introduction

Rural nonfarm economies (RNFEs) have been found to be a good source of income- booster for rural dwellers in many countries and have contributed to the reduction of poverty. (FAO, 2008) reports average nonfarm income shares in developing countries to be about "42% for Africa, 40% for Latin America and 32% for Asia". Nonfarm income share of household income is between 21 and 23% for Tanzania (Haggblade, Hazell, & Brown, 1989). The non-agricultural participation rate is estimated to be about 75% in Ghana and 93% in Malawi (B. Davis et al., 2010). More recently (Ackah, 2013; Owusu, Abdulai, & Abdul-Rahman, 2011) used data from northern Ghana and found positive effects of nonfarm activities on income and food security statuses of households. Participation in RNFEs, does not only provide a diversified source of income to shield households against any negative shocks, they have also provided farmers, the means to purchase inputs and or reinvest in their farms, although there is mixed findings in the literature as to the use of income by farmers from nonfarm activities (Reardon, Crawford, & Kelly, 1994).

Though a plethora of the literature establishes a connection between participating in rural nonfarm activities and welfare or poverty statuses of rural households, (Bezu, Barrett, & Holden, 2012; Davis, 2004), there are barriers to entry so not all rural dwellers can participate. While the barriers may differ depending on the activity, the high-yielding nonfarm activities seem to be an option for the few rural dwellers who can afford or are wealthy. (Reardon, 1997), and (Barrett, Reardon, & Webb, 2001) report that higher income households tend to have a greater share of their income from nonfarm sector. This suggests that, higher income households may be able to take advantage of the RNFE while low income households may not.

In this paper, we focus on the predictors of participation in the RNFE, and predictors of choice between rural wage employment and self-employment conditional on participation. Recent debates in development economics, on the relative desirability of wage employment and self-employment options, motivate this paper. This debate is whether poor citizens are just frustrated wage earners, or they are just frustrated entrepreneurs. It is important to note that self-employed rural dwellers are not necessarily entrepreneurs especially if they engage in the RNFE as a way of coping with risk temporarily unlike entrepreneurs who take risks and have plans for expansion.

Wage employments tend to pay more than self-employment in rural areas and some previous papers have established that, poor people opt for self-employment because there really isn't a choice for them. See for example (Contreras, Gillmore, & Puentes, 2017; Fields, 2014). Contreras et al found double selection. In that, some workers choose themselves to be self-employed while some are forced into self-employment because they couldn't get access to wage employments. As (Fields, 2014) notes, more research is needed to ascertain why households choose their type of employments. Understanding why households choose the type of employment is very vital to efforts aimed at eradicating extreme poverty in developing countries.

In my effort to learn some of the predictors of households' choice of self or wage employment upon entering the nonfarm economy, I restrict attention to three categories of households. The first category has households that did not engage in any nonfarm activity. The second category comprises of households that engaged in some nonfarm activity consisting of only wage employment. The third category comprises of households that engaged in some nonfarm activity consisting of only self-employment. Many previous works have only compared participating households to non-participating ones making this paper different. In addition, I use panel data to mitigate bias due to unobserved household characteristics. Many previous papers have used cross-sectional data. I enrich the literature by providing newer evidence on the hypothesized relationships between key variables and participation in the RNFE.

The results suggest wealth and land assets are key predictors of participation in the RNFE. They also predict self-employment over rural wage employment, conditional on participation. I find that, 1 standard deviation increase in non-agricultural wealth index, predicts an average increase in the likelihood of participating in rural self-employment by approximately 16%. This result is about 5% for likelihood of participating in rural wage employment. 1 standard deviation increase in non-agricultural wealth index, predicts an average increase in the likelihood of choosing self-employment over wage employment by approximately 5%. This latter result is conditional on participation. Households appear to age out of self-employment or that, rural self-employment appear to be engaged in by younger households.

There are always endogeneity issues to grapple with when studying participation in the RNFE. That is, do wealthy households tend to participate or there is a reverse causality in the sense that, households rather participate and later become wealthy. Households can be pulled into the RNFE for asset accumulation purposes and households can also be pushed into the RNFE as a way of coping with economic hardships. (Dimova & Sen, 2010) found that, households in the Kagera region of Tanzania, rather diversify income sources in order to accumulate assets and not for survival purposes. I find significant Fixed Effects results between participation decision and wealth index variables suggesting that, reverse causality could be of less concern. This could be so because it is not likely that within three years of the panel data, households have been able to participate in the RNFE and accumulated enough wealth that can provide enough within household variation needed for Fixed Effect estimations.

The RNFE needs to be studied in more details because aside being a source of asset accumulation, it provides a suitable platform for empowering the poor and vulnerable. In addition, rural dwellers are increasingly making efforts themselves to engage in nonfarm activities to increase their incomes (Winters et al., 2009). Governments are able to help the rural poor to overcome some of the barriers to entry in order to improve the livelihoods of poor households (Barrett et al., 2001; Delgado & Siamwalla, 1997), if we know the predictors and the barriers. Tanzania's emerging economy makes its rural economy an area of interest especially when recent studies have shown that rural dwellers adopt diversification as a pathway out of poverty, (De Weerdt, 2010).

The rest of the paper is organized as follows. Section 4.2 gives a description of the data used in the analysis. The estimation methods are presented in section 4.3. The results and discussions are in section 4.4. Section 4.5 has the conclusion.

Data Description and Patterns of Rural Income Diversification in Tanzania

The data used in this paper is the Tanzanian rural income generating activities data obtained from the FAO database. The FAO constructs this rural income generating activities data (RIGA) using the LSMS data made available by the World Bank. I use the first three waves collected in 2008/2009, 2010/2011 and 2012/2013. The survey was designed to be nationally, urban/rural and agro-ecologically representative. The households were clustered into 409 enumeration areas, which comprised of 2063 rural and 1,202 urban households in the first round. I use only the rural sample of the data that form a balanced panel. The World Bank database has the documentation on the details of the sampling methods and procedures.

Table 4-1 describes the pattern of income diversification and some demographic characteristics in the data set. The average household size in the three waves of the data is about 5 persons per household. These households on average also have about 25% of them headed by a female. The size of the labor force is approximately less than three people per household. Households in the rural areas do not appear to have land abundantly available because the average land size owned by the households is about 2 hectares. In terms of pattern of income diversification, about 22% of households had at least one of their members participate in agricultural wage activities in the first wave. This fraction became almost 30% by 2012/2013 in the third wave.

Non-agricultural wage on the other hand hovered around 23% in the second and third waves of the data, which indicates that non-agricultural wage employment, and diversification is still popular in rural Tanzania. In total, the percentage of households participating in the agricultural sector in general fell from almost 99% in 2008/2009 to 92% in the 2012/2013 wave. The pattern is similar with participation in livestock production.

Agricultural consisted about 70% of household income in the first wave and about 62% in the third wave. Non-agricultural share of household income stood at about 30% in the first wave and increased to about 38% in the third wave. These provide evidence that households in rural Tanzania are diversifying their sources of income. The balanced panel of rural households obtained for this analysis is 4500 and the standard deviations of our key variables namely non-agricultural wealth index, agricultural wealth index and land owned in hectares, from this 4500 balanced panel are 1, 0.95 and 2.446 respectively. Details of descriptive statistics for the subsample of the waves used in the regressions are contained in tables 4-2, 4-3 and 4-4. The final data used in this study vary depending on which regression is being run but only balanced panel is used throughout.

Estimating Participation in RNFE

Because the participation in RNFE is a binary variable, I model the decision to participate in nonfarm wage or self-employment using the linear probability Fixed Effect (LFE) and Correlated Random Effects (CRE) models. A natural place to start in terms of predicting the probability of a household participating in nonfarm wage or self-employment activities is to consider the linear probability model with Fixed Effects. This model has the potential of providing good estimates in addition to it not requiring a distributional assumption on the unobserved heterogeneity term conditional on observables (Wooldridge, 2010). Specifically, it could give estimates that are free of bias stemming from household-specific unobserved variables that are fixed between panel waves, such as ability or underlying wealth. I report results from Correlated Random Effects and those from Linear Fixed Effects probability models for comparison purposes.

More formally, we define the likelihood to participate in the RNFE following (Wooldridge, 2010) as follows

$$y_{it}^{*} = X_{it}\beta + c_i + e_{it}\ ,\ y_{it} = 1[y_{it}^{*} > 0] \tag{1}$$

where y_{it}^{*} is a latent variable that captures the marginal value of time to the household in the RNFE. y_{it} is the decision to participate or not in the RNFE. Then, the response probability for the linear probability model is

$$prob(y_{it} = 1|X_{it}, c_i) = X_{it}\beta + c_i, \tag{12}$$

The Probit model of this, takes the form

$prob(y_{it} = 1|X_{it}, c_i) = \Phi(X_{it}\beta + c_i)$ where X_{it} in the models above is a vector of covariates and c_i is the unobserved heterogeneity term. The Correlated Random Effects model which allows for dependence between c_i, and X_{it} takes the form

$prob(y_{it} = 1|X_{it}, c_i) = \Phi(X_{it}\beta + c_i)$, a_i/X_i~ Normal (0,σ_a^2)

c [25] $_i = \psi + \bar{X}_i\xi + a_i$.

The full model can be written as follows

$prob(y_{it} = 1|X_{it}, c_i) = \Phi(\psi_a + X_{it}\beta_a + \bar{X}_i\xi_a + c_i)$.

$\psi_a, \beta_a\ and\ \xi_a$ are obtained by multiplying $\psi, \beta, and\ \xi$ by $(1 + \sigma_a^2)^{-1/2}$. In this case, $\psi_a, \beta_a\ and\ \xi_a$ are estimated consistently in a pooled Probit of y_{it} on $X_{it}, \bar{X}_i$. I test for the joint significance of the coefficients of $\bar{X}_i$ as is normally done and find that they are jointly significant which means CRE use is justified.

Results and Discussions

Tables 4-2, 4-3 and 4-4 show the summary statistics of covariates used in the regressions. The tables contain information on per capita household expenditures among the three household categories considered in this paper. The three categories as a reminder are households that did

[25] I use the Mundlak version. The general form has just X_{it} and not their averages.

not engage in any nonfarm activity, households that engaged in some nonfarm activity consisting of only wage employment, and households that engaged in some nonfarm activity consisting of only self-employment. For simplicity, I refer to the first two categories of household henceforth as wage employment and self-employment households respectively.

From these tables, we see that the mean per capita expenditures are higher for non-agricultural wage employees in the balanced panel used. The mean per capita expenditure for the rural self-employed category is also higher in all survey rounds (for the balanced panel sample used) than for the category of households that did not participating in any non-agricultural activity.

The expected sign of covariates in predicting participation in rural nonfarm economy or in predicting what type of activity the household would choose when they participate, depends on whether the household is diversifying for accumulation of assets or income growth, or the household is diversifying income sources as a way of coping with risk. Some of the covariates can have either positive or negative expected signs depending on the motive for diversification. For example, if households engage in the RNFE as a way of coping with risks, then we would expect households to not participate in the RNFE when they have enough assets (that is a negative relationship between assets and participation decision) and vice versa.

Agricultural Wealth, Non-agricultural Wealth and Landholdings

While agricultural wealth makes it easy for households to expand their agriculture and may probably not engage in nonfarm activities, non-agricultural wealth plays an important role in predicting the participation decision of households in the RNFE. Once these households choose to participate, the wealth status of the household can also predict whether they self-finance or can access a loan and engage in self-employment or opt for rural wage employment. As mentioned in the introduction, households may be pushed into the RNFE as a way of getting means to survive and households could also be pulled into the RNFE as a way of accumulating assets. Availability of land to a household is expected to aid the household see an enhancement in their agricultural activities relative to households without land. Land however can assist households to participate in non-agricultural rural self-employment by serving as a source of collateral security. Thus, households with a good amount of land could have flexibility or ease when it comes to participating in the RNFE. Conditional on participating in the RNFE, these households can also choose self-employment because they can finance while those with little or no land may participate in the RNFE as wage earners if credit markets are imperfect.

(Taylor & Yunez-Naude, 2000) found land variable and farm engagements to be positively related using data from Mexico. They also found a negative relationship between land and rural wage employment participation. In addition, (Winters et al., 2009) used data covering 15 countries and found that availability of land makes it easy for households to engage in agricultural activities to improve their welfare. (Ackah, 2013) finds land endowment to be negatively correlated with RNFE participation.

Table 4-5 shows the results of predictors of participation in rural non-agricultural wage employment. I compare rural dwellers who participated in rural wage employment, with the group that did not participate in any non-agricultural employment. In table 4-6, I predict participation in the RNFE as a self-employed household. Here, I compare rural dwellers who participated in self-employment, against the group that did not participate in any non-agricultural employment. Then in Table 4-7, I show the predictors of choice between nonfarm self-employment and rural wage employment conditional on participating in the RNFE. This table enables us to learn about what type of rural nonfarm activity household choose to undertake once they opt to participate in the RNFE. All the households in tables 4-5 to 4-7 that participated in the RNFE may or may not have engaged in some agricultural activities.

While both Tables 4-5 and 4-6 show a positive relationship between non-agricultural wealth index and participation in rural wage employment, or rural self-employment, they both also show a negative relationship between agricultural wealth index and participation in rural wage employment and in rural self-employment. The relationship between participating in nonfarm self-employment and wealth, shown in Tables 4-5 and 4-6, appear to be in favor of the "pull factors" literature. The "pull factors" literature says the RNFE or activities may rather pull the wealthy or those with the means to participate in them (Haggblade, Hazell, & Reardon, 2010), as against the push factors argument, (Bardhan & Udry, 1999) which says households are pushed to diversify as a way of coping with risk.

The Linear Fixed Effects estimates are highly significant for the non-agricultural wealth index variable in Tables 4-5 and 4-6. Since Fixed Effects estimation uses within household variations only, it must be that there is significant within household variations in the data and thus gives credence to the estimates obtained from the Correlated Random Effects. The significance of the Linear Fixed Effects estimates may also suggest that reverse causality should be of less concern. It is less likely that households within three years of the survey rounds have accumulated enough assets as a result of being in the RNFE such that there is within household variation sufficient to yield significant Fixed Effect estimates. Thus the pull factors literature is supported here too.

The amount of land the household owns is significant and negative in the rural wage-employment equation (Table 4-5) but not significant in the rural self-employment regression in Table 4-6. The wealth variables may have absorbed the effects of the land variable in Table 4-6. The results on the land variable carry the expected signs in both Tables 4-5 and 4-6. The Pooled CRE results in Table 4-5 suggests that, 1 standard deviation increase in non-agricultural wealth index, predicts an average increase in the likelihood of participating in rural wage-employment by approximately 5%. On the other hand, 1 standard deviation increase in agricultural wealth index, predicts an average decrease in the likelihood of participating in rural wage employment by approximately 3%. Looking at Table 4-6, we can see that, 1 standard deviation increase in non-agricultural wealth index, predicts an average increase in the likelihood of participating in rural self-employment by approximately 16%. On the other hand, 1 standard deviation increase in agricultural wealth index, predicts an average decrease in the likelihood of participating in rural wage employment by approximately 3%. Altogether, the results seem consistent with the "pull factors" side of the RNFE participation literature.

Table 4-7 shows that, for households that participate in the RNFE, non-agricultural wealth is positively associated with choosing to be in rural self-employment relative to rural wage employment. This is not surprising because we would expect households with enough non-agricultural wealth to self-finance and thus are likely to choose self-employment over wage employment once they decide to participate in the RNFE. This result is also not so consistent with reverse causality since wage earners earn more than self-employed households on average (Tables 4-2, 4-3 and 4-4). From Table 4-7, we interpret the CRE results to mean, 1 standard deviation increase in non-agricultural wealth index, predicts an average increase in the likelihood of choosing rural self-employment over wage employment by approximately 5%.

Still on Table 4-7, the land variable is significantly and positively associated with participating households choosing self-employment over rural wage employment. 1 standard deviation increase in hectares of land owned by households, predicts an average increase in the likelihood of choosing rural self-employment over wage employment by approximately 11%.

This could suggest that, households with more land choose to do self-employment over wage employment because they can leverage their land assets to be able to self-finance their business. This is consistent with the argument that some rural nonfarm opportunities may be the sole preserve of asset owners since credit markets are imperfect in developing countries. The Linear Fixed Effects results are highly significant indicating a significant variation within households in terms of land holdings and that the result holds within households.

In sum, wealth or assets do not only predict participation, they predict whether households would choose self-employment over wage employment at least in this data. In other words, wealth or assets are key predictors of occupational choice by households conditional on participating in the RNFE. It appears that the assets allow households may have the flexibility and the wherewithal required to venture into some of the rural nonfarm activities. Households that are able to self-finance would do so and those that cannot but have collateral securities would be able to access funds to be able to finance their projects.

Demographic Characteristics and other Covariates

The demographics of a household can also predict participation in the RNFE. For example, the household head's gender can influence the ownership and distribution of resources, which in turn influences the productivity and income levels of the household. Similarly, the household head's marital status and education level can influence the allocation of resources in the household.

Education is expected to make a household more eligible to participate in the RNFE. Conditioned on participation, education is likely to make households choose high-paying wage employment than self-employment in the rural areas. Some previous studies like (Lanjouw, Quizon, & Sparrow, 2001) found that education aids participation in non-agricultural wage employment and even self-employment in Tanzania. They also find proximity to infrastructure as key determinant of nonfarm income of the peri-urban dweller in Tanzania. (Ackah, 2013) finds human capital to be positively related to nonfarm participation.

From our results in Tables 4-5 and 4-6, average education level in the household is significant with a positive sign. That is, higher education can predict participation in wage employment and self-employment but weakly significant in the self-employment equation as expected from the discussion above. The age of the head of the household variable has a negative association with participation in rural self-employment but no significant relationship with participation in rural wage employment as shown in Tables 4-5 and 4-6. The summary statistics in Tables 4-2, 4-3 and 4-4 show that in all survey rounds, age of no-non-agricultural employment category of rural dwellers is a bit higher than those in the self-employment, and the rural wage-employment categories. This suggests or predicts that, older people are less inclined to go into rural self-employment relative to agricultural activities. This could be because such old people would have garnered so much experience in farming so that, they are unwilling to move to the RNFE. This may also help us to understand that younger people are increasingly getting interested in the RNFE and are leaving the farm sector for the older people to undertake.

In Table 4-7, where we examine the choice of occupation upon participating in the RNFE, the age of the household head variable has a zero and insignificant coefficient. This suggests that age may not matter in what a household chooses to participate in conditional on participation.

Married household head variable has a negative and insignificant association with participation in rural wage employment and a negative and significant relationship with rural self-employment in Tables 4-5 and 4-6. This could also be because, married household heads tend to have enough hands to help on their farms and thus are less inclined to leave agriculture entirely.

In Table 4-7, we see that for the households that participate in the RNFE, the likelihood of choosing self-employment over wage employment is lower for household heads that are married. This could be perhaps because these married heads of households think about food security and how to provide for the family and thus would like to opt for an activity that provides quick and guaranteed source of income.

The household size variable has a positive and significant relationship with participating in rural wage and self-employment equations as shown in Tables 4-5 and 4-6. The bigger the size of the household, the more likely it is that a member is qualified to or is able to participate in the RNFE. This result is expected.

Although in Tables 4-5 and 4-6, we see no significant results in relation to distance from the nearest market and participation in nonfarm employment, the summary statistics show that those who are farther from the nearest market are those who did not engage in any non-agricultural employment for income. In Table 4-7 the result shows there is no relationship between distances of plot to the nearest market and the choice of rural nonfarm activity.

Availability of electricity at the dwelling place of the household variable is significant in the wage-employment regression in Table 4-5 and carries a positive sign. It also is significant and negative in the self-employment regression in Table 4-6. Electricity expansion, availability or coverage in the area may be taking place based on some criteria not observed in the data.

In Table 4-7 having electricity at the dwelling place has a negative relationship with the likelihood of choosing rural self-employment over rural wage employment. This suggests that wage employment may be lucrative than self-employment as portrayed by the summary statistics in Tables 4-2, 4-3 and 4-4 or that wage employees have electricity provided for them by their employers. It could also be that infrastructure presence is correlated with wage employment opportunities. Proximity to infrastructure in general has been found in (Winters, Davis, & Corral, 2002) to be a predictor of participation in the RNFE in Mexico.

Conclusion

I study the predictors of participation in rural wage, and rural self-employment. I focus on three categories of households and these are, households that did not engage in any nonfarm activity, households that engaged in some nonfarm activity consisting of only wage employment, households that engaged in some nonfarm activity consisting of only self-employment. I do not include households that participated in both rural wage and self-employment in our analysis. Aside studying the predictors of participating in the rural nonfarm economy, I also look at predictors of the choice between rural self-employment and wage employment conditional on participation.

Most previous studies have just compared participants in the rural nonfarm economy to non-participants. I enrich the literature by providing newer evidence on the relationships between key variables and households' decisions to partake in the rural nonfarm economy. The relationship between participating in nonfarm wage, self-employment and wealth, appear to be in favor of the "pull factors" literature. The "pull factors" literature says the rural nonfarm activities may rather pull the wealthy or those capable of venturing into it for asset accumulation purposes (Haggblade et al., 2010), against the push factors argument, (Bardhan & Udry, 1999) which says households are pushed to diversify as a way of coping with risk. Aside the wealth index variables, the amount of land available to households plays a significant role in a household's decision to go into the rural nonfarm economy.

Putting all together, we can learn that wealth and land assets predict household's engagement in agriculture and their continuous stay or their ability to move to nonfarm activities. The type of rural nonfarm activity they choose is also related to the availability of assets, thereby siding with the "pull factors" argument about why households engage in income diversification. Some significant Fixed Effects results suggest that, reverse causality in the form of households gaining wealth after joining the rural nonfarm economy, could be of less concern.

APPENDIX

Appendix: Tables for Essay 3

Table 4-1: Rural Tanzania Income Generating Activities

Survey Round:	2008/2009		2010/2011		2012/2013	
			Means (TZS) or Percentages (%)			
Variables	Sample size	Mean or %	Sample Size	Mean or %	Sample size	Mean or %
Household size	2055	5.43	2576	5.48	3144	5.26
Age of head	2055	47.22	2575	47.60	3143	47.00
Edu of head	2025	4.49	2563	4.67	3127	4.89
Female head (%)	2055	24.39	2489	25.61	3144	25.79
Married head (%)	2055	66.5	2576	55.05	3144	60.58
Household labor force	2055	2.52	2576	2.59	3144	2.51
Years of edu of labor force	2053	3.92	2576	3.40	2980	5.71
Per capita consumption exp	2055	33501.73	2576	517302.5	3144	708717
Ag. wage emp income	2055	22931.38	2622	43942.35	3209	103343.7
Non-ag wage emp income	2055	105820.2	2622	245225.9	3209	379923.4
Self-employment income	2055	215522	2622	347306.4	3209	481059.7
Total household income	2055	768844.9	2622	1984343	3209	1763192
% Ag. wage participation	2055	21.77	2622	27.77	3209	29.87
%non-Ag. wage participation	2055	14.67	2622	22.56	3209	22.82
% Ag. participation	2055	98.51	2622	94.322	3209	92.22

Table 4-1 (cont'd)

Share of Income Generating Activity						
Agricultural(%)	2050	70.22	2602	66.60	3194	61.70
On farm(%)	2050	65.94	2602	61.24	3194	54.08
Non ag (%)	2050	29.78	2602	33.39	3194	38.30
Nonfarm (%)	2050	19.24	2602	24.97	3194	25.86
Ag wage	2050	4.28	2602	5.37	3194	7.63
Non ag wage	2050	6.52	2602	10.16	3194	10.92
Land owned(ha)	2055	1.59	2576	1.83	3144	1.81
Household owns dwellings	2055	91.89	2576	86.68	3144	84.64

Survey weights applied.

Table 4-2: Means and Percentages (2008 Survey Data)

	Wage Employment (N=139)	Self-Employment (N=342)	No Non-Ag Employment (N=724)	Full Sample (N=1205)
Ag wealth index	-0.15	-0.04	-0.04	-0.06
Wealth index	0.55	0.25	-0.27	-0.03
Age of household head	43.89	44.90	49.82	47.74
Married household head	0.76	0.65	0.65	0.66
Household size	5.43	5.34	5.27	5.31
Average education of household	5.53	4.50	3.41	3.96
Land owned in hectares	0.90	1.66	1.47	1.46
Dwelling has electricity	0.12	0.04	0.02	0.03
Distance to nearest government school	0.33	0.19	0.10	0.15
Distance to nearest market	4.24	4.04	3.80	3.91
Per capita expenditure	44664.80	36570.19	30407.00	33800.89
Non ag wage employment income	803921.10	0.00	0.00	92734.46
Self-employment income	0.00	717001.20	0.00	203497.40
Transfer income	44562.59	44732.28	56120.44	51555.05
Other income	611.51	0.01	546.96	399.17

Source: Author’s computation from survey data. Means are in Tanzanian Shillings.
per capita expenditure is the total expenditure of the household divided by the household size.
Wealth indexes are computed by the FAO using a broad range of assets owned by the household. A principal component analysis is used in computing the indexes. The assets include television, radio, refrigerator, flooring of the house, tractor, plough, etc (for agricultural wealth indexes) See (Wealth Index Mapping in the horn of Africa-FAO 2011)

Table 4-3: Means and Percentages (2010 Survey Data)

	Wage Employment (N=159)	Self-Employment (N=369)	No Non-Ag Employment (N=677)	Full Sample (N=1205)
Ag wealth index	-0.22	-0.04	0.03	-0.03
Wealth index	0.06	0.01	-0.29	-0.15
Age of household head	43.11	44.61	50.68	47.82
Married household head	0.51	0.53	0.59	0.56
Household size	4.84	5.40	4.98	5.09
Average education of household	3.91	3.25	2.87	3.12
Land owned in hectares	1.24	2.21	2.01	1.97
Dwelling has electricity	0.11	0.04	0.03	0.04
Distance to nearest government school	0.26	0.19	0.18	0.20
Distance to nearest market	2.97	2.75	2.67	2.73
Per capita expenditure	687218.40	521153.30	441931.10	498556.60
Non ag wage employment income	996955.50	0.00	0.00	131548.50
Self-employment income	0.00	841496.60	0.00	257686.50
Transfer income	47185.64	47641.84	47994.34	47779.69
Other Income	5157.23	12555.56	1556.87	5400.00

Source: Author's computation from survey data. Means are in Tanzanian Shillings.
Per capita expenditure is the total expenditure of the household divided by the household size.
Wealth indexes are computed by the FAO using a broad range of assets owned by the household. A principal component analysis is used in computing the indexes. The assets include television, radio, refrigerator, flooring of the house, tractor, plough, etc. (for agricultural wealth indexes) See (Wealth Index Mapping in the horn of Africa-FAO 2011)

Table 4-4: Means and Percentages (2012 Survey Data)

	Wage Employment (N=137)	Self-Employment (N=363)	No Non-Ag Employment (N=705)	Full Sample (N=1205)
Ag wealth index	-0.16	0.05	-0.04	0.03
Wealth index	0.21	0.10	-0.27	-0.10
Age of household head	43.70	45.31	51.76	48.90
Married household head	0.68	0.61	0.62	0.63
Household size	5.18	5.77	5.09	5.31
Average education of household	4.40	3.53	3.00	3.32
Land owned in hectares	1.40	2.34	2.16	2.13
Dwelling has electricity	0.17	0.08	0.03	0.06
Distance to nearest government school	0.06	0.05	0.08	0.07
Distance to nearest market	2.45	1.64	2.27	2.10
Per capita expenditure	795294.10	684298.30	562346.20	625568.10
Non Ag wage employment income	1824722.00	0.00	0.00	207458.00
Self-employment income	0.00	1237116.00	0.00	372674.60
Transfer income	74595.13	67399.09	101726.70	88301.00
Other income	51576.64	86457.30	14936.17	40647.30

Source: Author's computation from survey data. Means are in Tanzanian Shillings.
per capita expenditure is the total expenditure of the household divided by the household size.
Wealth indexes are computed by the FAO using a broad range of assets owned by the household. A principal component analysis is used in computing the indexes. The assets include television, radio, refrigerator, flooring of the house, tractor, plough, etc (for agricultural wealth indexes) See (Wealth Index Mapping in the horn of Africa-FAO 2011)

Table 4-5: Predictors of participation in Non-Ag Wage Employment

	(1) Linear Fixed Effects	(2) CRE Probit Pooled	(3) CRE Probit Joint
Non-agricultural wealth index	0.043**	0.053***	0.050***
	(0.021)	(0.015)	(0.013)
Agricultural wealth index	-0.026	-0.029*	-0.029**
	(0.019)	(0.016)	(0.014)
Age of household head	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Head of household married	-0.009	-0.018	-0.016
	(0.027)	(0.020)	(0.020)
Household size	0.007	0.011**	0.011**
	(0.005)	(0.005)	(0.005)
Average years of education in household	0.033***	0.028***	0.028***
	(0.006)	(0.006)	(0.006)
Land owned in hectares	-0.014**	-0.027***	-0.027***
	(0.006)	(0.007)	(0.008)
Dwelling has electricity	0.154*	0.102**	0.104**
	(0.080)	(0.049)	(0.049)
KM from community to nearest government primary school	0.003	0.000	0.001
	(0.012)	(0.014)	(0.016)
KM from plot to nearest market	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Year dummy (2008 base year)	Yes	Yes	Yes
Time average of covariates	No	Yes	Yes
Constant	0.084		
	(0.075)		
Observations	1464	1464	1464
R^2	0.105		

Standard errors are in parentheses and are clustered at the household. Households with Non-agricultural wage employment are compared to those with no non-agricultural employment
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-6: Predictors of Participation in Non-Ag Self Employment

	(1) Linear Fixed Effects	(2) CRE Probit Pooled	(3) CRE Probit Joint
Non-agricultural wealth index	0.167***	0.163***	0.163***
	(0.021)	(0.021)	(0.015)
Agricultural wealth index	-0.035**	-0.031***	-0.032***
	(0.015)	(0.011)	(0.011)
Age head of household	-0.006***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)
Head of household married	-0.053**	-0.076***	-0.073***
	(0.025)	(0.020)	(0.019)
Household size	0.014***	0.016***	0.016***
	(0.005)	(0.005)	(0.005)
Average years of education in household	0.009	0.010	0.010*
	(0.006)	(0.006)	(0.006)
Land owned in hectares	0.004	0.003	0.003
	(0.005)	(0.005)	(0.005)
Dwelling has electricity	-0.155*	-0.118*	-0.130**
	(0.079)	(0.065)	(0.058)
KM from community to nearest government	0.009	0.009	0.009
primary school	(0.011)	(0.010)	(0.012)
KM from plot to nearest market	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Year dummy (2008 base year)	Yes	Yes	Yes
Time average of covariates	No	Yes	Yes
Constant	0.529***		
	(0.067)		
Observations	2589	2589	2589
R^2	0.103		

Standard errors are in parentheses and are clustered at the household. Households with Non-agricultural self-employment are compared to those with no non-agricultural employment

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-7: Predictors of Choice of Non-Agricultural Self-Employment

	(1) Linear Fixed Effects	(2) CRE Probit Pooled	(3) CRE Probit Joint
Non-agricultural wealth index	0.099***	0.054*	0.064***
	(0.038)	(0.029)	(0.025)
Agricultural wealth index	-0.001	-0.045	-0.031
	(0.043)	(0.037)	(0.039)
Age of household head	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)
Head of household married	-0.070	-0.089*	-0.082*
	(0.063)	(0.050)	(0.047)
Household size	-0.006	0.007	0.005
	(0.014)	(0.016)	(0.013)
Average years of education in household	-0.007	-0.004	-0.005
	(0.016)	(0.015)	(0.012)
Land owned in hectares	0.023	0.047**	0.043*
	(0.015)	(0.023)	(0.023)
Dwelling has electricity	-0.359**	-0.212**	-0.234**
	(0.139)	(0.102)	(0.092)
KM from community to nearest government primary school	-0.044**	-0.040**	-0.039*
	(0.018)	(0.020)	(0.021)
KM from plot to nearest market	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
Year dummy (2008 base year)	Yes	Yes	Yes
Time average of covariates	No	Yes	Yes
Constant	0.864***		
	(0.173)		
Observations	372	372	372
R^2	0.113		

Standard errors are in parentheses and are clustered at the household. hh denotes household. Households with Non-agricultural wage employment are compared to those with non-agricultural self employment. Households may have some agricultural activities but not both wage and self-employment.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

BIBLIOGRAPHY

BIBLIOGRAPHY

Ackah, C. (2013). Nonfarm employment and incomes in rural ghana. *Journal of International Development*, *25*(3), 325–339. https://doi.org/10.1002/jid.1846

Ahlin, C. (2009). Matching for Credit: Risk and Diversification in thai Microcredit Groups. *Michigan State University Working Paper*, (March), 1–49.

Ahlin, C., & Jiang, N. (2008). Can micro-credit bring development? *Journal of Development Economics*, *86*(1), 1–21. https://doi.org/10.1016/j.jdeveco.2007.08.002

Ahlin, C., & Townsend, R. M. (2007a). Selection into and across credit contracts: Theory and field research. *Journal of Econometrics*, *136*(2), 665–698. https://doi.org/10.1016/j.jeconom.2005.11.013

Ahlin, C., & Townsend, R. M. (2007b). Using Repayment Data To Test Across Models Of Joint Liability Lending, *117*(1999).

Ahlin, C., & Waters, B. (2016). Dynamic Microlending under Adverse Selection: Can it Rival Group Lending? *Journal of Development Economics*, *121*, 237–257. https://doi.org/10.1016/j.jdeveco.2014.11.007

Ali, D. A., & Deininger, K. (2015). Is There a Farm Size–Productivity Relationship in African Agriculture? Evidence from Rwanda. *Land Economics*, *91*(2), 317–343. https://doi.org/10.3368/le.91.2.317

Armenda, B., Aghion, Â. De, & Gollier, C. (2000). Peer Group Formation in an Adverse Selection Model Ã. *The Economic Journal*, *110*(March 1996), 632–643.

Assunção, J. J., & Braido, L. H. B. (2007). Testing household-specific explanations for the inverse productivity relationship. *American Journal of Agricultural Economics*, *89*(4), 980–990. https://doi.org/10.1111/j.1467-8276.2007.01032.x

Assunção, J. J., & Ghatak, M. (2003). Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity. *Economics Letters*, *80*(2), 189–194. https://doi.org/10.1016/S0165-1765(03)00091-0

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22–53. https://doi.org/10.1257/app.20130533

Bardhan, P., & Udry, C. (1999). *Development Microeconomics*.

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship.*Journal of Development Economics*, *51*(2), 193–215. https://doi.org/10.1016/S0304-3878(96)00412-9

Barrett, C. B., & Bellemare, M. F. (2010). Reconsidering Conventional Explanations of the

Inverse Productivity – Size Relationship. *World Development*, *38*(1), 88–97. https://doi.org/10.1016/j.worlddev.2009.06.002

Barrett, C. B., Reardon, T., & Webb, P. (2001). *Nonfarm income diversification and household livelihood strategies in rural Africa: Concepts, dynamics, and policy implications*. *Food Policy* (Vol. 26). https://doi.org/10.1016/S0306-9192(01)00014-8

Benjamin, D. (1992). Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica*, *60*(2), 287–322.

Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, *46*(1), 51–84. https://doi.org/10.1016/0304-3878(94)00048-H

Benjamin, D., & Brandt, L. (2002). Property Rights, Labour Markets, and Efficiency in a Transition Economy: The Case of Rural China. *The Canadian Journal of Economics / Revue Canadienne d'Economique*, *35*(4), 689–716. https://doi.org/10.1111/1540-5982.00150

Benjamin, D., Dupas, P., Foster, A., Hamoudi, A., Hotz, V. J., Einav, L., … Financial, C. U. (2014). No Title.

Berry, B. R. A., & Cline, W. R. (1979). AGRARIAN STRUCTURE AND PRODUCTIVITY IN DEVELOPING COUNTRIES.

Besley, T. (1994). How do Market Failures Justify Interventions in Rural Credit Markets ?, *9*(1), 27–47.

Besley, T., & Coate, S. (1995). Group Lending, Repayment Incentives and Social Collateral. *Journal of Development Economics*, *46*(1), 1–18. https://doi.org/10.1016/0304-3878(94)00045-E

Bevis, L. E. M., & Barrett, C. B. (2017). Close to the Edge : Do Behavioral Explanations Account for the Inverse Productivity Relationship ?

Bezu, S., Barrett, C. B., & Holden, S. T. (2012). Does the Nonfarm Economy Offer Pathways for Upward Mobility? Evidence from a Panel Data Study in Ethiopia. *World Development*, *40*(8), 1634–1646. https://doi.org/10.1016/j.worlddev.2012.04.019

Bhalla, Surjit; Roy, P. (1988). Mis-Specification in Farm Productivity Analysis: The Role of Land Quality. *Oxford Economic Papers*, *40*(1), 55–73. https://doi.org/10.2307/2663254

Bhole, B., & Ogden, S. (2010). Group lending and individual lending with strategic default. *Journal of Development Economics*, *91*(2), 348–363. https://doi.org/10.1016/j.jdeveco.2009.06.004

Binswanger, H. P., Deininger, K., & Feder, G. (1995). Power, distortions, revolt and regorm in agricultural land relations. *Handbook of Development Economics*, *3*, 2659–2772. https://doi.org/10.1016/S1573-4471(95)30019-8

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size-productivity relationship. *Journal of Development Economics*, *103*(1), 254–261. https://doi.org/10.1016/j.jdeveco.2013.03.004

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. *Oxford Economic Papers*, *36*(1), 131–145. https://doi.org/10.2307/2662637

Carter, M. R., & Wiebe, K. D. (1990). Access to Capital and Its Impact on Agrarian Structure and Productivity in Kenya. *American Journal of Agricultural Economics*, *72*(5), 1146–1150. https://doi.org/10.2307/1242523

Chapoto, A., Mabiso, A., & Bonsu, A. (2013). Agricultural Commercialization, Land Expansion, and Homegrown Large-scale Farmers: Insights from Ghana, (August).

Chen, Z., Huffman, W. E., & Rozelle, S. (2011). Inverse relationship between productivity and farm size: The case of China. *Contemporary Economic Policy*, *29*(4), 580–592. https://doi.org/10.1111/j.1465-7287.2010.00236.x

Chowdhury, P. R. (2007). Group-Lending with Sequential Financing, Contingent Renewal and Social Capital. *Journal of Development Economics*, *84*(1), 487–506. https://doi.org/10.1016/j.jdeveco.2006.01.001

Collier, P., & Dercon, S. (2014). African Agriculture in 50 Years : Smallholders in a Rapidly Changing World ? *World Development*, *63*(June 2009), 92–101. https://doi.org/10.1016/j.worlddev.2013.10.001

Contreras, D., Gillmore, R., & Puentes, E. (2017). Self-Employment and Queues for Wage Work: Evidence from Chile. *Journal of International Development*, *29*(4), 473–499. https://doi.org/10.1002/jid.3074

Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2009). Microfinance Meets the Market. *Journal of Economic Perspectives*, *23*(1), 167–192. https://doi.org/10.1257/jep.23.1.167

Davis, B., Winters, P., Carletto, G., Covarrubias, K., Quiñones, E. J., Zezza, A., … DiGiuseppe, S. (2010). A Cross-Country Comparison of Rural Income Generating Activities. *World Development*, *38*(1), 48–63. https://doi.org/10.1016/j.worlddev.2009.01.003

Davis, J. R. (2004). The rural non-farm economy, livelihoods and their diversification: Issues and options. *Natural Resources Institute*, 49. https://doi.org/10.2139/ssrn.691821

de Quidt, J., Fetzer, T., & Ghatak, M. (2016). Group lending without joint liability. *Journal of Development Economics*, *121*, 217–236. https://doi.org/10.1016/j.jdeveco.2014.11.006

De Weerdt, J. (2010). Moving out of Poverty in Tanzania: Evidence from Kagera. *Journal of Development Studies*, *46*(2), 331–349. https://doi.org/10.1080/00220380902974393

Deininger, K., Savastano, S., & Xia, F. (2017). Smallholders' land access in Sub-Saharan Africa: A new landscape? *Food Policy*, *67*, 78–92. https://doi.org/10.1016/j.foodpol.2016.09.012

Delgado, C. L., & Siamwalla, A. (1997). Rural economy and farm income diversification. In *Plenary session of the XXIII International conference of agricultural economists, August 10-16* (pp. 3–25).

Dimova, R., & Sen, K. (2010). Is household income diversification a means of survival or a means of accumulation? Panel data evidence from Tanzania. *BWPI Working Paper*, (April 2010).

Dorward, A. (1999). Farm Size and Productivity in Malawian Smallholder Agriculture. *Journal of Development Studies*, *35*(5), 141–161. https://doi.org/10.1080/00220389908422595

FAO. (2008). *The state of food and agriculture, 2008*. *Food and Agriculture Organization of United States*.

Feder, G. (1985). the Relationship Between Farm Size and Farm Productivity:the Role of Family Labor, Supervision and Credit Constrains. *Journal of Development Economics*, *18*(2), 297–313.

Fields, G. S. (2014). Self-employment and Poverty in Developing countries. *IZA World of Labor*, (May), 1–10. https://doi.org/10.15185/izawol.60

Gangopadhyay, S., Ghatak, M., & Lensink, R. (2005). Joint liability lending and the peer selection effect. *Economic Journal*, *115*(506), 1005–1015. https://doi.org/10.1111/j.1468-0297.2005.01029.x

Ghatak, M. (1999). Group Lending, Local Information and Peer Selection. *Journal of Development Economics*, *60*(1), 27–50. https://doi.org/10.1016/S0304-3878(99)00035-8

Ghatak, M. (2000). Screening by the Company You Keep: Joint Liability Lending and the Peer Selection Effect. *Economic Journal*, *110*(465), 601–631. https://doi.org/10.1111/1468-0297.00556

Gourlay, S., Dillon, A., McGee, K., & Oseni, G. (2016). Land Measurement Bias and Its Empirical Implications Evidence from a Validation Exercise. *World Bank Policy Research Working Paper*, *7597*(March).

Haggblade, S., Hazell, P., & Brown, J. (1989). Farm-nonfarm linkages in rural sub-Saharan Africa. *World Development*, *17*(8), 1173–1201. https://doi.org/10.1016/0305-750X(89)90232-5

Haggblade, S., Hazell, P., & Reardon, T. (2010). The Rural Non-farm Economy: Prospects for Growth and Poverty Reduction. *World Development*, *38*(10), 1429–1441. https://doi.org/10.1016/j.worlddev.2009.06.008

Heltberg, R. (1998). Rural Market Imperfections Productivity Relationship : and the Farm Size-Evidence from Pakistan, *26*(10), 1807–1826.

Holden, S. T., & Bezu, S. (2016). Preferences for land sales legalization and land values in Ethiopia. *Land Use Policy*, *52*, 410–421. https://doi.org/10.1016/j.landusepol.2016.01.002

Holden, S. T., & Fisher, M. (n.d.). Can Area Measurement Error Explain the Inverse Farm Size Productivity Relationship ?

International Finance Corporation. (2012). Innovative Agricultural SME Finance Models, (November), 143.

Jacoby, H. G. (1993). Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra. *The Review of Economic Studies*, *60*(4), 903–921. https://doi.org/10.2307/2298105

Jayne, T. S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F. K., … Kachule, R. (2016). Africa's changing farm size distribution patterns: the rise of medium-scale farms. *Agricultural Economics (United Kingdom)*, *47*, 197–214. https://doi.org/10.1111/agec.12308

Jayne, T. S., Chapoto, A., Sitko, N., Nkonde, C., & Chamberlin, J. (2014). Eoreclosing a Smallholder Agricultural Expansion Strategy ?, *67*(2), 35–54.

Karlan, D., & Zinman, J. (2009). Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. *Econometrica*, *77*(6), 1993–2008. https://doi.org/10.3982/ECTA5781

Katzur, T., & Lensink, R. (2012). Group lending with correlated project outcomes. *Economics Letters*, *117*(2), 445–447. https://doi.org/10.1016/j.econlet.2012.06.032

Kawasaki, K. (2010). The costs and benefits of land fragmentation of rice farms in Japan. *Australian Journal of Agricultural and Resource Economics*, *54*(4), 509–526. https://doi.org/10.1111/j.1467-8489.2010.00509.x

Kevane, M. (1996). Agrarian Structure and Agricultural Practice: Typology and Application to Western Sudan. *American Journal of Agricultural Economics*, *78*(1), 236–245. https://doi.org/http://www.blackwellpublishing.com/journal.asp?ref=0002-9092

Kimhi, A. (2006). Plot size and maize productivity in Zambia : is there an inverse relationship ?, *35*, 1–9.

Lamb, R. L. (2003). Inverse productivity : land quality , labor markets , and measurement error, *71*, 71–95. https://doi.org/10.1016/S0304-3878(02)00134-7

Lanjouw, P., Quizon, J., & Sparrow, R. (2001). Non-agricultural Earnings in Peri-Urban Areas of Tanzania: Evidence from Household Survey Data. *Food Policy*, *26*(4), 385–403. https://doi.org/10.1016/S0306-9192(01)00010-0

Larson, D. F. (2012). Should African Rural Development Strategies Depend on Smallholder Farms ? An Exploration of the Inverse Productivity Hypothesis, (September).

Larson, D. F., Otsuka, K., Matsumoto, T., & Kilic, T. (2014). Should African rural development strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis. *Agricultural Economics (United Kingdom)*, *45*(3), 355–367. https://doi.org/10.1111/agec.12070

Lau, L. J., & Yotopoulos, P. a. (1971). A test for relative efficiency and application to Indian agriculture. *The American Economic Review*, *61*(1), 94–109.

Li, G., Feng, Z., & Fan, L. (2013). Re-examining the inverse relationship between farm size and efficiency The empirical evidence in China, *5*(4), 473–488. https://doi.org/10.1108/CAER-09-2011-0108

Maes, J. P., & Reed, L. R. (2012). *State of the Microcredit Summit Campaign Report 2012. Microcredit Summit Campaign: Washington, DC*. Retrieved from http://www.ruralfinance.org/fileadmin/templates/rflc/documents/1253177264086_SOCR2009_English.pdf

Mellor, J. W., & others. (1982). Agricultural Growth-Structures and Patterns, 216–228. Retrieved from http://ageconsearch.umn.edu/bitstream/182449/2/IAAE-CONF-142.pdf

Owusu, V., Abdulai, A., & Abdul-Rahman, S. (2011). Non-farm Work and Food Security among Farm Households in Northern Ghana. *Food Policy*, *36*(2), 108–118. https://doi.org/10.1016/j.foodpol.2010.09.002

Ramana, N. V. (2004). Agricultural Finance by Microfinance Institutions Problems and the Way Forward. *BASIX UNCTAD Geneva*.

Reardon, T. (1997). Using evidence of household income diversification to inform study of the rural nonfarm labor market in Africa. *World Development*, *25*(5), 735–747. https://doi.org/10.1016/S0305-750X(96)00137-4

Reardon, T., Crawford, E., & Kelly, V. (1994). Links between non-farm income and farm investment in African households: adding capital market perspectives. *American Journal of Agricultural Economics*, *76*(5), 1–15. https://doi.org/10.2307/1243412

Sen, A. K. (1962). An Aspect of Indian Agriculture. *Economic Weekly*, *14*(4–6), 243–246.

Sezu, S., & Holden, S. (2014). Are Rural Youth in Euthiopa Abandoning Agriculture? *World Development*, *64*, 259–272.

Stiglitz, J. E. (1990). Peer monitoring and credit markets. *World Bank Economic Review*, *4*(3), 351–366. https://doi.org/10.1093/wber/4.3.351

Stiglitz, J. E., & Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *American Economic Review*, *71*(3), 393–410

Tassel, E. Van. (1999). Group lending under asymmetric information, *60*, 3–25.

Taylor, J. E., & Yunez-Naude, A. (2000). The Returns from Schooling in a Diversified Rural Economy. *American Journal of Agricultural Economics*, *82*(2), 287–297. https://doi.org/10.1111/0002-9092.00025

Varian, H. R. (1990). Monitoring Agents with Other Agents. *Journal of Institutional and Theoretical Economics*, *146*(1, The New Institutional Economics Different Approaches to the Economics of Institutions), 153–174.

Winters, P., Davis, B., Carletto, G., Covarrubias, K., Quiñones, E. J., Zezza, A., … Stamoulis, K. (2009). Assets, Activities and Rural Income Generation: Evidence from a Multicountry Analysis. *World Development*, *37*(9), 1435–1452. https://doi.org/10.1016/j.worlddev.2009.01.010

Winters, P., Davis, B., & Corral, B. (2002). Asset, Activities and Income Genration in Rural Mexico: Factoring in Social and Public Capital, *27*, 139–156.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. *MIT Press*. https://doi.org/10.1515/humr.2003.021

Zaibet, L. T., & Dunn, E. G. (1998). Land Tenure, Farm Size, and Rural Market Participation in Developing Countries: The Case of the Tunisian Olive Sector. *Economic Development and Cultural Change*, *46*(4), 831–848. https://doi.org/10.1086/452376