AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Occupational Choice and Agricultural Labor Exits in Sub-Saharan Africa

Ellen B. McCullough

Invited paper presented at the 5th International Conference of the African Association of Agricultural Economists, September 23-26, 2016, Addis Ababa, Ethiopia

Copyright 2016 by [authors]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Occupational Choice and Agricultural Labor Exits in Sub-Saharan Africa

Ellen B. McCullough
Charles H. Dyson School of Applied Economics and Management
Cornell University

January 15, 2016

Abstract

The process of economic development is characterized by rising output per agricultural worker and the exit of labor from agriculture to other sectors, which together result in rising incomes and falling incidence of poverty. This paper explores the relationship between labor productivity and the occupational choices that underlie the structural transformation process. I model households' decisions to participate in different activities – farming, wage employment, and self employment – through operation of a household non-farm enterprise. I estimate a structural, polytomous model of occupational choice using nationally representative household survey datasets from Tanzania, matched geospatially to several other relevant datasets. Then, I simulate the response of occupational choice to stylized productivity shocks to farming, wage employment, and self employment. I find that participation in farming is not responsive to productivity shocks of any sort. This is most likely because farming participation rates are already quite high. Wage and self employment participation do respond to wage and self employment productivity shocks, respectively. These results highlight the importance of investing in improved smallholder farmer productivity, especially along the intensive margins of farming participation and especially in places with low population density and poor market access, where farming productivity gains are the only ones to impact households.

Acknowledgments

I am grateful to Chris Barrett, Victoria Prowse, David Just, and Nancy Chau for useful comments on earlier versions of this manuscript. I'm also very grateful to Siobhan Murray of the World Bank, who performed several geographic matches with the household survey data using the households' exact geolocations, which are withheld from the publicly accessible datasets in order to preserve respondents' confidentiality. I thank participants of the Western Economics Association International graduate student workshop, other members of Chris Barret's research group, Karen Brooks, Luc Christiaensen, Paul Christian, and Prabhu Pingali for helpful comments. Any errors are my responsibility solely.

Occupational Choice and Agricultural Labor Exits in Sub-Saharan Africa

1 Introduction

Economic development is characterized, almost universally, by rising output per agricultural worker and the movement of labor from agriculture to other sectors, which together result in rising incomes and falling incidence of poverty (Timmer 2009). African countries are mostly in the early stages of this structural transformation process, with large cross-sector productivity gaps and large labor shares still in agriculture (Gollin, Lagakos, and Waugh 2014). Recently, though, growth has been observed in annual output per worker across Sub-Saharan Africa. In the aggregate, labor exits from agriculture to other sectors explain about half of the observed increases in annual output per worker (McMillan and Harttgen 2014).

Growth in agricultural labor productivity is closely associated with poverty reduction, both through the direct effects on the many workers who participate in the agricultural sector, and indirectly, because it leads to growth in non-agriculture sectors and lowers food prices through increased per capita food supplies (De Janvry and Sadoulet 2010). Few debate the importance of farming to poor households, simply because farming is the occupation in which the poor participate with the highest frequency (Christiaensen, Demery, and Kuhl 2011).

What is under debate in Sub-Saharan Africa today, however, is the scope for achieving structural change through smallholder-focused interventions in Sub-Saharan Africa. Some agriculture-skeptics argue that smallholder farmers are weak agents for labor productivity growth of the magnitude necessary to trigger large scale poverty reduction due to low baseline productivity and poor prospects for improving labor productivity within agriculture (e.g., Dercon 2013; Collier and Dercon 2014; Dercon and Gollin 2014). By extension, these skeptics question the role of agricultural interventions in poverty reduction.

Historically, technology-led agricultural productivity growth has been the essential lever for launching structural transformation (e.g., Johnston and Mellor 1961; World Bank 2008; Christiaensen, Demery, and Kuhl 2011). The economy-wide labor productivity growth that accompanied the widespread adoption of high-yielding varieties in South and East Asia and Latin America during the Green Revolution serves as evidence (Evenson and Gollin 2003). And most economists have long rejected the idea that economic growth can be spurred in poor economies while agriculture remains stagnant (Ranis 2004). Nevertheless, development experts have highlighted the importance of interventions that raise labor productivity more generally and in other sectors of the economy.

Today's debate on agriculture's role in overall economic growth in Sub-Saharan Africa hinges on the potential for raising labor productivity in agriculture and in other sectors, and on the impacts of rising labor productivity. In this case, labor productivity in a sector refers to the net returns achieved per worker who participates in that sector. Many types of interventions are associated with rising labor productivity – agricultural technology, improved education, improved infrastructure, etc. The effects of labor productivity enhancing interventions can play out on the intensive margins, for workers who remain in the same occupation as productivity changes, and on the extensive margins, as workers shift occupations in response to productivity changes.

The occupational choice decisions that underlie the structural transformation process play out among many households and farms in heterogeneous settings. While there is empirical regularity in the aggregate relationships between agricultural productivity growth, non-farm growth, agricultural labor exits, overall economic growth, and poverty reduction; the micro-economic processes that underlie these relationships are not well understood (Foster and Rosenzweig 2007). To my knowledge, no empirical study has explicitly examined the micro-economic dimensions of agricultural transformation in Sub-Saharan Africa in the context of occupational choice and technological change.

One major reason for research scarcity on this topic has been, until very recently, lack of datasets that cover relevant farming and non-farming activities of households in both urban and rural areas, including household-managed non-farm enterprises and farms as well as wage labor. Taking advantage of newly available, innovative Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) datasets, I examine the role that improved agricultural technology plays in fostering structural change in African economies. I match LSMS-ISA datasets with a number of other relevant datasets using geo-referenced household locations. I then model annual household returns to participation in farming, non-farm self employment, and wage employment. I find that, in farming, latent labor productivity for households is closely related to household size, the cost of hired labor, land owned, precipitation, and soil nutrients. Self employment latent labor productivity is closely related to peri-urban status, the age of the household head, and ownership of productive assets. And wage employment latent labor productivity is closely related to market access, male headedness of the household, education wtihin the household, and local wage rates.

I use imputed latent labor productivity measures to estimate a household level polytomous model of occupational choice. Predicted occupational choices closely match actual occupational choice shares for observations left out of the estimation sample, and for different sub-populations within the estimation sample. Finally, I simulate the welfare impacts of doubling labor productivity in farming, self employment and wage employment, respectively. I estimate these impacts both along the intensive margins of participation, for households that do not change occupational choices, and along the extensive margins of participation, for households that do change occupational choices. The lion's share of welfare effects are experienced by households that do not shift occupational choice. Participation in farming is overall non-responsive to any of the productivity shocks. Some entry into self employment is seen for the self employment labor productivity shock, and into wage employment for the wage labor productivity shock. Households tend to enter into self and wage employment without exiting farming. The results suggest that agricultural labor productivity growth can lead to large welfare gains because so many households participate in farming, without impacting the probability that households participate in farming.

2 Model

In Sub-Saharan Africa, workers outside of agriculture tend to have higher returns per worker per year (McCullough 2015; Gollin, Lagakos, and Waugh 2014; McMillan and Harttgen 2014). This occurs not because activities outside of agriculture are inherently more productive per hour of labor worked, but because workers outside of agriculture tend to supply more hours of labor per year, while the agricultural sector houses a large reservoir of underemployed workers (McCullough 2015).

Because ability to participate in fuller-employment activities outside of agriculture seems to be a very important determinant of annual worker returns and household expenditures per capita, this paper focuses on the extensive margin of labor supply (choice of occupation) rather than the intensive margin (hours worked per year in each occupation). I use a discrete choice framework not only because the extensive margins of labor supply are of greater interest than the intensive margins in the structural change framework, but also because labor supply is difficult to measure on the intensive margin, with measurement error differing systematically across occupational choices.

While self employment is not as commonly included in occupational choice models as is wage employment, I allow for it as an occupational choice because it is a very common one in the Sub-Saharan African setting. Furthermore, the shift of labor from self employment to wage employment is a key characteristic of the development process (Behrman 1999), and one that is associated with labor productivity growth even when workers do not change sectors as they shift from self to wage employment (McCaig and Pavcnik 2013). Here, self employment does not include own production of household goods, such as child rearing, but rather the operation of household-managed enterprises intended to generate income for the household.[1] I also allow households to participate in multiple activities simultaneously, reflecting the reality of occupational choices observed in this setting (Barrett, Reardon, and Webb 2001; Davis et al. 2010).

I assume a representative household makes its occupational choice of participation ($P = 1$) or non-participation ($P = 0$) in each of three activities: farm operation (F), wage employment (WE),

[1] Virtually all households have at least one member who engages in the production of household goods on the extensive margin, so participation in household good production would not be very interesting to model empirically, at least at the household level.

and self employment (SE). Allowing for each binary option in that triplet, the choice set contains $8 = 2^3$ possibilities. Households derive utility from income per household member (income per adult equivalent is Y_i, with s_i denoting household size).

I use a basic household random utility model with discrete occupational choices to derive estimable equations for structural model parameters, where household utility has both observed and unobserved components, and households are assumed to select the option that brings it the highest utility. Random utility occupational choice models are widely used in labor economics to study the effects of policies and taxes on labor supply (e.g., Keane and Wolpin 1997; Keane and Moffitt 1998; van Soest, Das, and Gong 2002). I model occupational choices at the household level rather than the individual level because, for two of the three activities available to households – farming and self employment – returns are only observable at the household level.[2] In this formulation, utility received by household i from decision j (u_{ij}) is known to the decision-maker (household i) but not to the researcher. Household i chooses option k if and only if $u_{ik} \geq u_{ij} \; \forall \; k \neq j$ and $u_{ik} > u_{ij}$ for at least one $k \neq j$. The household observes its own utility (u_{ij}) across choices, which can be decomposed into a component observed by the researcher (U_{ij}) and an unobserved compoenent (ε_{ij}). Some distributional assumptions made on ε_{ij} are required for maximum likelihood estimation of model parameters. The assumptions made here are discussed in section 3.

The household's decision follows:

$$\max_{P_{ij}=(P^F,P^{WE},P^{SE})} \quad u_i(P_{ij}) = \alpha \cdot Y_{ij}^{.5} + \gamma_j \cdot C_i + \delta_j + \varepsilon_{ij} \tag{1}$$

s.t.

$$Y_i \equiv \frac{1}{s_i}(\Pi_i + R_i) \tag{2}$$

$$\Pi_i \quad \equiv \quad P_i^F \cdot \Pi_i^F \quad + P_i^{SE} \cdot \Pi_i^{SE} \quad + \quad P_i^{WE} \cdot \Pi_i^{WE} \tag{3}$$

$$P_i^a \in \{0,1\} \qquad \forall \quad a \in \{F, WE, SE\} \tag{4}$$

Each household's income (Y_i) is determined by the process defined in equation 2. For each household and occupational choice, the corresponding income is determined by the net returns (profits) to participating in farming, wage employment, and self employment (Π^F, Π^{WE} , and Π^{SE}, respectively). Income also includes non-labor income sources (R), which do not vary across occupational choices and are derived from public and private transfers and other sources. The index variable j refers to each of the 8 unique combinations of participation in the three different activities from which the household selects its occupational choice. The household's choice of occupation is

[2] Modeling individual occupational choices would be an interesting extension, which would allow for closer examination of age and gender patterns. It requires use of some assumptions on how returns to participation in farming and self employment vary within the household.

influenced by additional choice predictors (C_i), such as the household head's parent's education level, that influence a household's selection into an occupation apart from affecting the returns to participation.

There is no leisure consumption in the model. Rather, any dis-utility associated with supplying labor to occupation is reflected in the occupation-specific preference shifters (δ_j).I use a functional form that is monotonically increasing and concave in income.[3] I do not impose *a priori* that utility is decreasing in labor supply. In this model, both willful non-participation in the labor force and unemployment (unsuccessfully attempting to participate in wage employment or other activities) are observed equally, as non-supply of labor. It is not possible to distinguish between these outcomes empirically.

Net returns to participation in an activity are determined by a flexibly specified indirect profit function, which is the dual of a multi-input production function. Consider a total of K farm inputs and outputs, hereafter the netput vector. The profit function takes the input and output price vector as arguments. Here, I use a flexible Generalized Leontief form to specify the returns to activity $a \in \{F, SE, WE\}$. This functional form is advantageous for its flexibility. This process is described in equation 5.

$$\Pi_i^a(P_i^a) = \sum_{k=1}^{K} \beta_k^a x_k^{1/2} + \sum_{k=1}^{K} \sum_{m=1}^{M} \beta_{km}^a (x_k^{1/2} x_m^{1/2}) + e_{ai} \tag{5}$$

Here, x_k refers to the k^{th} variable in the netput vector, which includes the variables that proxy for household shadow prices and relevant context variables that condition household returns to participation in an activity. For example, mean rainfall is used as a control for returns to farming. All of the variables are interacted with each other in the specification. I model returns to sector participation using a stylized profit function rather than an expenditure function or production function for several reasons. First, it allows me to avoid modeling endogenous input use decisions, which then lead to an infinite choice set of inputs used and occupational choices. Rather, the stylized profit function takes prices and key context variables as arguments, which are observable for households that participate in an activity and for those that do not. Second, this approach is relevant in the developing country setting, where there is strong evidence of input and output market frictions, and different households face different prices (Dillon and Barrett 2014; De Janvry, Fafchamps, and Sadoulet 1991; Barrett 2007). Rather than restricting the choice set of occupations available to different households, I allow shadow prices and returns to vary as a function of geographically determined and household specific observable variables. The specific instruments included for shadow prices are discussed in section 4.

[3]The utility function parameters are quite robust to different specifications that do not impose that utility is concave or monotonically increasing in income.

Neither the profit function nor the occupational choice model explicitly includes the fixed costs associated with entering or exiting an occupation. These costs are simply not available in the data. When fixed costs associated with participating in an activity are ignored, discrete choice labor supply models tend to under-predict non-participation in that activity (van Soest, Das, and Gong 2002). Occupation-specific preference shifters (δ_j) can pick up these fixed costs when they are not observed directly. The consequence is that one cannot then disentangle the effects of fixed costs versus alternative-specific preference heterogeneity on participation choices. This must be considered when interpreting the vector of preference shifting parameters.

This model rests on several assumptions. It is a static model, and therefore does not allow for borrowing or saving. Risk and uncertainty are not featured in the profit functions for farm and non-farm enterprises, though risks that households associate with different occupations, and preferences for those risks, can be absorbed into the occupation-specific preference shifters. At this point, general equilibrium effects on wages and prices are not explored. Relevant equilibrium effects in the structural change context include employment effects resulting in changing wage rates, changes in relative output prices due to non-homothetic demand and non-tradability of goods and services consumed (or closed markets). These equilibrium effects are certainly of interest in future studies. The partial equilibrium estimates remain interesting and relevant in the short term.

I estimate a static, rather than dynamic, model because the time interval between survey rounds is fairly short (2-3 years). Transition matrices between the first and second survey rounds for farm, self and wage employment are shown in Figure 1. Overall, a plurality of households never participate in wage labor markets and more households appear to exit wage labor employment than enter it. Conversely, more households enter self employment than exit it. The largest categories are households who do not change participation in any given activity across survey rounds. This is particularly true for farming, where a large majority (about two thirds) of households farmed in both survey rounds. Furthermore, there is not a lot of temporal variation in many of the variables used to estimate returns to participation. The focus is on explaining, in a pooled cross-section, observed patterns of occupational choice within the framework of structural change processes, and then to address how these patterns might change in different circumstances.

3 Estimation

Estimation proceeds in two stages. In the first stage, I estimate profit function parameters. In the second, I estimate the parameters of an occupational choice model using imputed profits. One major challenge in estimating returns to participation is, of course, that returns are only observed for households that elect to participate. I control for selection effects by estimating returns to participation on the full sample of participants and non-participants, using a Heckman selection model (Heckman 1979). For each activity (farming, wage employment, and self employment), I estimate annual returns per household as a function of the x variables described in equation 5 and the selection variables (C) described in equation 1. The estimation equation follows. Equations 6 and 7 are estimated jointly, and u_1 and u_2 are have a correlation coefficient of ρ.

$$\Pi_i^a(P_i^a) \;=\; \sum_{k=1}^{K} \beta_k^a x_k^{1/2} + \sum_{k=1}^{K}\sum_{m=1}^{M} \beta_{km}^a (x_k^{1/2} x_m^{1/2}) + u_1 \tag{6}$$

$$\text{and } \Pi^a \text{ is observed if:} \lambda_j \cdot C_i + u_2 > 0 \tag{7}$$

Then, using the estimated β parameters, I impute returns to participation in farming, wage, and self employment for all households, regardless of their participation. Imputed returns are then used to generate for each household a vector of incomes, one for each of the 8 possible choices. I assume that non-participation in activity a results in a profit of 0 for that activity.

For the second stage estimation, I use a mixed logit model in order to avoid the strong independence from irrelevant assumptions that occur with multinomial logit models. By estimating the preference shifters as random coefficients, I allow for preference heterogeneity and correlation of errors across choices. The random coefficient is δ_{kj}, and it is estimated at the lowest administrative level above the household. Because there is only one observation per household in the sample, it is not tractable to estimate the random coefficient at the household level. This approach is akin to an error components model, with δ_{kj} serving as a structured component of the unobserved utility (Train 2002). The remaining component of the unobserved utility, error term ε_{ij}, is assumed to be independent and identically distributed according to the extreme value (Gumbel) distribution.

After integrating out the random error, the probability of each choice is then given by equation 8. The index term w refers to ward level, which is the lowest administrative level observed in the data. Since there is only one observation per household, I estimate the random parameters at the ward level rather than the household level. Because of the considerable computational demands of exact maximum likelihood estimation, I use maximum simulated likelihood to estimate α, γ, $\bar{\delta}$, and Σ (Gu, Hole, and Knox 2013).

$$\text{Prob}(P_{ijw} = 1) = \int \Big(\frac{e^{\alpha Y_{ij}^{.5} + \gamma_j' c_i} \cdot e^{\delta_{jw}}}{\sum_k e^{\alpha Y_{ik}^{.5} + \gamma_k' c_i} \cdot e^{\delta_{kw}}} \Big) \phi(\delta) d\delta \tag{8}$$

$$\delta \sim \mathcal{N}(\bar{\delta}, \Sigma) \tag{9}$$

Following estimation, I predict choice probabilities for each household and choice by simulating R draws, drawing values of δ_{wj} from the distribution $f(\delta|\bar{\delta}, \Sigma)$ and ε_{ij} from the Gumbel distribution.

The marginal effect of a choice variable or profit function variable on participation in an activity can be derived from equation 8. The parameters for all options appear in the probability equation for each option. It is not straightforward, ex ante, to predict how occupational choices will vary with profit function variables that appear in profit functions for multiple activities. With the functional forms specified, the marginal effects of each profit and choice variable are allowed to vary across households.

4 Data and variables

I estimate the model using household level data from the Tanzanian National Panel Survey, which is part of the LSMS-ISA dataset. These nationally representative, multi-topic and multi-purpose surveys allow for construction of occupational choice, time use, and returns to participation variables. They also include relevant covariates, such as firm and farm inputs and outputs, infrastructure and market access, and household characteristics. I estimate the model using the 2010-11 round of data.

For each household[4], I generate labor supply variables based on individual level, activity-specific time recall variables over the 12 month period preceding the survey date. I then classify households by their corresponding occupational choices $P_i = (P^F, P^E, P^M)$, with participation defined as positive supply of hours by a household member to a given activity and non-participation defined as no supply of labor to the activity. Because I am interested in the annual returns to participation per household, and in the intensive margins of labor supply and occupational choice, I do not differentiate between households who supply different hours of labor to the same activity. If households run an enterprise without any member supplying any labor to it, or if a household operates a farm without any household member supplying any labor to it, I do not consider this participation from a labor supply perspective. . Participation rates and average per capita incomes are tabulated by occupational choice in Table 1.

Besides participation, the other dependent variable in the model is returns to participation. The net returns to self employment in a farm enterprise are the gross value of output, including the value of own-consumed or non-marketed farm products, net costs incurred, which include purchase of inputs, non-farm hired labor, machinery, etc. The net returns to self employment in a non-farm enterprise consist of gross firm procedes over the 12-month recall period minus costs incurred. Wage labor net returns consist of the total gross wages earned during the 12-month recall period by household members who worked as laborers during the period. For ease of interpretation, I convert all local currency based measures to constant international dollars using the purchasing

[4]In this survey, households are defined as groups of individuals who live together and share meals.

power parity conversion factor for private consumption from the World Banks World Development Indicators.

In order to estimate the second stage of the model, it is important to observe all of the first stage covariates not just for households' chosen occupations, but also for non-chosen occupations. The imputed incomes for non-chosen options are reflected in the denominator of each choice probability equation, as depicted in equation 8. Therefore, the first stage estimation of returns to activity participation uses variables that can be observed regardless of the households' occupational choice.

In the agricultural profit functions, the contextual variables are derived from multiple datasets . A general control for agricultural yield potential was created by matching low-technology yield potential estimates from the gridded Global Agro-Ecological Zones dataset with household locations.[5] Yield potential estimates are aggregated across the most country's most widely grown crops[6] using crop area weights generated from crop maps contained in the Harvest Choice dataset.[7] Farm technology is proxied by crop model generated yield predictions for a high technology use scenario and a low technology use scenario. These yield predictions are generated part of the Global Agro-Ecological Zones dataset, using the same crop area weights to create cross-crop measures of technological potential.[8]

Additional agricultural variables from georeferenced sources include mean rainfall during the wettest quarter of the year (from the National Oceanographic and Atmospheric Association), average slope (from the US Geological Survey), soil nutrient retention capacity and workability (from FAO) and the share of land under irrigation (FAO). Using the LSMS-ISA dataset, I generate a locally smoothed estimate of daily median wages paid per hired farm worker as a proxy for labor prices. The smoothing technique used for median wages and for other variables in this section involves generating a variable at the smallest administrative unit above the household. In the case of Tanzania, this is the ward. If at least three observations of the variable are not available at the ward level, I use the next higher administrative level to generate the statistic. In the case of wages, I use median instead of mean wages in order to reduce the influence of outliers. Land prices are proxied by population density and total area of land owned by the household. Availability of machinery is proxied by the availability of tractors, as described by the locally smoothed rate of tractor use by survey respondents. Table 2 contains summaries of the farm profit variables used in profit modeling, tabulated across the eight occupational choices.

The self employment profit variables include the locally smoothed median average annual cost per worker hired by an enterprise, generated from the LSMS-ISA dataset, as a proxy for labor costs. Access to productive capital is proxied by one index of non-agricultural productive assets

[5] http://www.fao.org/nr/gaez/en/

[6] For Tanzania, this list includes maize, paddy rice, cassava, banana, sweet potato, sugar, and cotton.

[7] http://harvestchoice.org/

[8] http://www.fao.org/nr/gaez/en/

and another one for agricultural productive assets. Durable household goods like televisions and mattresses are not included in the index. The prevalence of energy inputs is proxied by nighttime light intensity, taken from the Defense Meteorological Satellite data. Table 3 summarizes enterprise profit variables by occupational choice.

Wage labor profit variables include locally smoothed median annual returns to wage employment and participation rates for the agricultural, industry, and service sectors. These are meant to proxy for the demand for wage employment. The wage labor variables are summarized by occupational choice in Table 4. Additionally, the maximum educational level attained by any household member is also included, on the premise that the most educated household member is the one most likely to secure wage employment outside of the household.

In all of the profit functions, a common set of demographic and geographic variables is included. This includes a dummy equal to one for urban households, as included in the LSMS dataset. It also includes a dummy for peri-urban households, which are assumed to be those that can travel to a population center of at least 500,000 people within two hours. The household's travel time to the nearest town of 500,000 or above, its network distance to the nearest town of 100,000 or above, and its network distance to the nearest major road are also included. Network distances are generated using maps of transport routes . Household level common demographic variables included in profit functions are the number of household members between the ages of 16 and 65, the age of the household head, a dummy variable the equals one if the household head is female, and the average years of education among household adults, excluding indivduals under the age of 25 who may still be enrolled in school. These common profit variables are summarized by occupational choice in Table 5.

Finally, a few additional variables are also included as predictors of households' occupational choices apart from their effects on profits. These include a measure of the household's incoming transfers received from public and private sources. Demographic variables include household size, the share of household members who are dependent (below the age of 15 or over the age of 65), and a dummy that equals one if the household head's father attended school. Finally, I include the average length of the agricultural growing season, which was generated using MODIS global vegetation phenology dataset. These variables are summarized by occupation choice in Table 5. Basic summary statistics for all covariates are contained in the appendix in Table A.1.

5 Results

I estimate the model using the 2010-11 round of Tanzania survey data.Table 6 depicts the marginal effects each variable on annual household returns to participation in farm employment, self employment, and wage employment. The selection variable coefficients are shown in Table 7. The first stage model fit is fairly pretty good, with pseudo-R^2 values of 0.32 for farming, 0.20 for self employment, and 0.43 for wage employment.

For farming, the variables that have a marginal effect on annual profits that is significantly different from zero, after controlling for selection, are: household size (positive), land owned (positive), precipitation during the wettest quarter of the year (positive), hired labor costs (negative), and soil nutrient retention (negative). For self employment, the marginal effects that are significantly different from zero include: non-agricultural productive assets index (positive), peri-urban dummy, with rural as the base case (negative), and the age of the household head (negative). For wage employment , the marginal effects that are significant include the years of education by the most educated household member (positive), the average local wages for a service sector worker (positive), the distance to the nearest major road (negative), and a dummy for female headed households, with male-headed households as the base case (negative). I have included tables in the appendix that depict all of the coefficients for each variable interaction, in matrix form (Table A.5 for farming, Table A.6 for self employment, and Table A.7 for wage employment).

The parameter estimates for the second stage occupational choice model are depicted in Table 8. The income parameter (α), and the parameters that describe the multi-variate normal distribution of the alternative-specific coefficients are shown. The parameter estimates are consistent with scale heterogeneity across choices.

In Table 9, I show the average marginal effect of each profit function variable on the probability of selecting each choice, along with the standard deviation of the marginal effects. The profit function variables affect occupational choice through income effects. They are calculated by differentiating the closed form solution of the probability of participating in an occupation with respect to the each x variable. This expression is evaluated at each data point, drawing simulated δ coefficients using the estimated parameters of the multivariate normal distribution. Similarly in Table 10, I show the average marginal effect of each selection variable on the probability of choosing each occupational choice.

The choice shares predicted by the model match very closely with the participation shares observed in the data. Figure 3 shows the actual particpation shares compared with the predicted shares, along with a box plot of the 5th to the 95th percentiles of prediction probabilities. There is good fit across the entire estimation sample and within specific subsamples, of the dataset. Figure 4 shows a scatterplot of predicted probabilities onto actual participation shares for sixteen different subsets of the population. There is a very good fit between predicted probabilities and actual choice shares within all groups for which fit was checked. Next, I performed a validation exercise by estimating the model on a subset of data, randomly dropping one fifth of the sample enumeration areas. Figure 5 depicts a comparison between predicted probabilities and actual choice shares for the enumeration areas not used in model estimation, showing a fairly close fit.

Occupational choices do not vary greatly over agroclimatic potential, as characterized by a cross-crop index of medium-technology yield levels (Figure 6). Self and wage employment are much more common in high population density areas than in low population density areas. And farming is much more common in low population density areas than in high population density areas (Figure 7). This is consistent with high population density areas featuring larger markets for those operating household non-farm enterprises. High population density areas are also more likely to have surplus labor supply and more wage labor employment opportunities. Households located nearer to population centers of at least 100k people are more likely to participate in wage and self employment than are households located farther from these population centers (Figure 8). Those located further away from population centers are more likely to farm, or to farm in addition to participating in wage or self employment.

6 Policy Simulations

Understanding the sensitivity of occupational choice to labor productivity growth in different sectors has important implications for prioritizing between and targeting delivery of development interventions. A major pathway by which technology-led agricultural labor productivity improvement has resulted in poverty reduction, historically, has been through the eventual reallocation of labor out of agriculture. These pathways are very context-specific, however, and depend on returns to different income-earning opportunities that households face in lieu of, or in addition to, farming.

I simulate three stylized labor productivity shocks in order to understand how these interventions are likely to affect welfare, and the relative importance of shifting occupational choices vis a vis within-sector welfare gains. The first relates to farm labor productivity. I double each households' imputed measure of farm labor productivity. In the second simulation, I double each households' imputed measure of self employment productivity. And, in the third simulation, I double each households' imputed measure of wage labor productivity. For each of the simulations, I also run variants targeting households with low and high agroclimatic potential, low and high population density, and low and high levels of market access, as measured by the travel distance to the nearest population center of at least 500,000 people. With respect to each context variable, "low" and "high" are defined as below and above the median value observed in the sample, respectively.

For each simulation, I generate a new values of returns to participating in each occupational choice. These are based on the first stage prediction of labor productivity for each household, which are conditioned on all of the variables that are arguments in the profit function for each activity. Using the newly simulated imputed income for each occupational choice, I predict new choice probabilites for each household. I compare the probability of each occupational choice between the baseline and the simulated policy intervention. Then, I compare the baseline welfare with simulated welfare, decomposing welfare changes into those that take place along the intensive margin of participation (without changing occupational choice) and those that take place along the extensive margin of participation (due to a change in occupational choice).

For each of the three simulations, figure 9 shows a box plot of the probability difference in participation in farming, self employment, and wage employment, respectively. Farming participation is not very responsive to any of the simulations. Houses, on average, face a small increase in the probability of participating in farming when farm labor productivity is doubled. Increased self and wage labor productivity are associated with very small decreases in farming participation. Self employment participation increases by 1.5 percentage points when self employment income doubles, and it decreases by less than half of a percentage point when wage labor productivity doubles. Similarly, wage labor participation increases by 1.5 percentage points when wage labor productivity doubles, and self employment decreases by less than half of a percentage point. Households tend to respond to productivity shocks by entering into the activity whose productivity was shocked without exiting from baseline activities in which the household participated (figure 10). The self employment simulation is associated with households that only farm adding self employment and households that do nothing adding self employment. Similarly, the wage employment simulation is associated with households that only farm adding wage employment, and households that participate in farming and self employment adding wage employment (figure 11).

The expected welfare effects associated with each simulation closely mirror the probabilities of falling into each category, as depicted in figure 12. The self and wage labor producivity simulations are associated with higher average welfare gains in the population than the farm productivity simulation even though farming participation rates are much higher than self and wage employment participation rates, mostly because self and wage employment comprise a high share of incomes for households that do participate in them.

The probability of changing occupation is examined more closely in different contexts in figures 13, 15, and 17. The relative welfare gains are examined more closely in different contexts in figures 14, 16, and 18. In the places with better market access, households facing self employment productivity shocks are more likely to exit wage labor, and households facing wage labor productivity shocks are more likely to exit self employment. This result could arise from lower underemployment or greater returns to specialization in good market access areas. A closer look at the welfare gains to productivity shocks shows that the farm productivity simulation has the greatest impact on farming households in remote areas (figure 14). The impacts of self and wage employment productivity shocks, on the other hand, are greatest in places with good market access.

As with good market access areas, households with higher population density tended to be more likely to exit wage labor as they entered self employment due to a self employment productivity shock, or to exit self employment as they enter wage labor due to a wage labor productivity shock (figure 15). The expected welfare gains to farming productivity improvement are higher in low population density areas, while the gains to wage and self employment productivity shocks are higher in high population density areas (figure 16). Welfare gains and the probability of changing occupations do not vary over agroclimatic potential (figures 17 and 18).

Together, these findings suggest that improved agricultural productivity has a very important role to play, especially for households in remote areas with low population density. They also suggest that increased labor productivity in wage labor and self employment is likely to pull farming households into those activities. However, entry into wage and self employment is not associated with labor exits from agriculture. Income diversification at the household level is likely to play a critical first step in the structural transformation process, and in the eventual shift of labor out of agriculture.

References

Barrett, C.B. 2007. "Displaced distortions: Financial market failures and seemingly inefficient resource allocation in low-income rural communities." In *Development Economics Between Markets and Institutions: Incentives for growth, food security and sustainable use of the environment*. Wageningen: Wageningen Academic Publishers, vol. 4.

Barrett, C.B., T. Reardon, and P. Webb. 2001. "Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications." *Food Policy* 26:315–331.

Behrman, J. 1999. "Labor markets in developing countries." *Handbook of Labor Econonomics, volume 3* 3:2859–2939.

Christiaensen, L., L. Demery, and J. Kuhl. 2011. "The (evolving) role of agriculture in poverty reduction - An empirical perspective." *Journal of Development Economics* 96:239–254.

Collier, P., and S. Dercon. 2014. "African agriculture in 50 years: smallholders in a rapidly changing world?" *World Development* 63:92–101.

Davis, B., P. Winters, G. Carletto, K. Covarrubias, E.J. Quiñones, A. Zezza, K. Stamoulis, C. Azzarri, and S. DiGiuseppe. 2010. "A Cross-Country Comparison of Rural Income Generating Activities." *World Development* 38:48–63.

De Janvry, A., M. Fafchamps, and E. Sadoulet. 1991. "Peasant household behaviour with missing markets: some paradoxes explained." *Economic Journal* 101:1400–1417.

De Janvry, A., and E. Sadoulet. 2010. "Agricultural Growth and Poverty Reduction : Additional Evidence." *World Bank Reserach Observer* 25:1–20.

Dercon, S. 2013. "Agriculture and development: revisiting the policy narratives." *Agricultural Economics* 44:183–187.

Dercon, S., and D. Gollin. 2014. "Agriculture in African Development: Theories and Strategies." *Annual Review of Resource Economics* 6:471–492.

Dillon, B., and C.B. Barrett. 2014. "Agricultural Factor Markets in Sub-Saharan Africa An Updated View with Formal Tests for Market Failure." *World Bank Policy Research Working Paper* No. 7117.

Evenson, R.E., and D. Gollin. 2003. "Assessing the Impact of the Green Revolution, 1960 to 2000." *Science* 300:758–762.

Foster, A., and M. Rosenzweig. 2007. "Economic development and the decline of agricultural employment." In T. P. Schultz and J. Strauss, eds. *Handbook of Development Economics*. Amsterdam: North-Holland, vol. 4, chap. 47.

Gollin, D., D. Lagakos, and M.E. Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129:939–993.

Gu, Y., A.R. Hole, and S. Knox. 2013. "Fitting the generalized multinomial logit model in Stata." *Stata Journal* 13:382–397.

Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

Johnston, B., and J. Mellor. 1961. "The role of agriculture in economic development." *American Economic Review* 51:566–593.

Keane, M., and R. Moffitt. 1998. "A structural model of multiple welfare program participation and labor supply." *International Economic Review* 39:553–589.

Keane, M.P., and K.I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy* 105:473–522.

McCaig, B., and N. Pavcnik. 2013. "Moving out of Agriculture: Structural Change in Vietnam." *National Bureau of Economic Research Working Paper Series* No. 19616.

McCullough, E. 2015. "Labor productivity and employment gaps in Sub-Saharan Africa." *Policy Research Working Paper No. 7234*, pp. .

McMillan, M., and K. Harttgen. 2014. "What is driving the African Growth Miracle?" *National Bureau of Economic Research Working Paper Series* No. 20077.

Ranis, G. 2004. "The Evolution of Development Thinking: Theory and Policy." Unpublished.

Timmer, C.P. 2009. *A world without agriculture : the structural transformation in historical perspective*. Washington, D.C.: AEI Press.

Train, K.E. 2002. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

van Soest, A., M. Das, and X. Gong. 2002. "A structural labour supply model with flexible preferences." *Journal of Econometrics* 107:345–374.

World Bank. 2008. "World Development Report 2008: Agriculture for Development." Working paper, International Bank for Reconstruction and Development, Washington, D.C.

Table 1: Tabulation of Occupational Choices

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Rural								
Number HHs	110	871	127	553	119	348	68	251
Share HHs	0.0450	0.356	0.0519	0.226	0.0486	0.142	0.0278	0.103
Per capita consumption, usd	1,054	697.2	1,374	803.5	2,026	669.0	1,269	847.6
(sd)	1,208	484.5	1,340	519.3	1,517	494.3	1,033	605.0
Urban								
Number HHs	96	72	303	123	232	65	189	72
Share HHs	0.0833	0.0625	0.263	0.107	0.201	0.0564	0.164	0.0625
Per capita consumption, usd	2,392	921.4	2,072	1,243	2,554	1,431	2,113	1,201
(sd)	1,821	557.3	1,586	1,134	1,821	1,228	1,629	822.9

Tables

Table 2: Summary statistics for farm variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Yield Potential low (cross-crop ind)	0.475	0.423	0.457	0.435	0.475	0.452	0.465	0.427
(sd)	0.139	0.179	0.121	0.169	0.122	0.169	0.103	0.166
Yield Potential high (cross-crop ind)	0.556	0.506	0.543	0.520	0.551	0.538	0.558	0.519
(sd)	0.159	0.205	0.133	0.187	0.137	0.195	0.125	0.190
Rate improved maize seed use (mean smth)	0.136	0.0822	0.170	0.0973	0.164	0.0878	0.184	0.0998
(sd)	0.203	0.145	0.198	0.162	0.197	0.165	0.193	0.174
Cost hired lbr (med smth, USD/day)	3.826	1.797	4.551	2.074	4.212	1.884	4.844	2.116
(sd)	2.705	1.493	2.740	1.709	2.711	1.534	2.750	1.739
Rate of tractor use (mean smth)	0.0129	0.0452	0.00371	0.0376	0.00875	0.0383	0.00761	0.0301
(sd)	0.0452	0.110	0.0216	0.102	0.0457	0.0885	0.0343	0.0805
Land owned (ha, RIGA)	0.230	1.811	0.166	1.874	0.0427	1.419	0.155	1.723
(sd)	0.620	1.979	0.868	2.375	0.246	1.849	0.674	1.956
Mean precip wettest qrtr (mm, NOAA CPC)	599.8	585.9	578.4	572.1	599.9	579.0	565.7	558.8
(sd)	167.9	193.1	128.8	185.5	146.1	191.9	118.7	165.0
Slope (pct, USGS)	3.798	6.066	3.361	5.081	3.749	5.710	3.046	5.407
(sd)	3.277	6.071	2.638	4.838	3.052	5.286	1.909	5.253
Soil nutrient retention capacity (FAO)	1.330	1.515	1.407	1.544	1.425	1.525	1.440	1.644
(sd)	0.770	0.853	1.109	0.933	0.994	0.719	1.240	1.101
Soil workability (FAO)	1.272	1.682	1.228	1.675	1.208	1.535	1.249	1.731
(sd)	0.902	1.119	1.140	1.184	1.008	0.854	1.269	1.246
Share land irrigated (percent, FAO)	0.00303	0.00765	0.00609	0.00488	0.00385	0.00753	0.00297	0.00436
(sd)	0.0160	0.0319	0.0322	0.0220	0.0249	0.0351	0.0161	0.0212

Table 3: Summary statistics for self-employment variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Cost/hired worker (med smth, USD)	1,034	521.3	1,283	549.8	1,354	542.5	1,347	539.2
(sd)	932.8	564.9	1,037	593.2	1,100	581.5	1,081	579.7
Nighttime light ave coverage(DMSP F16)	1,392	34.74	1,778	145.8	1,868	71.59	2,042	207.1
(sd)	1,804	162.4	1,771	629.3	1,741	312.4	1,753	739.2
Financial service available	0.549	0.348	0.579	0.395	0.593	0.429	0.576	0.399
(sd)	0.499	0.477	0.494	0.489	0.492	0.495	0.495	0.491
Productive non-ag asset ind, fact 1	0.675	0.644	1.079	0.885	1.085	0.705	1.279	1.066
(sd)	0.647	0.543	0.813	0.698	0.835	0.589	0.883	0.824
Productive ag-related asset ind, fact 1	0.0234	0.0331	0.0249	0.0397	0.0256	0.0330	0.0290	0.0382
(sd)	0.0155	0.0401	0.0221	0.0598	0.0295	0.0437	0.0399	0.0544
Net returns from ent (USD)	0	0	4,861	2,052	0	0	4,514	1,888
(sd)	0	0	5,525	3,447	0	0	5,370	3,363

Table 4: Summary statistics for wage employment variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Max educ in hh (yrs)	7.908	6.133	9.416	7.428	10.64	7.421	10.56	8.576
(sd)	5.266	3.906	3.880	3.660	4.444	4.048	3.974	3.720
Nighttime light intensity(DMSP F16)	15.36	1.371	19.81	2.710	20.84	1.848	22.63	3.763
(sd)	17.47	3.081	16.49	6.786	16.03	4.135	15.93	7.888
Returns/ag worker (med smth, USD)	687.7	259.2	726.5	320.0	736.9	343.6	744.6	348.1
(sd)	583.2	333.9	504.2	384.2	568.1	443.5	528.3	421.3
Returns/ind worker (med smth, USD)	1,979	978.0	2,485	1,127	2,462	1,028	2,570	1,173
(sd)	1,302	731.3	1,316	989.4	1,290	848.3	1,412	966.9
Returns/ser worker (med smth, USD)	2,576	1,447	2,874	1,515	3,034	1,382	3,020	1,667
(sd)	2,507	1,785	2,650	1,795	2,623	1,781	2,665	2,188
Partpn in ag emplmt (med smth sh)	0.0562	0.123	0.0263	0.124	0.0308	0.209	0.0321	0.200
(sd)	0.0990	0.134	0.0734	0.139	0.0776	0.181	0.0755	0.192
Partpn in ind emplmt (med smth sh)	0.0740	0.0281	0.0958	0.0435	0.107	0.0612	0.125	0.0621
(sd)	0.0893	0.0661	0.103	0.0789	0.101	0.0991	0.103	0.0956
Partpn in ser emplmt (med smth sh)	0.279	0.116	0.344	0.145	0.441	0.203	0.423	0.243
(sd)	0.193	0.131	0.186	0.146	0.191	0.175	0.175	0.195
Net returns from market (USD)	1,347	113.2	701.0	180.5	4,992	1,701	4,418	2,071
(sd)	3,641	1,008	2,559	1,486	5,170	3,394	5,132	3,753

Table 5: Summary statistics for profit choice variables and controls included in profit models for all occupations, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Urban	0.466	0.0764	0.705	0.182	0.661	0.157	0.735	0.223
(sd)	0.500	0.266	0.457	0.386	0.474	0.365	0.442	0.417
Peri-urban dummy	0.0388	0.0286	0.0837	0.0607	0.0598	0.0557	0.0428	0.0929
(sd)	0.194	0.167	0.277	0.239	0.238	0.230	0.203	0.291
Hrs travel to nrst town >500k (LSMS-ISA)	5.120	8.288	3.632	7.658	3.793	6.986	2.766	6.790
(sd)	4.770	4.522	4.342	4.624	4.296	4.114	3.521	4.412
Network dist to nrst town >100k (km, LSMS-ISA)	71.94	141.6	49.97	134.2	44.90	131.6	39.00	136.4
(sd)	83.00	94.00	76.51	99.30	72.28	99.87	67.83	101.9
Dist nrst major rd (km, LSMS-ISA)	13.65	24.88	5.836	22.32	6.480	20.59	4.849	16.96
(sd)	23.70	25.37	13.88	25.82	15.73	22.91	12.37	21.20
People per square km, 2005 (ln, HC)	6.092	4.453	6.944	4.660	7.210	4.764	7.296	4.890
(sd)	2.511	1.156	1.996	1.327	1.815	1.153	1.776	1.320
Number hh members 16-65	1.830	2.583	2.581	2.966	2.587	2.976	3.661	3.706
(sd)	1.192	1.593	1.482	1.859	1.574	1.529	2.044	1.947
Age of head	44.96	52.17	41.42	46.75	38.47	47.16	43.37	47.40
(sd)	18.17	16.76	13.60	14.48	12.65	14.90	13.08	14.39
Female head	0.432	0.238	0.323	0.250	0.211	0.228	0.233	0.180
(sd)	0.497	0.426	0.468	0.433	0.408	0.420	0.424	0.384
Yrs educ, adults (ave)	6.918	4.507	8.178	5.894	9.456	5.737	8.871	6.525
(sd)	4.908	3.336	3.424	3.279	4.248	3.644	3.660	3.414
Transfers recvd (USD)	117.2	73.55	67.24	67.25	81.81	83.84	82.88	71.91
(sd)	175.3	137.5	152.5	129.7	184.0	136.2	195.7	143.4
Household size	3.364	5.382	4.421	5.889	4.034	5.872	5.630	6.755
(sd)	2.261	2.946	2.509	3.390	2.634	2.830	2.954	3.213
HH dependent share	0.303	0.414	0.247	0.359	0.176	0.347	0.215	0.323
(sd)	0.334	0.268	0.242	0.233	0.209	0.227	0.199	0.206
Head's father attended school	0.461	0.310	0.665	0.395	0.701	0.429	0.642	0.498
(sd)	0.500	0.463	0.472	0.489	0.459	0.495	0.480	0.501
Years educ, head	7.083	4.536	8.281	5.914	9.744	5.690	8.844	6.570
(sd)	5.252	3.918	3.924	3.900	4.615	4.187	4.540	4.393
Ave length of season (days, MOD12Q2)	182.4	174.0	183.8	172.7	184.1	179.2	186.4	177.6
(sd)	15.85	24.98	16.68	25.02	17.15	26.03	15.69	25.62

Table 6: Marginal effects of profit function variables on returns to participation in farming, self employment, and wage employment (in USD per year). These are based on the Generalized Leontief profit function specification estimated using a Heckman selection model. The selection parameter estimates are shown in Table 7.

	Farm Margins at means	(SE)	Self Emplm't Margins at means	(SE)	Wage Emplm't Margins at means	(SE)
Urban (sq rt)	135.842	105.371	-710.422	952.864	-323.012	746.940
Peri-urban dummy (sq rt)	-11.949	678.455	-5,809.333	4,826.420	1,515.902	3,555.486
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-87.897	53.434	-987.280	531.530	92.147	328.373
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	14.805	8.933	-28.612	91.065	-64.368	59.958
Dist nrst major rd (km, LSMS-ISA) (sq rt)	15.000	18.173	77.699	234.559	-237.193*	114.297
People per square km, 2005 (ln, HC) (sq rt)	-204.729	122.090	1,013.498	1,526.430	348.260	897.252
Number hh members 16-65 (sq rt)	360.126**	64.025	804.815	750.843	508.537	395.601
Age of head (sq rt)	43.922	25.068	-419.253	273.632	42.620	145.741
Female head (sq rt)	-54.241	56.824	-201.497	538.188	-797.558*	341.084
Yrs educ, adults (ave) (sq rt)	66.084	35.784	-42.117	379.546	-154.296	624.761
Yield Potential low (cross-crop ind) (sq rt)	-185.002	460.278				
Yield Potential high (cross-crop ind) (sq rt)	-94.349	436.320				
Rate improved maize seed use (mean smth) (sq rt)	-184.248	136.459				
Cost hired lbr (med smth, USD/day) (sq rt)	-121.834	67.761				
Rate of tractor use (mean smth) (sq rt)	407.344	323.785				
Land owned (ha, RIGA) (sq rt)	383.472**	40.049				
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	30.411*	12.090				
Slope (pct, USGS) (sq rt)	59.790	43.512				
Soil nutrient retention capacity (FAO) (sq rt)	-217.076	119.507				
Soil workability (FAO) (sq rt)	216.567	130.760				
Share land irrigated (percent, FAO) (sq rt)	2,556.904**	899.619				
Cost/hired worker (med smth, USD) (sq rt)			24.064	28.378		
Nighttime light ave coverage(DMSP F16) (sq rt)			-30.658	44.336		
Financial service available (sq rt)			1,003.185	513.653		
Productive non-ag asset ind, fact 1 (sq rt)			3,171.157**	1,041.906		
Productive ag-related asset ind, fact 1 (sq rt)			-4,581.303	7,078.165		
Max educ in hh (yrs) (sq rt)					1,980.455**	586.246
Nighttime light intensity(DMSP F16) (sq rt)					49.682	163.018
Returns/ag worker (med smth, USD) (sq rt)					-16.314	35.770
Returns/ind worker (med smth, USD) (sq rt)					-2.169	17.224
Returns/ser worker (med smth, USD) (sq rt)					12.166	11.823
Partpn in ag emplmt (med smth sh) (sq rt)					-313.959	1,021.808
Partpn in ind emplmt (med smth sh) (sq rt)					-359.235	1,006.978
Partpn in ser emplmt (med smth sh) (sq rt)					254.652	803.682
N	3,599		3,599		3,599	
N (non-censored)	2407		447		1342	
Mean profits (USD/yr)	1123.93		3341.16		3767.53	
R2 (adj)	0.327		0.256		0.362	

* $p < 0.05$; ** $p < 0.01$

Table 7: Selection variable coefficients for returns to participation by activity (Generalized Leontief profit function specification with selection). Parameters were estimated using a Heckman Selection Model. The profit accompanying function parameters are shown in Table 6.

	Farm Select	(SE)	Self Emplm't Select	(SE)	Wage Emplm't Select	(SE)
Transfers recvd (USD)	-0.001**	0.000	-0.000*	0.000	0.000	0.000
Household size	0.099**	0.010	0.085**	0.009	0.053**	0.008
HH dependent share	0.623**	0.106	-0.124	0.126	-0.797**	0.101
Head's father attended school	-0.035	0.055	0.153*	0.062	0.112*	0.048
Years educ, head	-0.063**	0.006	0.009	0.007	0.049**	0.006
Urban	-1.391**	0.055	0.186**	0.063	0.558**	0.051
Peri-urban dummy	-0.619**	0.103	0.323**	0.114	0.367**	0.098
Ave length of season (days, MOD12Q2)	-0.011**	0.001	-0.002*	0.001	0.004**	0.001
Lambda	-390.54		-977.66		-1,697.31	
Sigma	800.66		3,568.35		3,804.23	
P value comparison test	0.00		0.00		0.00	
N	3,599		3,599		3,599	
N (censored)	1,192		3,152		2,257	
N (non-censored)	2407		447		1342	

* $p < 0.05$; ** $p < 0.01$

Table 8: Second stage occupational choice model parameter coefficients. The base case is non-participation in all activities. The left column shoes coefficients in the utility function for income and income squared. The matrix below it shows alternate-specific coefficients for each of the selection variables, the mean random coefficient for each alternative, and the standard deviation for each alternative (the diagonal of the variance-covariance matrix desctibing the multi-variate normal distribution of δ.

	MSL						
Income per capita sq rt. (predicted '000)	0.414*						
	(0.175)						
Income per capita sq rt. X urban	-0.290						
	(.065)						
	farm	self	farm,self	wage	wage,farm	wage,self	farm,wage,self
Transfers recvd (USD)	-2.199**	-1.231*	-2.504**	-0.409	-1.622**	-0.775	-2.021**
Household size	0.477**	0.335**	0.566**	0.293**	0.568**	0.551**	0.673**
HH dependent share	0.404	-0.904*	-0.339	-2.072**	-0.591	-2.550**	-1.371**
Head's father attended school	-0.016	0.549**	0.126	0.540**	0.295	0.466*	0.481*
Years educ, head	-0.091**	0.002	-0.034	0.074**	-0.052*	0.015	-0.021
Urban	-1.194**	1.266**	-0.239	0.791**	-0.403	1.373**	0.081
Peri-urban	-1.370**	1.736**	-0.377	1.252**	-0.790	0.985	-0.013
Ave length of season	-0.019**	-0.005	-0.024**	-0.004	-0.012**	0.001	-0.017**
δ (mean)	3.054**	-0.614	2.880**	-0.992	0.375	-3.182**	0.188
δ (sd)	(0.791)	(0.884)	(0.881)	(0.920)	(0.846)	(1.045)	(0.896)
N	3599						
Log-likelihood	-5773.68						
Pseudo R2	.1225						
AIC	907.37						
BIC	1489.08						

Table 9: Average marginal effect of each profit function variable on the probability of participating in each occupation. The standard deviation of each average marginal effect estimate is shown below in parentheses.

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
Common Profit Function Variables									
Urban	-0.00275	-0.177	0.0798	0.0164	0.0199	-0.00865	0.0542	0.0252	0.00141
	(0.0446)	(0.0933)	(0.0635)	(0.0742)	(0.0405)	(0.0385)	(0.0493)	(0.0436)	(0.0952)
Peri-urban	0.000879	-0.168	0.112	-0.0183	0.0636	-0.0217	0.0186	0.0245	0.00160
	(0.0602)	(0.0875)	(0.0944)	(0.0818)	(0.0648)	(0.0512)	(0.0384)	(0.0504)	(0.104)
Travel time (>500k)	0.000359	0.0000617	-0.00260	-0.000973	0.00210	0.000813	-0.000198	0.0000514	-0.0000763
	(0.00189)	(0.00510)	(0.0111)	(0.00608)	(0.00885)	(0.00450)	(0.00115)	(0.000869)	(0.00632)
Network Dist (>100k)	-0.00000177	0.0000306	-0.0000108	-0.0000136	0.0000103	-0.00000314	-0.000000205	-0.00000845	0.000000410
	(0.0000396)	(0.000114)	(0.000118)	(0.0000911)	(0.000128)	(0.0000793)	(0.0000197)	(0.0000247)	(0.0000899)
Dist to major rd	0.000118	0.000464	0.000185	0.000411	-0.000457	-0.000412	-0.000260	-0.0000548	-0.00000801
	(0.00196)	(0.0102)	(0.00240)	(0.00494)	(0.00306)	(0.00593)	(0.00124)	(0.00354)	(0.00508)
Pop density	0.0000761	-0.0000872	-0.000549	0.00310	-0.000483	-0.00107	-0.000144	-0.000620	0.0000284
	(0.0186)	(0.0374)	(0.00696)	(0.219)	(0.0507)	(0.0718)	(0.0101)	(0.0353)	(0.0884)
HH indep pop	-0.0164	0.0521	-0.00224	-0.0170	-0.00151	-0.0167	-0.000778	-0.00409	-0.0000129
	(0.165)	(0.414)	(0.0950)	(0.185)	(0.0215)	(0.153)	(0.00857)	(0.0527)	(0.185)
Age hh head	-0.0000162	0.0000402	-0.0000365	-0.0000901	0.0000411	0.0000635	-0.00000387	-0.00000132	0.000000378
	(0.000222)	(0.000407)	(0.000249)	(0.000491)	(0.000214)	(0.000215)	(0.0000544)	(0.000121)	(0.000287)
Female head dummy	0.00120	0.00120	-0.000352	0.000657	0.000108	-0.00161	-0.000548	-0.000822	-0.0000964
	(0.00490)	(0.00866)	(0.00664)	(0.00821)	(0.00457)	(0.00579)	(0.00164)	(0.00282)	(0.00598)
Ave hh educ (yrs)	0.0398	0.213	0.0271	0.107	-0.0382	-0.253	-0.0178	-0.0738	-0.00230
	(0.173)	(0.686)	(0.146)	(0.340)	(0.186)	(0.861)	(0.0780)	(0.224)	(0.458)
Farm Profit Function Variables									
Low tech yield potential	-0.000520	-0.00592	-0.0000453	-0.000662	0.000365	0.00436	0.000191	0.00183	-0.0000105
	(0.0154)	(0.188)	(0.00753)	(0.0478)	(0.0140)	(0.101)	(0.00423)	(0.0477)	(0.0815)
High tech yield potential	0.00175	0.000149	0.000462	0.00112	0.000170	-0.00246	0.0000616	-0.000710	-0.0000336
	(0.0138)	(0.129)	(0.00597)	(0.0343)	(0.00959)	(0.0700)	(0.00328)	(0.0328)	(0.0563)
Maize tech use (share)	0.0146	-0.0366	0.0345	-0.0245	0.0123	-0.00404	0.0123	0.00635	0.00115
	(0.729)	(1.852)	(0.380)	(0.651)	(0.277)	(0.480)	(0.143)	(0.247)	(0.784)
Farm wages (cost/day)	0.00000548	-0.000144	0.00000969	0.0000299	0.0000387	0.0000178	0.0000196	0.0000495	0.00000349
	(0.000897)	(0.00224)	(0.000481)	(0.000837)	(0.000342)	(0.000771)	(0.000147)	(0.000300)	(0.000976)
Tractor use (share)	-0.354	1.424	-0.0931	-0.236	-0.203	-0.181	-0.137	-0.283	0.00981
	(1.948)	(6.339)	(1.797)	(2.450)	(1.157)	(1.652)	(0.511)	(0.860)	(2.781)
Land owned (ha)	-0.0311	0.0944	-0.0225	-0.00146	-0.0216	-0.00108	-0.0107	-0.00685	0.00158
	(0.177)	(0.409)	(0.0745)	(0.160)	(0.0746)	(0.0950)	(0.0280)	(0.0437)	(0.179)
Rainfall	-0.000000755	0.00000383	-0.00000102	-0.000000347	-0.000000770	-1.90e-08	-0.000000470	-0.000000294	5.84e-08
	(0.00000604)	(0.0000207)	(0.00000342)	(0.00000906)	(0.00000299)	(0.00000526)	(0.00000139)	(0.00000263)	(0.00000890)

(*Continued on next page*)

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
(*Continued from previous page*)									
Slope	-0.0000353	0.000192	-0.0000293	-0.0000232	-0.0000373	-0.0000168	-0.0000222	-0.0000245	0.00000224
	(0.000398)	(0.00257)	(0.000238)	(0.000566)	(0.000224)	(0.000775)	(0.000127)	(0.000561)	(0.00103)
Soil nutrients	0.0307	-0.157	0.0378	0.0219	0.0229	0.00803	0.0127	0.0128	-0.00291
	(0.253)	(1.128)	(0.316)	(0.434)	(0.164)	(0.265)	(0.0679)	(0.127)	(0.478)
Soil workability	-0.0280	0.157	-0.0397	-0.0224	-0.0235	-0.00619	-0.0127	-0.0127	0.00299
	(0.221)	(1.172)	(0.360)	(0.446)	(0.179)	(0.280)	(0.0701)	(0.128)	(0.498)
Irrigation (share)	-2.100	11.31	-4.033	-0.307	-2.506	0.409	-1.378	-0.844	0.180
	(10.30)	(35.56)	(16.94)	(14.17)	(8.618)	(17.63)	(3.581)	(4.815)	(17.72)
Self Employment Profit Function Variables									
Cost per firm worker	-0.000000968	-0.00000453	0.00000163	0.00000521	-0.00000116	-0.00000180	0.000000389	0.00000102	4.41e-08
	(0.00000458)	(0.0000170)	(0.00000748)	(0.0000209)	(0.00000428)	(0.00000665)	(0.00000234)	(0.00000424)	(0.0000111)
Nighttime lights	-0.0000372	-0.000318	0.0000550	0.000338	-0.0000253	-0.000131	0.0000199	0.0000920	0.00000231
	(0.000188)	(0.000784)	(0.000247)	(0.000893)	(0.000142)	(0.000324)	(0.0000779)	(0.000225)	(0.000504)
Financial services available	0.000142	0.0000488	-0.000200	-0.0000363	0.0000105	-0.000104	0.00000308	0.0000677	-0.0000174
	(0.00189)	(0.00607)	(0.00368)	(0.00679)	(0.00201)	(0.00231)	(0.000713)	(0.00157)	(0.00383)
Asset index (non-ag)	0.00280	-0.00501	0.000742	0.0182	-0.00851	-0.00753	0.00176	0.000132	0.000207
	(0.569)	(0.668)	(0.407)	(1.188)	(0.121)	(0.190)	(0.0466)	(0.175)	(0.551)
Asset index (ag)	-0.0236	-0.0906	0.0421	0.0975	-0.0201	-0.0295	0.00696	0.0170	0.00165
	(0.0954)	(0.425)	(0.143)	(0.454)	(0.0605)	(0.105)	(0.0224)	(0.0595)	(0.244)
Wage Employment Profit Function Variables									
Max educ in hh (yrs)	-0.0226	-0.119	-0.0111	-0.0483	0.0192	0.130	0.00958	0.0397	0.00123
	(0.0993)	(0.387)	(0.0599)	(0.159)	(0.0926)	(0.455)	(0.0417)	(0.122)	(0.244)
Nighttime lights	-0.00323	-0.0238	-0.000162	-0.0128	0.00445	0.0245	0.00127	0.00900	0.000132
	(0.0211)	(0.0619)	(0.0246)	(0.0396)	(0.0402)	(0.0665)	(0.00944)	(0.0295)	(0.0440)
Wages (ag worker)	-0.000000355	0.000000220	-0.000000282	0.000000238	0.000000566	-0.000000283	0.000000129	-0.000000171	2.82e-08
	(0.00000444)	(0.0000145)	(0.00000558)	(0.00000910)	(0.00000861)	(0.0000165)	(0.00000186)	(0.00000649)	(0.00000979)
Wages (ind worker)	0.000000654	0.00000231	0.000000454	0.00000153	-0.00000109	-0.00000238	-0.000000271	-0.00000105	-2.24e-08
	(0.00000330)	(0.00000872)	(0.00000314)	(0.00000588)	(0.00000575)	(0.00000928)	(0.00000127)	(0.00000415)	(0.00000609)
Wages (service worker)	-0.000000238	-0.000000674	-0.000000382	-0.000000586	0.000000597	0.000000745	0.000000198	0.000000349	1.62e-08
	(0.00000178)	(0.00000360)	(0.00000168)	(0.00000238)	(0.00000281)	(0.00000408)	(0.000000690)	(0.00000171)	(0.00000265)
Participation share (ag)	-0.242	-0.413	-0.532	-0.269	0.727	0.466	0.206	0.157	0.0276
	(0.954)	(1.572)	(2.031)	(1.147)	(2.611)	(2.022)	(0.531)	(0.764)	(1.686)
Participation share (ind)	0.435	2.122	0.294	1.255	-0.724	-2.290	-0.185	-0.861	-0.0233
	(1.678)	(4.157)	(1.410)	(2.369)	(2.726)	(4.889)	(0.589)	(1.583)	(3.114)
Participation share (ser)	-0.0289	-0.0584	-0.000996	0.0263	0.00570	0.0801	-0.00265	-0.0228	0.00160
	(0.432)	(1.833)	(0.361)	(0.914)	(0.618)	(2.216)	(0.179)	(0.644)	(1.159)

Table 10: Average marginal effect of each selection variable on the probability of participating in each occupation variable margins. The standard deviation of each average marginal effect estimate is shown below in parentheses.

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
Income transfers (‘000 USD)	0.0706	-0.0565	-0.00942	-0.110	0.0649	0.0298	0.0264	-0.0159	-1.35e-10
	(0.0681)	(0.0482)	(0.0643)	(0.0510)	(0.0552)	(0.0415)	(0.0341)	(0.0280)	(0.0765)
HH Size	-0.0217	-0.00683	-0.00717	0.0119	-0.00903	0.00751	0.0105	0.0148	7.23e-11
	(0.0206)	(0.0125)	(0.00998)	(0.00745)	(0.00935)	(0.00506)	(0.0161)	(0.00856)	(0.0172)
HH Dependency share	0.0470	0.174	0.0304	0.0137	-0.0820	-0.0178	-0.0940	-0.0714	1.00e-10
	(0.0694)	(0.0946)	(0.0834)	(0.0613)	(0.0775)	(0.0421)	(0.0944)	(0.0597)	(0.111)
Head’s father school (dum)	-0.0166	-0.0442	0.0200	-0.0121	0.0146	0.0110	0.00508	0.0221	-2.65e-12
	(0.0221)	(0.0242)	(0.0167)	(0.0155)	(0.0132)	(0.0127)	(0.00574)	(0.0171)	(0.0271)
Head’s education (years)	0.000580	-0.0101	-0.000103	0.00192	0.00647	-0.000929	0.000697	0.00151	1.35e-12
	(0.00228)	(0.00551)	(0.00344)	(0.00413)	(0.00592)	(0.00198)	(0.00178)	(0.00228)	(0.00575)
Urban	-0.00275	-0.177	0.0798	0.0164	0.0199	-0.00865	0.0542	0.0252	0.00141
	(0.0446)	(0.0933)	(0.0635)	(0.0742)	(0.0405)	(0.0385)	(0.0493)	(0.0436)	(0.0952)
Peri-Urban	0.000879	-0.168	0.112	-0.0183	0.0636	-0.0217	0.0186	0.0245	0.00160
	(0.0602)	(0.0875)	(0.0944)	(0.0818)	(0.0648)	(0.0512)	(0.0384)	(0.0504)	(0.104)
Growing season length	0.000513	-0.000536	0.000260	-0.00141	0.000286	0.000439	0.000608	-0.000164	-4.19e-13
	(0.000528)	(0.000485)	(0.000507)	(0.000631)	(0.000369)	(0.000505)	(0.000606)	(0.000327)	(0.000814)

Figures

Figure 1: Transition matrices for households between 2008 and 2011, by activity. Farming is shown, followed by self employment and wage employment.

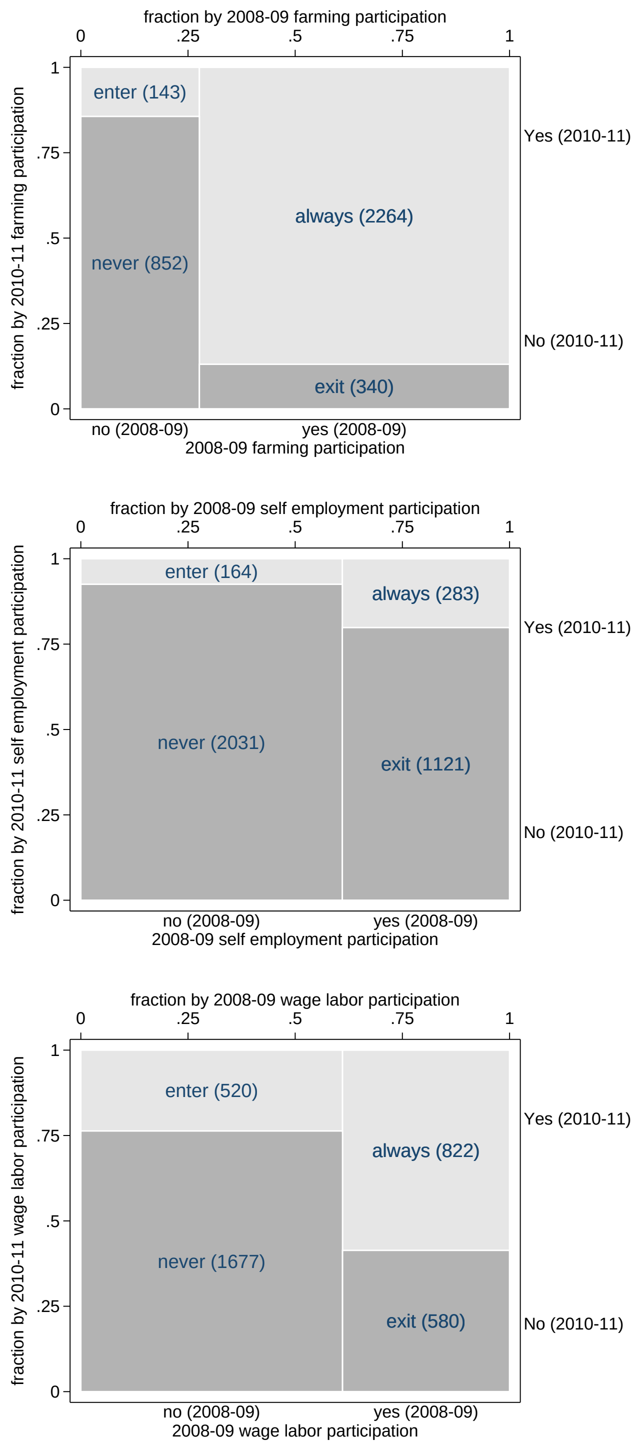


Figure 2: Average labor productivity (annual net returns per worker) across occupational choices.

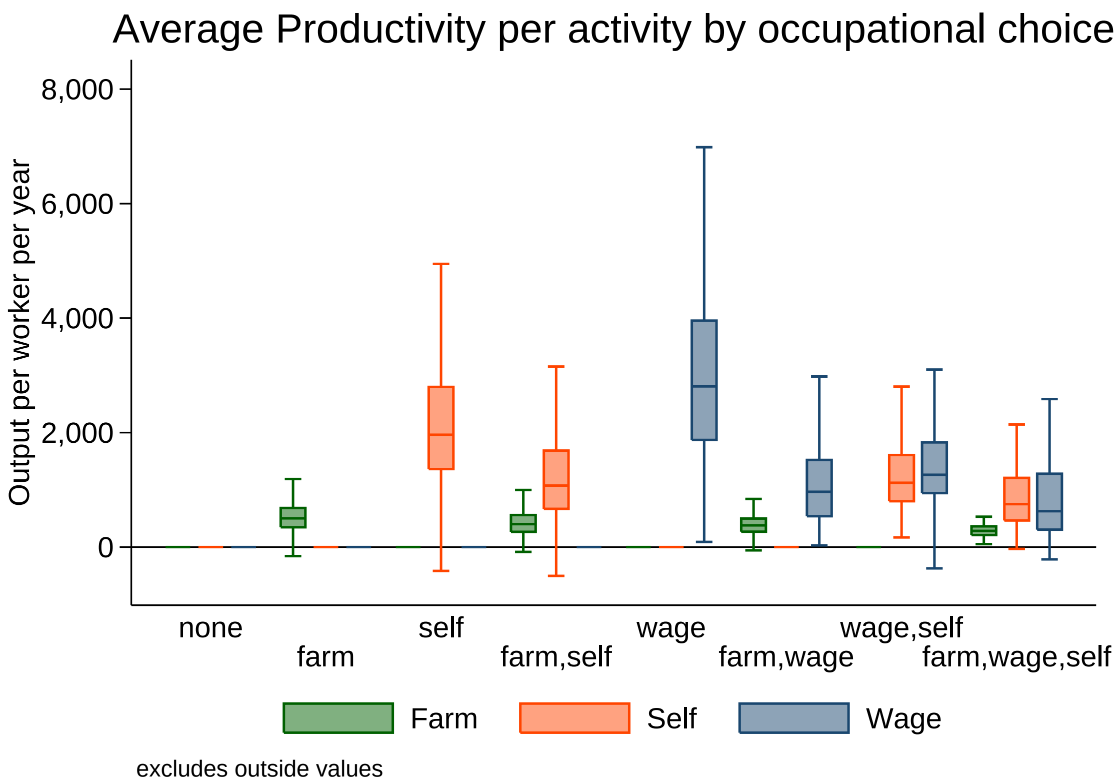


Figure 3: Comparison of predicted choice probabilities and actual choice shares.

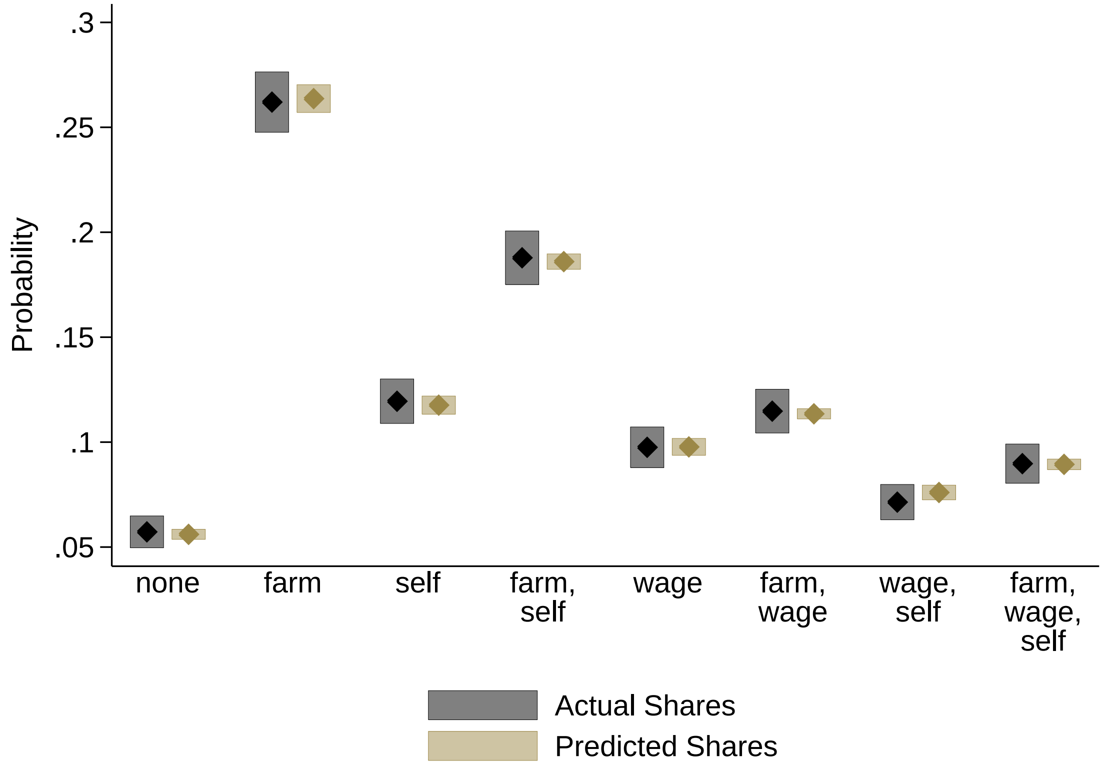


Figure 4: Comparison of predicted choice probabilities and actual choice shares by population sub-groups.

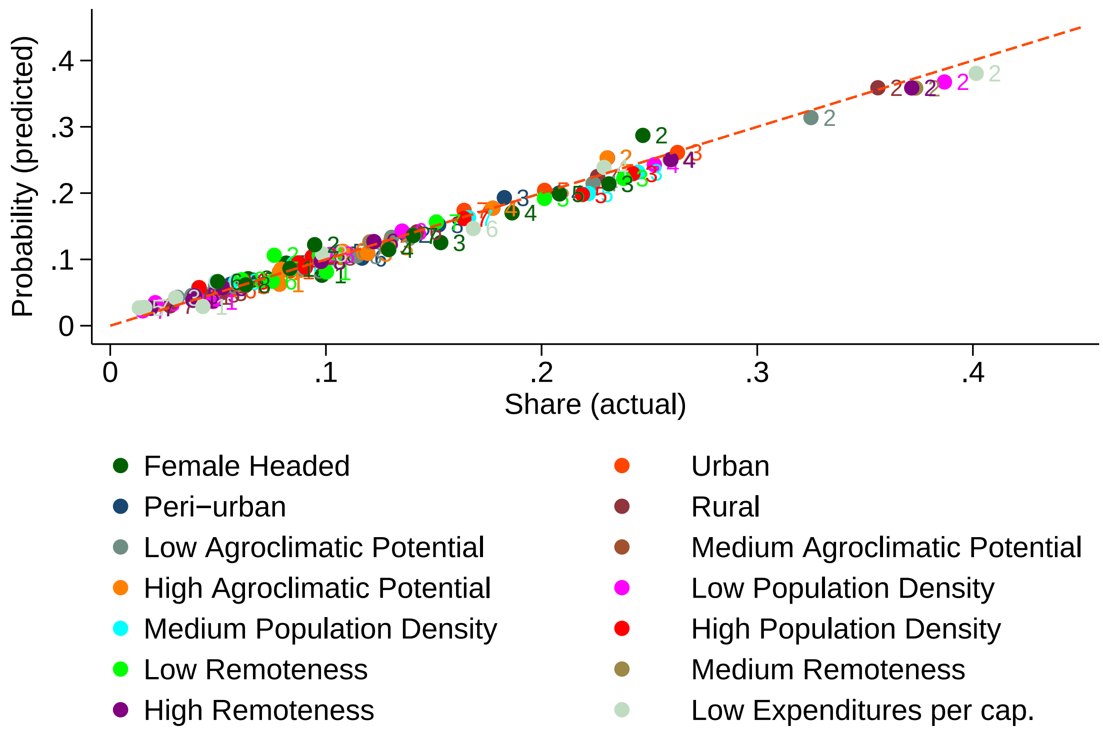


Figure 5: Validation exercise: comparison of predicted choice probabilities with actual choice shares for 20% of sample that was randomly omitted from the estimation sample.

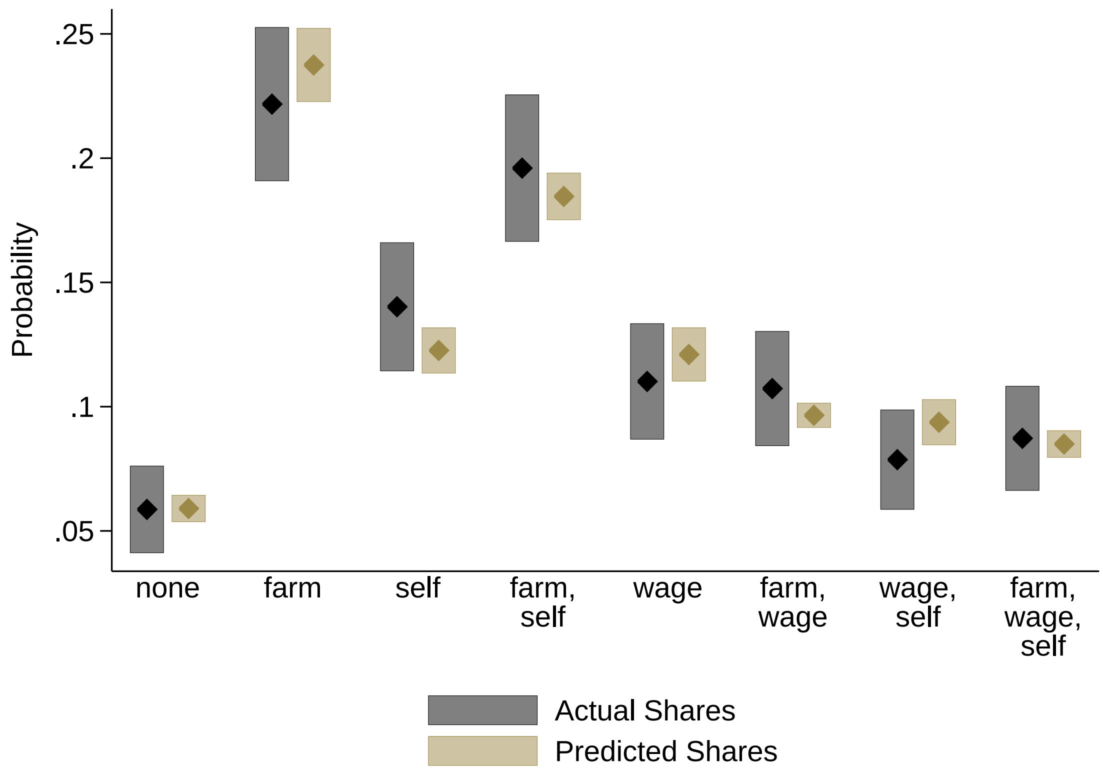


Figure 6: Conditional probability of participation in each occupation over agroclimatic potential. The probability of each choice corresponding with a specific level of agroclimatic potential is the vertical distance between the line above and the line below the area labeled with that choice.

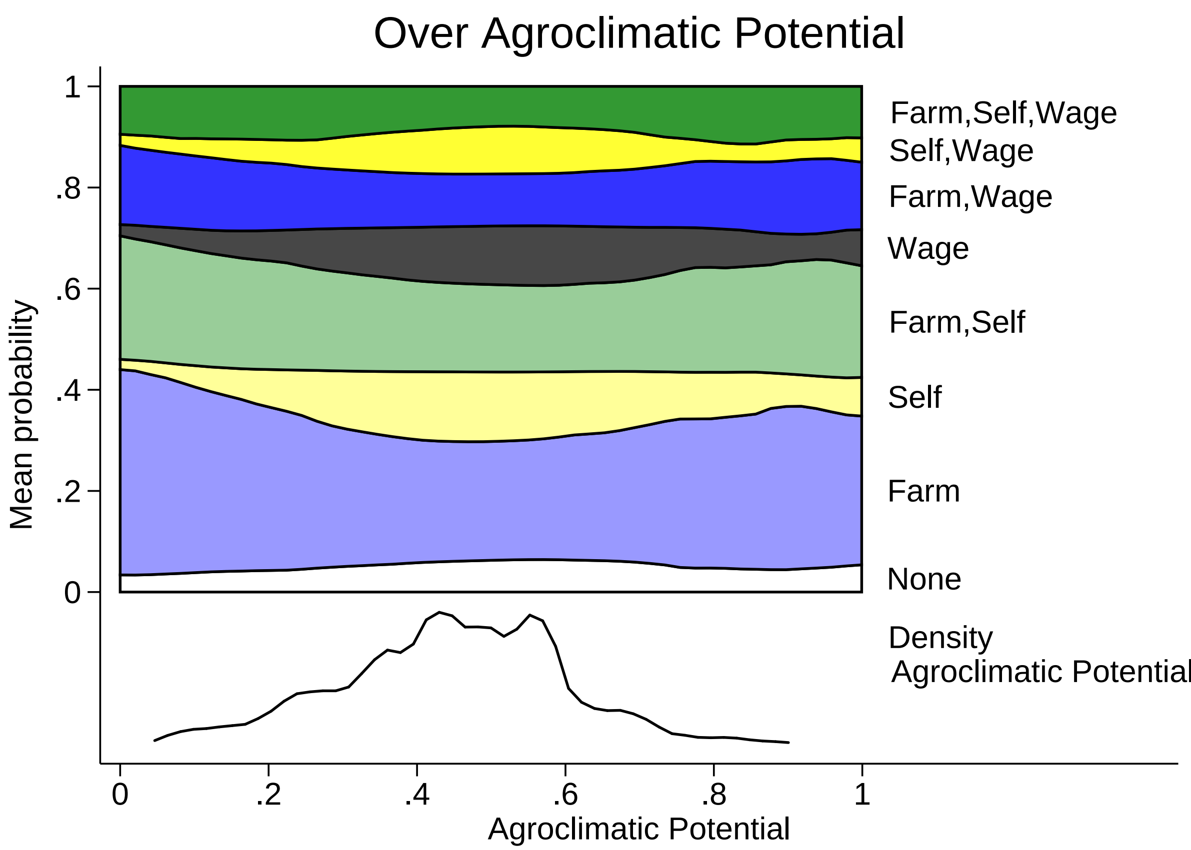


Figure 7: Conditional probability of participation in each occupation over population density. The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.

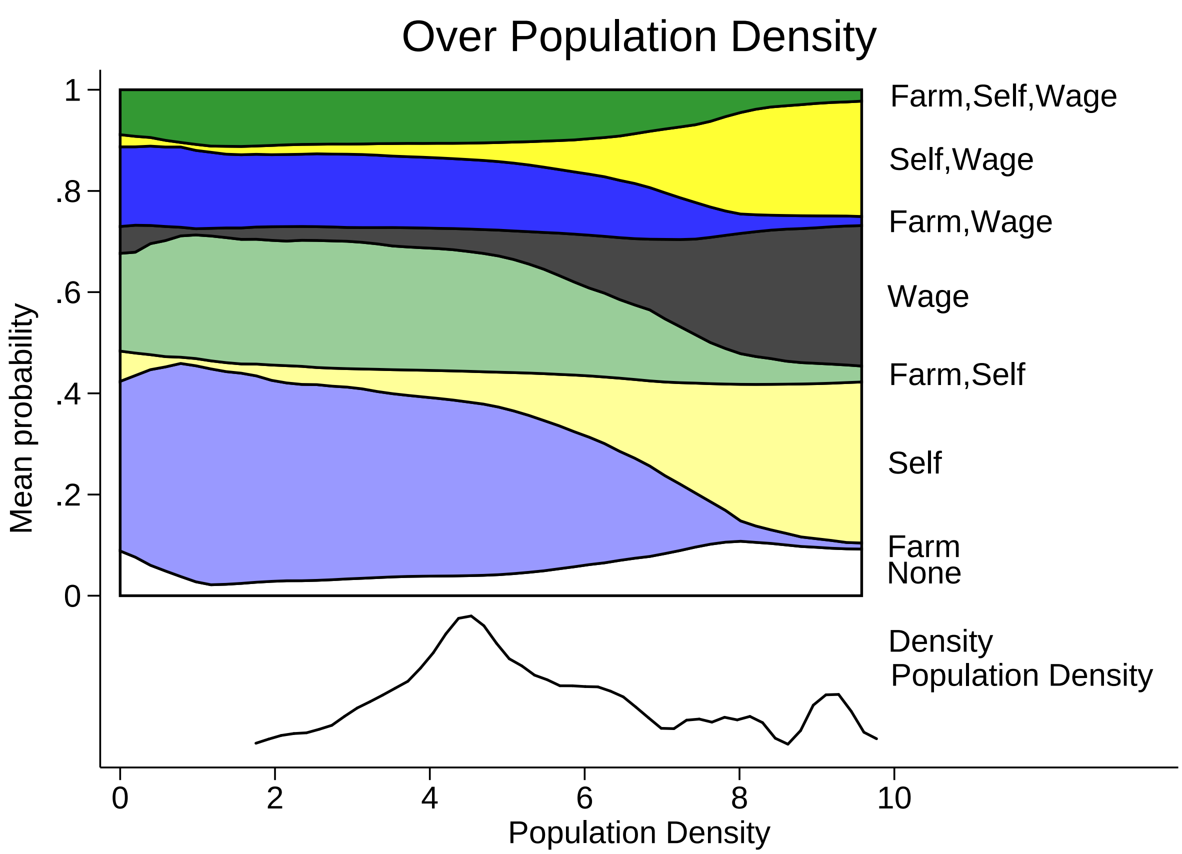


Figure 8: Conditional probability of participation in each occupation over remoteness (log of distance in km to the nearest population center of >500k). The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.

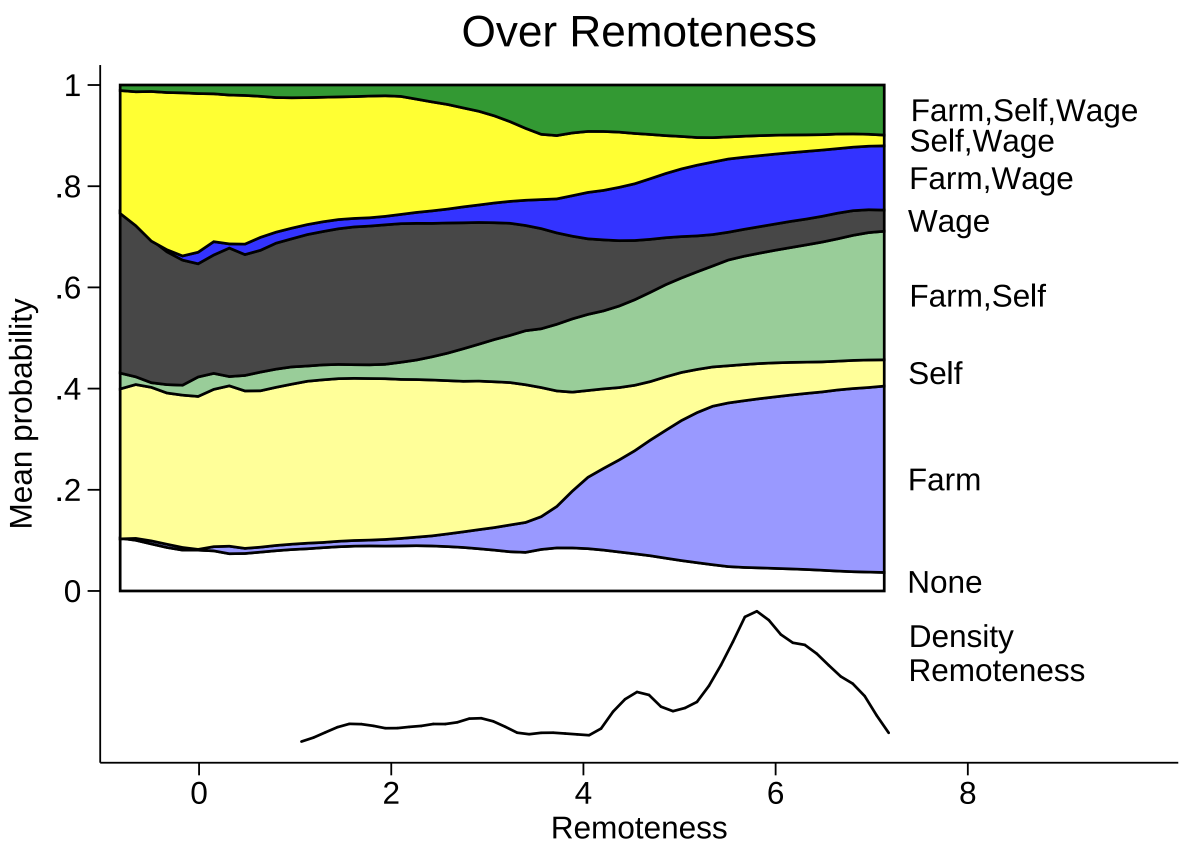


Figure 9: Difference in probability of participation in farming, self employment and wage employment across labor productivity simulations. In the farm simulation (depicted in green), farm labor productivity was doubled. In the self employment simulation (depicted in blue), self employment labor productivity was doubled. In the wage employment simulation (depicted in red), wage labor productivity was doubled. The diamonds show the mean difference in probability that households participate in each activity for each simulation, and the bars above and blow each diamond depict the 95% confidence intervals around the population mean.

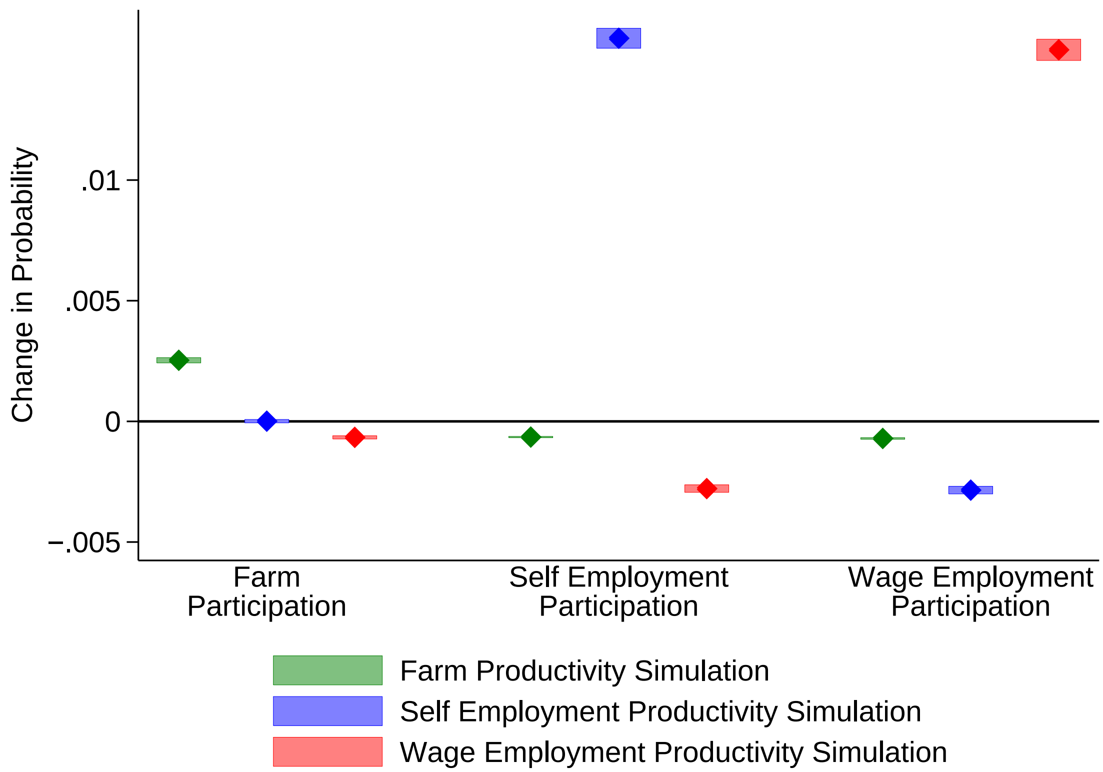


Figure 10: Difference in probability of participating in each occupational choice (combination of farming, self employment, and wage employment) across labor productivity simulations described in Figure 9.

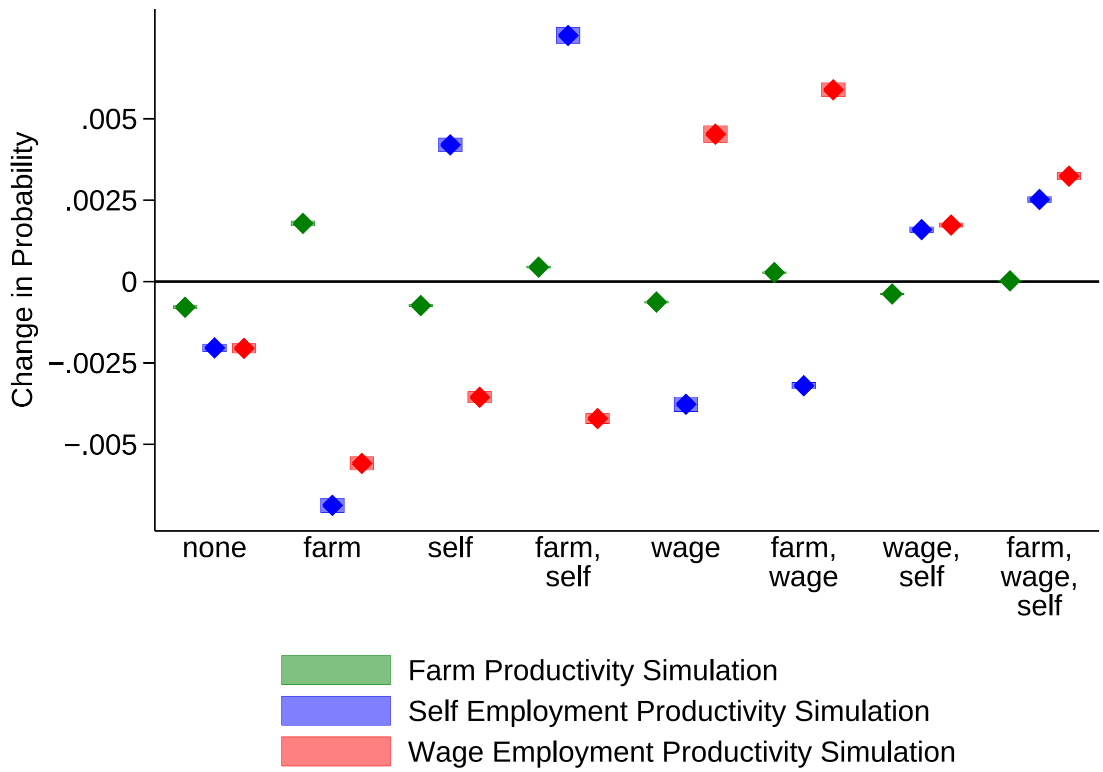


Figure 11: The probability that each household falls into the category of entering, always participating in, or exiting each each occupational choice (combination of farming, self employment, and wage employment) across the three labor productivity simulations described in Figure 9.

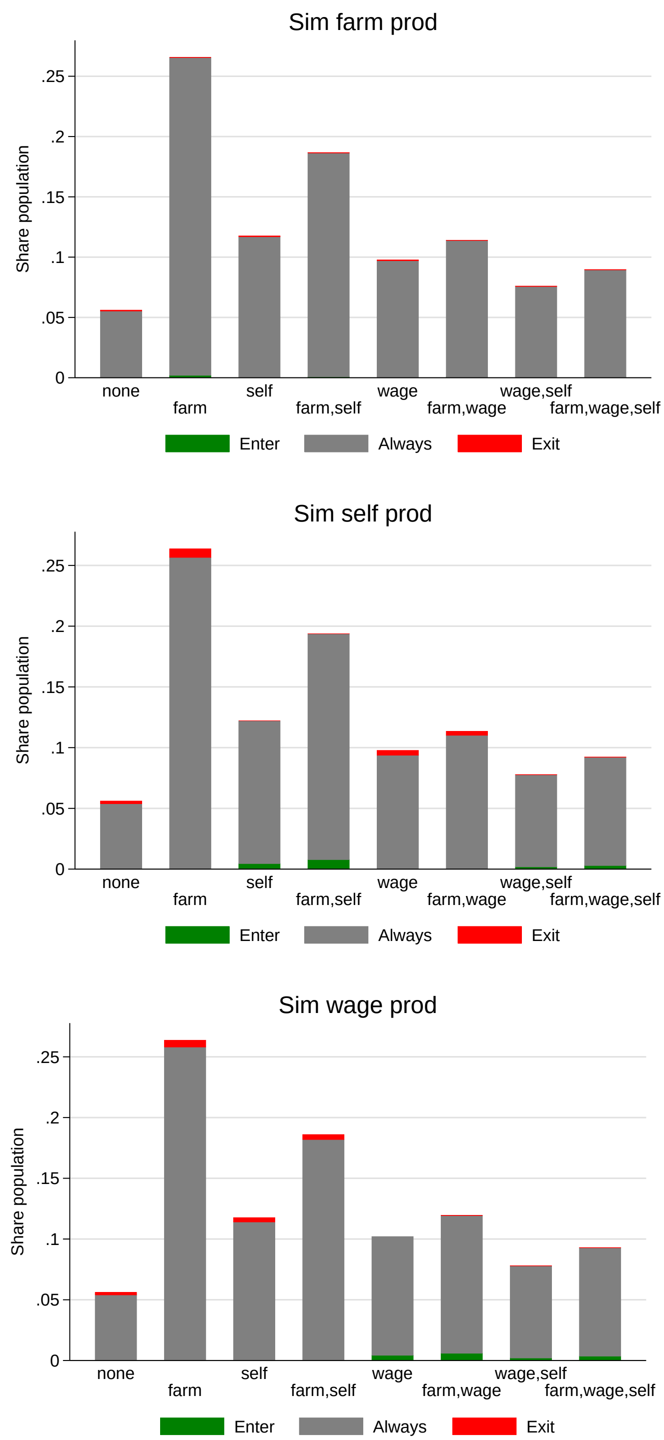


Figure 12: Welfare effects (% change in utility between simulated policy intervention and the baseline) of each policy simulation. For each of the three activities – farming, self employment and wage employment – the average welfare gains are decomposed into welfare gains from entry into the activity (when non-participating households shift to participation), exit from the activity (when participating households cease participation), and gains that occur within the participation margin (households participate before and after the intervention).

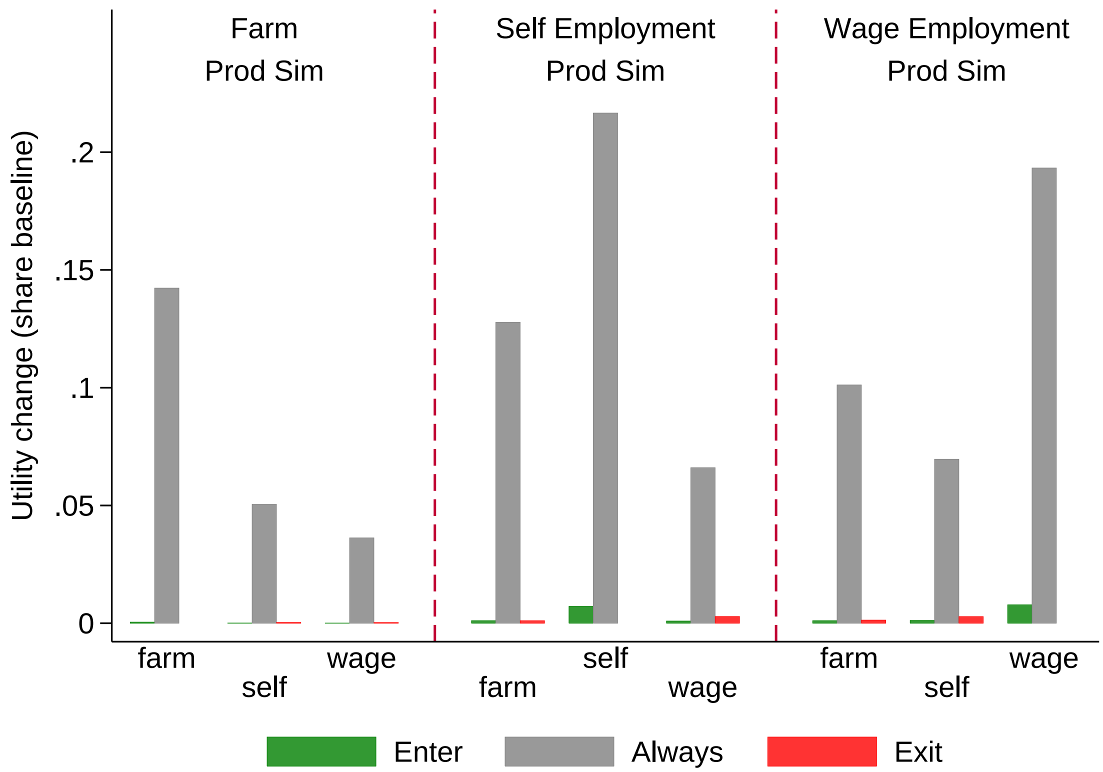


Figure 13: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on remoteness. The density of remoteness is shown underneath each regression.

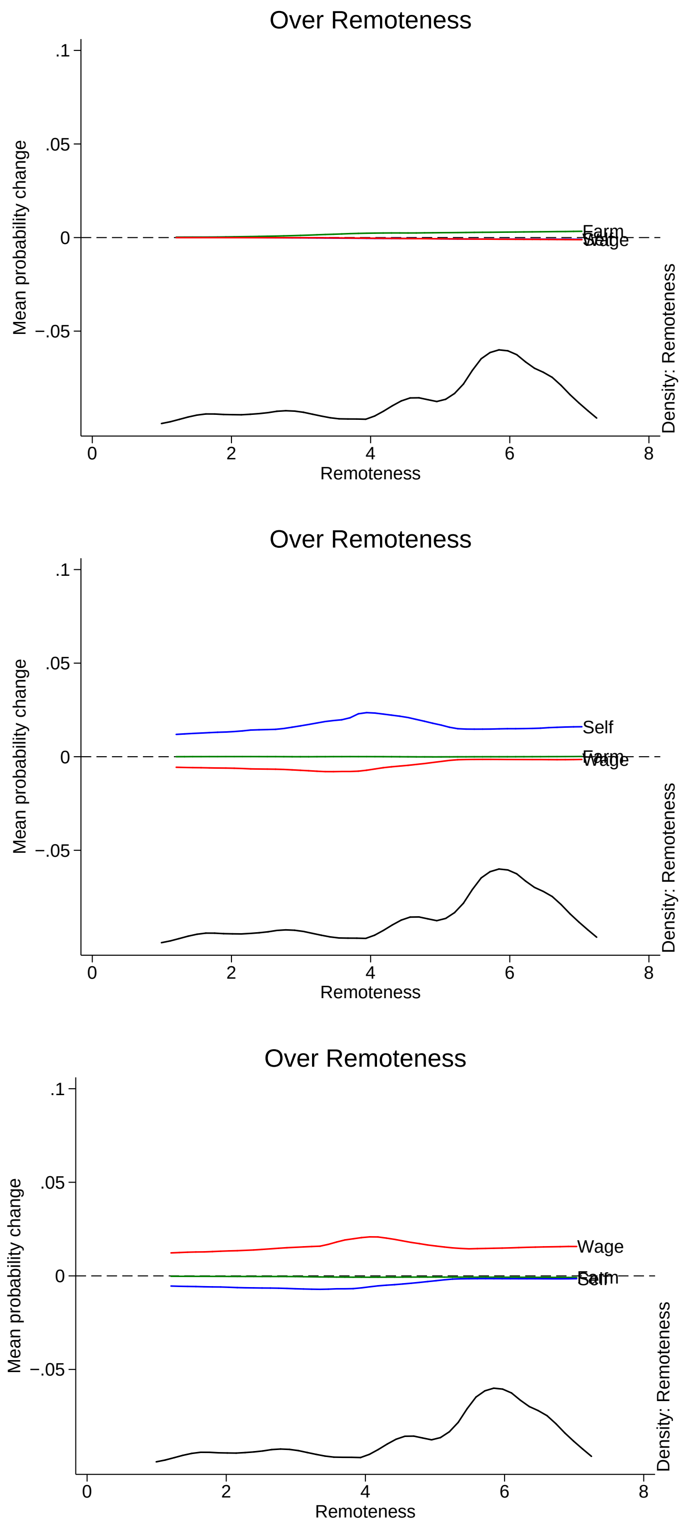


Figure 14: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over remoteness.

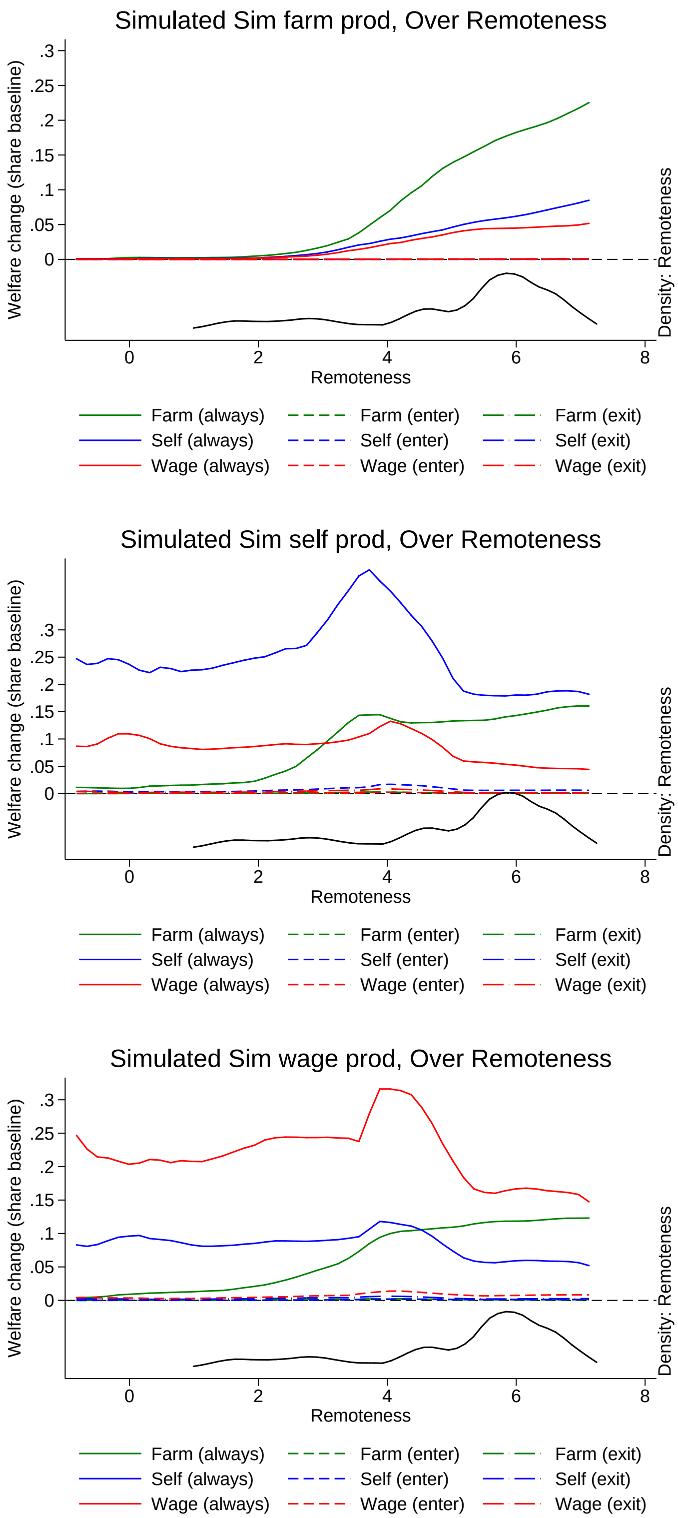


Figure 15: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on population density. The density of population density is shown underneath each regression.

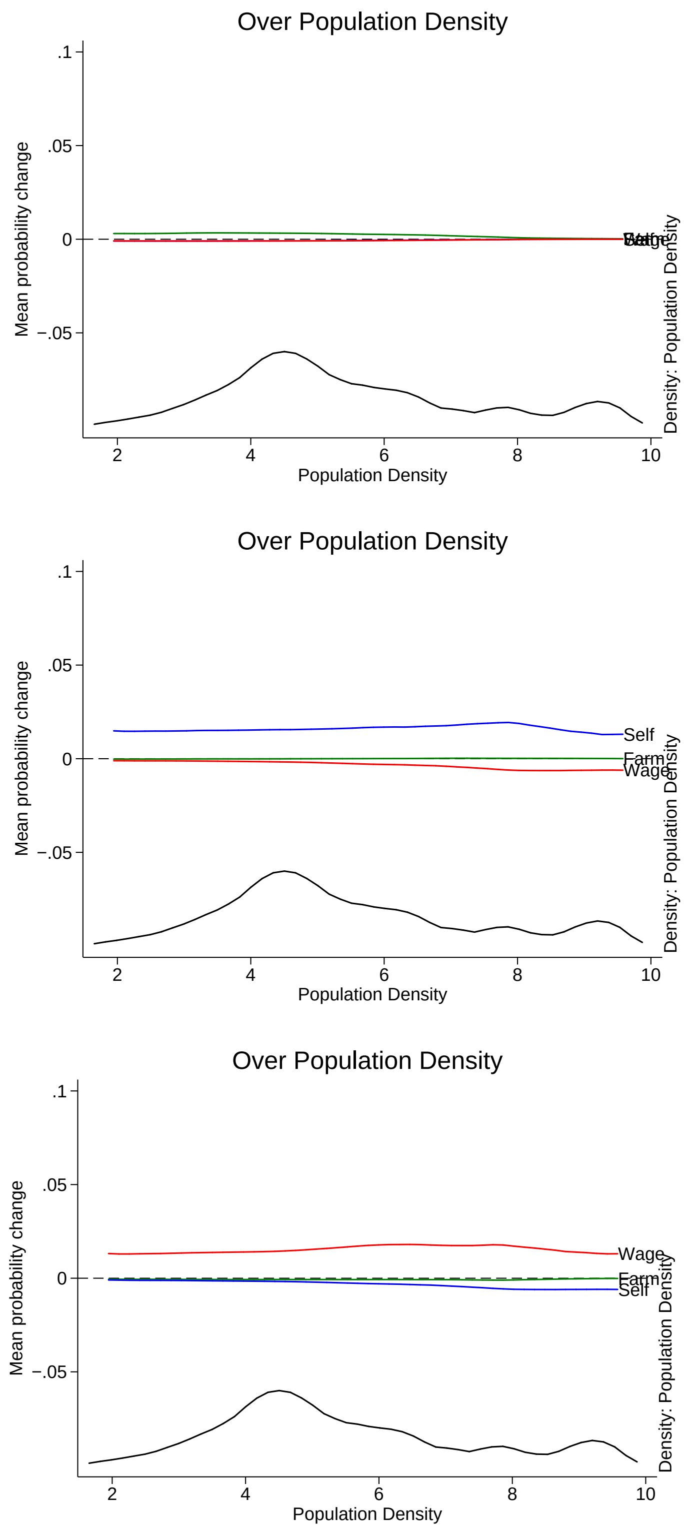


Figure 16: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over population density.

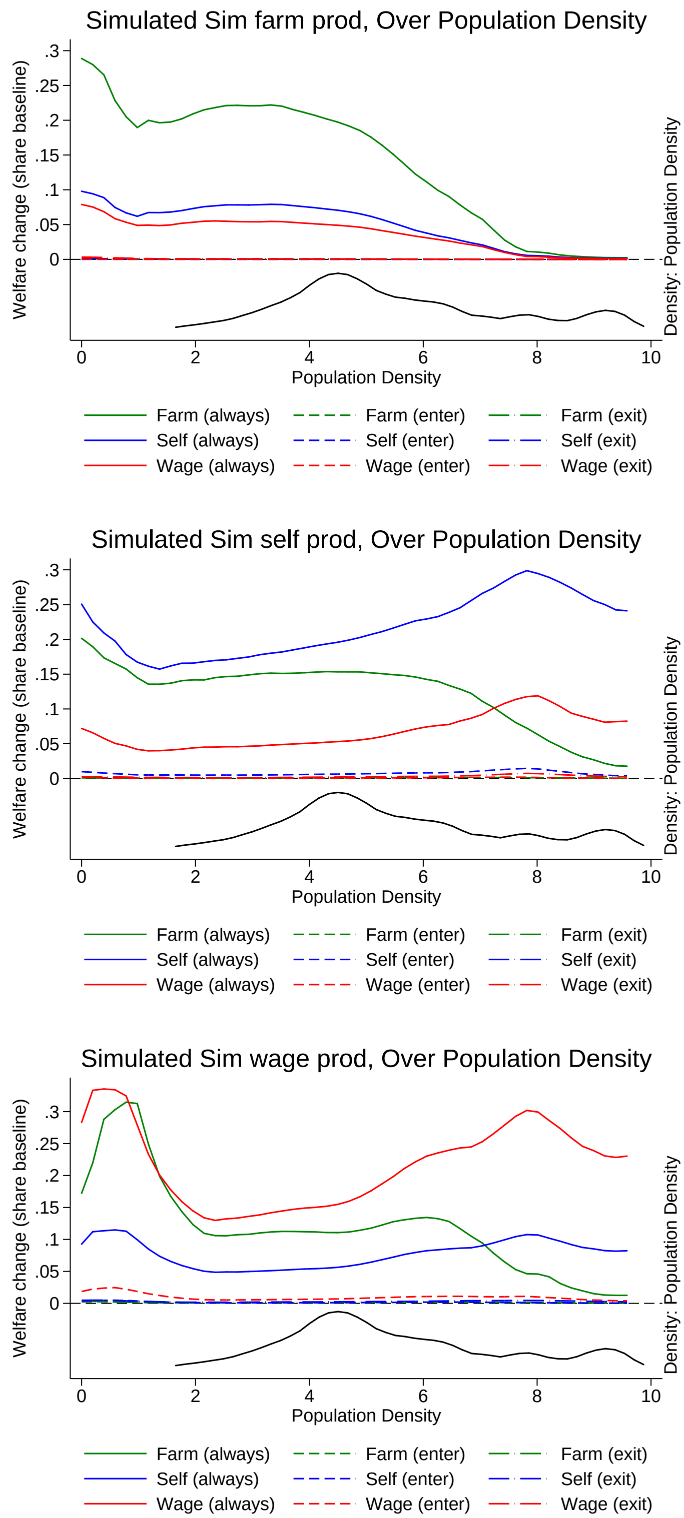


Figure 17: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on agroclimatic potential. The density of agroclimatic potential is shown underneath each regression.

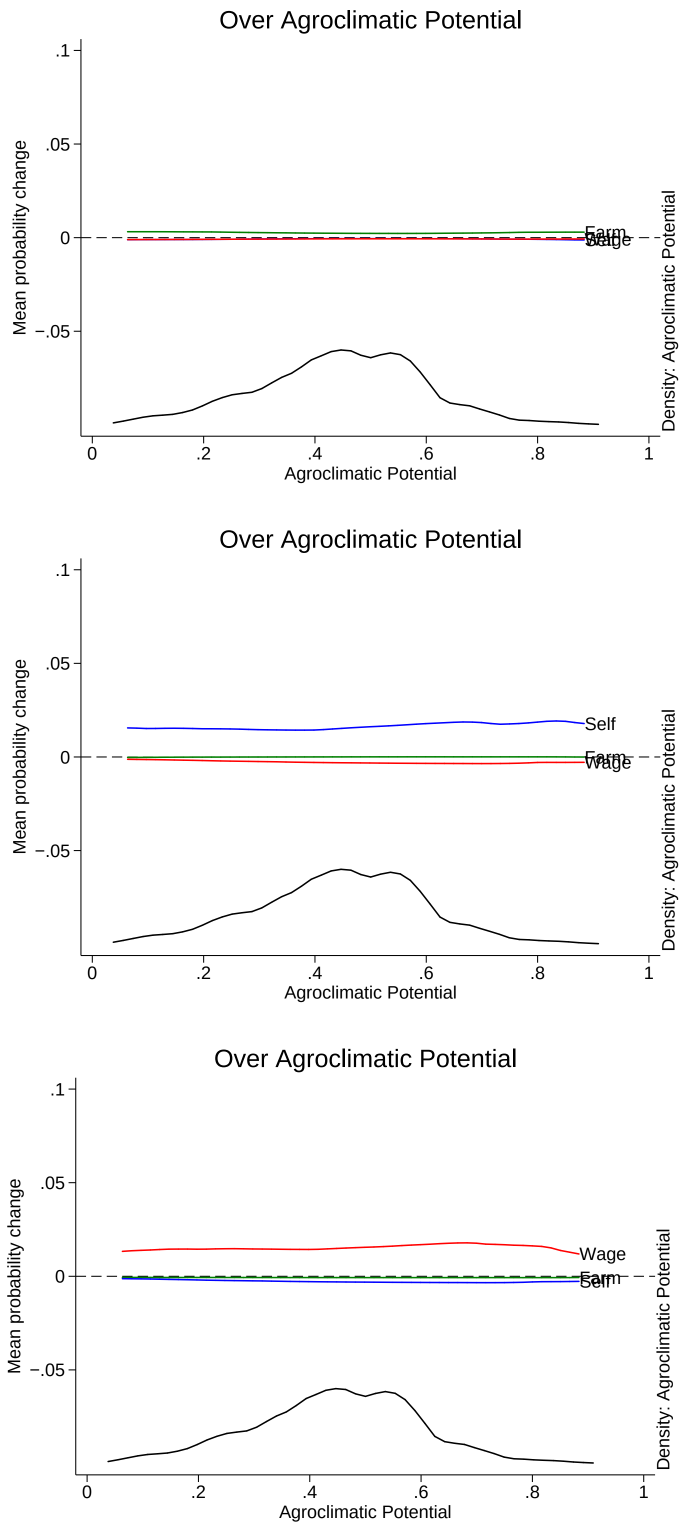
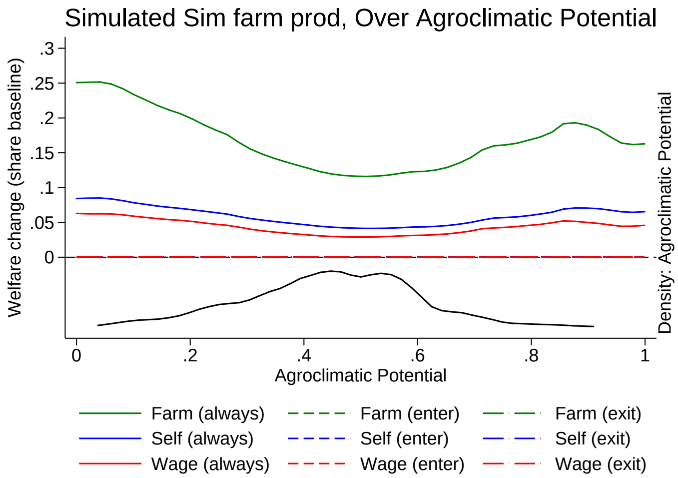
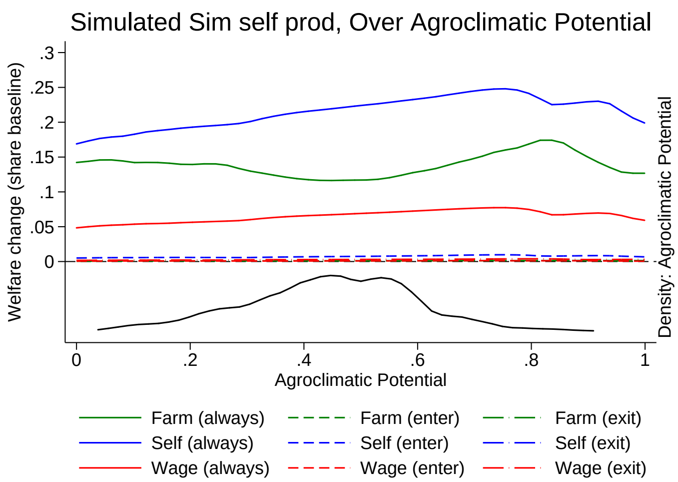
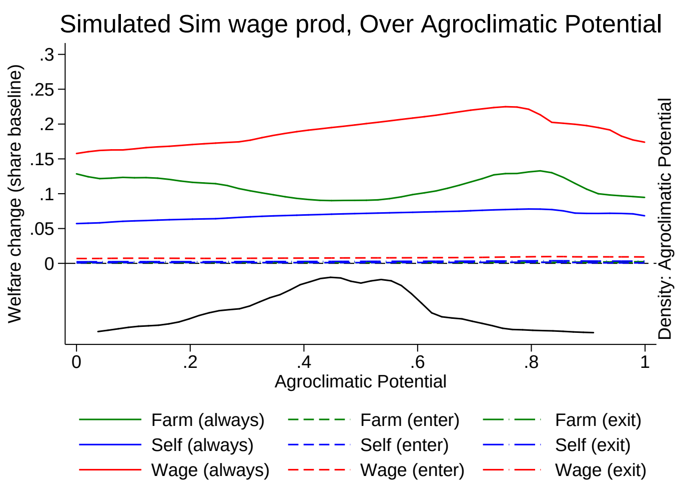


Figure 18: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over agroclimatic potential.







A Appendix

Table A.1: Summary statistics of regressors

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Transfers recvd (USD)	3,599	76.62	150.8	0	1,315
Household size	3,599	5.313	3.062	1	35
HH dependent share	3,599	0.324	0.255	0	1
Head's father attended school	3,599	0.469	0.499	0	1
Years educ, head	3,599	6.518	4.530	0	20
Urban	3,599	0.320	0.467	0	1
Peri-urban dummy	3,599	0.0547	0.227	0	1
Ave length of season (days, MOD12Q2)	3,599	178.2	23.11	131	234
Hrs travel to nrst town >500k (LSMS-ISA)	3,599	6.315	4.812	0	20.35
Network dist to nrst town >100k (km, LSMS-ISA)	3,599	106.9	99.62	0.240	546.4
Dist nrst major rd (km, LSMS-ISA)	3,599	17.05	23.32	0	135.4
People per square km, 2005 (ln, HC)	3,599	5.430	1.914	0	9.578
Number hh members 16-65	3,599	2.835	1.735	0	24
Age of head	3,599	46.49	15.63	18	105
Female head	3,599	0.252	0.434	0	1
Yrs educ, adults (ave)	3,599	6.461	3.969	0	20
Yield Potential low (cross-crop ind)	3,599	0.444	0.158	0	0.999
Yield Potential high (cross-crop ind)	3,599	0.529	0.179	0	1
Rate improved maize seed use (mean smth)	3,599	0.116	0.176	0	1
Cost hired lbr (med smth, USD/day)	3,599	2.786	2.379	0.194	10.12
Rate of tractor use (mean smth)	3,599	0.0286	0.0856	0	0.800
Land owned (ha, RIGA)	3,599	1.192	1.890	0	19.83
Mean precip wettest qrtr (mm, NOAA CPC)	3,599	579.9	172.3	231	1,440
Slope (pct, USGS)	3,599	4.886	4.848	0	46.60
Soil nutrient retention capacity (FAO)	3,599	1.496	0.956	0	7
Soil workability (FAO)	3,599	1.513	1.128	0	7
Share land irrigated (percent, FAO)	3,599	0.00566	0.0275	0	0.288
Cost/hired worker (med smth, USD)	3,599	791.3	851.6	2.049	4,973
Nighttime light ave coverage(DMSP F16)	3,599	683.4	1,372	0	4,764
Financial service available	3,599	0.450	0.498	0	1
Productive non-ag asset ind, fact 1	3,599	0.876	0.734	0	4.281
Productive ag-related asset ind, fact 1	3,599	0.0322	0.0433	0	0.341
Max educ in hh (yrs)	3,599	7.993	4.298	0	20
Nighttime light intensity(DMSP F16)	3,599	8.313	13.61	0	47.64
Returns/ag worker (med smth, USD)	3,599	449.9	487.6	14.94	2,135
Returns/ind worker (med smth, USD)	3,599	1,525	1,237	149.4	5,124
Returns/ser worker (med smth, USD)	3,599	1,974	2,251	25.62	17,079
Partpn in ag emplmt (med smth sh)	3,599	0.109	0.147	0	1
Partpn in ind emplmt (med smth sh)	3,599	0.0632	0.0934	0	0.667
Partpn in ser emplmt (med smth sh)	3,599	0.233	0.202	0	0.833

Table A.2: Comparison of farm profit coefficients with and without selection, Generalized Leontief specification. The first model includes only farm participants (no selection). The second and third columns present results from the second and first stages of a Heckman selection model, respectively. The marginal effects of profit function variables are shown.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	-171.761	101.024	135.842	105.371		
Peri-urban dummy (sq rt)	-324.229	786.995	-11.949	678.455		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-36.261	48.643	-87.897	53.434		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	10.518	7.465	14.805	8.933		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	2.188	13.251	15.000	18.173		
People per square km, 2005 (ln, HC) (sq rt)	-163.464	107.090	-204.729	122.090		
Number hh members 16-65 (sq rt)	438.735**	52.232	360.126**	64.025		
Age of head (sq rt)	20.156	19.963	43.922	25.068		
Female head (sq rt)	-88.432	47.877	-54.241	56.824		
Yrs educ, adults (ave) (sq rt)	37.378	27.615	66.084	35.784		
Yield Potential low (cross-crop ind) (sq rt)	293.084	367.483	-185.002	460.278		
Yield Potential high (cross-crop ind) (sq rt)	-689.259	371.876	-94.349	436.320		
Rate improved maize seed use (mean smth) (sq rt)	-50.670	148.237	-184.248	136.459		
Cost hired lbr (med smth, USD/day) (sq rt)	-167.123*	66.562	-121.834	67.761		
Rate of tractor use (mean smth) (sq rt)	747.074*	290.517	407.344	323.785		
Land owned (ha, RIGA) (sq rt)	440.026**	33.071	383.472**	40.049		
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	14.367	10.947	30.411*	12.090		
Slope (pct, USGS) (sq rt)	23.909	37.129	59.790	43.512		
Soil nutrient retention capacity (FAO) (sq rt)	-148.596	105.878	-217.076	119.507		
Soil workability (FAO) (sq rt)	2.062	98.096	216.567	130.760		
Share land irrigated (percent, FAO) (sq rt)	1,987.293*	860.207	2,556.904**	899.619		
Transfers recvd (USD)					-0.001**	0.000
Household size					0.099**	0.010
HH dependent share					0.623**	0.106
Head's father attended school					-0.035	0.055
Years educ, head					-0.063**	0.006
Urban					-1.391**	0.055
Peri-urban dummy					-0.619**	0.103
Ave length of season (days, MOD12Q2)					-0.011**	0.001
N	2,407		3,599		3,599	
R2 (adj)	0.343		0.327			
Lambda			-390.54			
Sigma			800.66			
P value comparison test			0.00			
N (non-censored)			2407			

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

Table A.3: Comparison of self employment profit coefficients with and without selection, Generalized Leontief specification. First stage Generalized Leontief estimation of enterprise profits. The first model includes only enterprise participants (no selection). The second and third columns present results from the second and first stages of a Heckman selection model, respectively. The marginal effects of profit function variables are shown.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	378.283	1,021.206	-710.422	952.864		
Peri-urban dummy (sq rt)	-3,672.948	4,960.063	-5,809.333	4,826.420		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-1,263.600	662.847	-987.280	531.530		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	2.817	105.276	-28.612	91.065		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	136.590	269.690	77.699	234.559		
People per square km, 2005 (ln, HC) (sq rt)	1,162.827	1,666.534	1,013.498	1,526.430		
Number hh members 16-65 (sq rt)	1,474.964*	712.782	804.815	750.843		
Age of head (sq rt)	-764.553*	317.143	-419.253	273.632		
Female head (sq rt)	-156.952	584.838	-201.497	538.188		
Yrs educ, adults (ave) (sq rt)	93.217	444.364	-42.117	379.546		
Cost/hired worker (med smth, USD) (sq rt)	26.940	30.811	24.064	28.378		
Nighttime light ave coverage(DMSP F16) (sq rt)	-33.909	48.258	-30.658	44.336		
Financial service available (sq rt)	424.410	517.335	1,003.185	513.653		
Productive non-ag asset ind, fact 1 (sq rt)	2,821.179**	942.155	3,171.157**	1,041.906		
Productive ag-related asset ind, fact 1 (sq rt)	-6,728.617	7,857.741	-4,581.303	7,078.165		
Transfers recvd (USD)					-0.000*	0.000
Household size					0.085**	0.009
HH dependent share					-0.124	0.126
Head's father attended school					0.153*	0.062
Years educ, head					0.009	0.007
Urban					0.186**	0.063
Peri-urban dummy					0.323**	0.114
Ave length of season (days, MOD12Q2)					-0.002*	0.001
N	447		3,599		3,599	
R2 (adj)	0.258		0.256			
Lambda			-977.66			
Sigma			3568.35			
P value comparison test			0.00			
N (non-censored)			447			

* $p < 0.05$; ** $p < 0.01$

Table A.4: Comparison of wage employment profit coefficients with and without selection, Generalized Leontief specification. The first model includes only wage market participants (no selection). The second and third columns present results from the second and first stages of a Heckman selection model, respectively. The marginal effects of profit function variables are shown.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	-240.086	843.255	-323.012	746.940		
Peri-urban dummy (sq rt)	2,187.252	3,290.466	1,515.902	3,555.486		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-1.983	383.008	92.147	328.373		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	-45.371	72.980	-64.368	59.958		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	-354.843*	158.223	-237.193*	114.297		
People per square km, 2005 (ln, HC) (sq rt)	1,355.943	1,129.243	348.260	897.252		
Number hh members 16-65 (sq rt)	1,045.595**	341.813	508.537	395.601		
Age of head (sq rt)	58.691	148.335	42.620	145.741		
Female head (sq rt)	-922.449**	308.528	-797.558*	341.084		
Yrs educ, adults (ave) (sq rt)	974.635	553.673	-154.296	624.761		
Max educ in hh (yrs) (sq rt)	1,883.221**	555.952	1,980.455**	586.246		
Nighttime light intensity(DMSP F16) (sq rt)	-46.184	185.853	49.682	163.018		
Returns/ag worker (med smth, USD) (sq rt)	5.589	60.635	-16.314	35.770		
Returns/ind worker (med smth, USD) (sq rt)	3.539	15.342	-2.169	17.224		
Returns/ser worker (med smth, USD) (sq rt)	26.232*	10.957	12.166	11.823		
Partpn in ag emplmt (med smth sh) (sq rt)	231.177	1,331.533	-313.959	1,021.808		
Partpn in ind emplmt (med smth sh) (sq rt)	203.108	709.598	-359.235	1,006.978		
Partpn in ser emplmt (med smth sh) (sq rt)	-491.600	961.593	254.652	803.682		
Transfers recvd (USD)					0.000	0.000
Household size					0.053**	0.008
HH dependent share					-0.797**	0.101
Head's father attended school					0.112*	0.048
Years educ, head					0.049**	0.006
Urban					0.558**	0.051
Peri-urban dummy					0.367**	0.098
Ave length of season (days, MOD12Q2)					0.004**	0.001
N	1,342		3,599		3,599	
R2 (adj)	0.371		0.362			
Lambda			-1697.31			
Sigma			3804.23			
P value comparison test			0.00			
N (non-censored)			1342			

* $p < 0.05$; ** $p < 0.01$

Table A.5: All coefficients, farm profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_farm_s																				
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	rypot3_low	rypot3_high	tech_use_maize	cpd_farmhire	tractor_share	landowned	rainfall_mean	slope	nutrientretention	soilworkability	irrigation_sh
Level_effect	705	1940	-667	275	406	-1484	1214	479	914	140	11977	-8801	-453	691	660	1234	290	-396	55	702	9870
sd	1217	3701	719	101	187	1185	693	335	709	303	4643	5174	1825	954	2424	524	135	445	1319	1383	7966
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	-9.01	370	230	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	126	445	47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	-31.9	-73.9	-20.3	-1.44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	20.2	51.9	11	1.39	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
dist_road	59.9	-94.2	-30.1	-1.97	-7.57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	32.2	85.7	18.7	3.2	3.99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
popdensity_ln	-9.75	-113	239	-22.6	-38	-69.6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	322	491	139	23.6	36.5	118	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
pop_indep	-101	257	125	-2.66	-33.9	-163	-31.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	121	196	64.4	10.2	15.8	138	45.9	.	.	.	.	.	.	.	.	.	.	.	.	.	.
age_head	-11.7	-27.3	-9.05	2.14	-.988	48.1	-60.7	-14.3	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	59	94.4	30.4	4.65	7.62	65.3	43.1	14.9	.	.	.	.	.	.	.	.	.	.	.	.	.
female_head	-67.6	-101	40.2	-29.3	8.47	126	-96.6	-35.1	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	133	232	75.3	11.8	19.1	161	85.1	43.1	.	.	.	.	.	.	.	.	.	.	.	.	.
educ_yrs_ave	15.5	65.8	42.8	-11.7	2.47	-34.6	4.18	.628	-16	14.3	.	.	.	.	.	.	.	.	.	.	.
sd	72.2	97.1	35	5.44	8.22	67.5	40	17.7	41.9	19	.	.	.	.	.	.	.	.	.	.	.
rypot3_low	1232	1197	468	-166	-79.3	-3762	-460	-21.3	39.3	-38.6	1225	.	.	.	.	.	.	.	.	.	.
sd	1056	1683	550	92.7	158	1198	427	222	530	235	2073	.	.	.	.	.	.	.	.	.	.
rypot3_high	-8.5	-933	-407	109	88.8	3904	314	-249	-301	-187	-1172	122	.	.	.	.	.	.	.	.	.
sd	994	1872	557	90.5	158	1155	453	227	535	245	3568	2493	.	.	.	.	.	.	.	.	.
tech_use_maize	-349	-384	401	19.8	-106	-101	-143	117	-188	96.8	-1526	1486	397	.	.	.	.	.	.	.	.
sd	283	510	165	26.2	50.3	407	170	79.9	197	96.1	1312	1313	494	.	.	.	.	.	.	.	.
cpd_farmhire	-212	29.6	-53	14.1	-42.6	163	61.4	61.5	165	-18.4	-253	92.3	-132	-76.3	.	.	.	.	.	.	.
sd	158	397	87	17.1	25.1	212	89.2	41.2	103	47.7	728	695	214	96.5	.	.	.	.	.	.	.
tractor_share	-153	-1739	-130	71.2	-112	-1018	263	188	-476	234	2217	-172	-708	-406	-1263	.	.	.	.	.	.
sd	543	1167	303	39.3	67.4	659	214	109	265	116	2008	2067	592	456	777	.	.	.	.	.	.
landowned	-214	-237	-55.9	.624	18	-7.11	-85.2	11	127	42.3	230	-12.8	51.2	-221	-131	-26.9	.	.	.	.	.
sd	99.9	148	49.6	8.02	12.4	110	56.4	28.1	71.7	29.5	378	386	128	71.9	154	26	.	.	.	.	.
rainfall_mean	-10.9	-61.1	-30.1	-2.99	-1.82	-.47	-7.88	-1.55	-35.3	-1.15	-104	43.8	7.75	-18.6	-101	-17.7	-1.69	.	.	.	.
sd	24.4	89.7	17.3	2.02	4.14	27.5	13.1	6	14.2	6.81	99.9	103	37.9	18.9	62	11.5	2.07	.	.	.	.
slope	-78.1	-183	30.1	6.87	-12.6	274	-23.9	-18.8	12.6	43.9	271	-71.2	33.7	-25	176	17.2	-.379	-19.1	.	.	.
sd	95.5	234	49.2	7.44	12.4	109	50.2	23.6	57.1	27.1	387	382	135	67.6	201	40	10	22.1	.	.	.
nutrientretention	-204	-572	17.3	-8.49	-66.5	155	-58.5	7.85	342	131	-1261	510	-397	-608	1203	-99.9	-11.2	-254	1176	.	.
sd	235	505	115	20.3	35	237	124	58.5	141	61.9	997	1055	325	201	523	98.4	25	92.1	471	.	.
soilworkability	194	480	-299	-17.9	71.5	141	240	-102	-252	-149	655	-274	-643	212	545	-23.3	59.6	85.3	-968	-192	.
sd	220	570	117	18.8	35.7	247	124	56.6	140	63.1	1030	951	302	182	455	94.9	26.2	84.6	415	261	.
irrigation_sh	-165	10672	584	-213	153	393	84.4	-241	-1677	50.8	-4493	5915	2809	252	-5130	352	-295	-242	1841	-2970	-6172
sd	872	7714	816	146	246	1779	581	280	624	286	4898	4250	1515	825	3068	430	172	444	2118	1456	2788

Table A.6: All coefficients, self employment profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_ent_s														
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	cpw_enthire	lightsum	fs_available	assets_na	assets_ag
Level_effect	-6789	-1157	-3707	-221	-635	-5511	-362	2003	3292	19.4	143	92.4	-5089	6611	-11083
sd	4416	6879	2370	436	742	4896	3631	1779	3705	1528	168	151	2947	4565	25551
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	-75.5	-4161	302	.	.	.	.	.	.	.	.	.	.	.	.
sd	447	2028	201	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	19.3	275	-5.44	9.09	.	.	.	.	.	.	.	.	.	.	.
sd	78	281	50.8	7.36	.	.	.	.	.	.	.	.	.	.	.
dist_road	-88.7	192	-7.44	.886	-12.1	.	.	.	.	.	.	.	.	.	.
sd	151	406	84.9	18.1	22.4	.	.	.	.	.	.	.	.	.	.
popdensity_ln	863	2369	891	24.4	61.2	-281	.	.	.	.	.	.	.	.	.
sd	1638	2617	762	130	234	873	.	.	.	.	.	.	.	.	.
pop_indep	1495	1720	168	41.2	62.4	313	464	.	.	.	.	.	.	.	.
sd	659	1208	346	67	120	1207	279	.	.	.	.	.	.	.	.
age_head	252	-450	-93.1	-5.19	70.3	188	-61.5	-138	.	.	.	.	.	.	.
sd	293	540	153	28.8	56.5	525	248	88.7	.	.	.	.	.	.	.
female_head	1108	-977	-149	.295	-6.9	-10.4	547	-42.9	.	.	.	.	.	.	.
sd	655	1133	347	69	123	1291	563	250	.	.	.	.	.	.	.
educ_yrs_ave	-155	-482	-19.3	-13.3	52.9	271	-358	-49.1	-71.3	-37.7	.	.	.	.	.
sd	382	586	201	35.5	64.2	562	308	126	293	125	.	.	.	.	.
cpw_enthire	32.9	-18.2	21.2	-1.91	2.14	-40.6	-40.7	-1.52	-50.2	18.8	-.302	.	.	.	.
sd	25	94.5	14.9	2.95	5.71	67.4	22.4	9.99	21.7	12.1	.556	.	.	.	.
lightsum	-74.6	-96.5	-45.3	3.68	-9.64	23	-2.31	1.78	-22.8	4.54	.647	-.696	.	.	.
sd	39.4	54.7	22.6	4.07	8.35	61.2	23.5	10.7	24.2	14	1.16	1.1	.	.	.
fs_available	173	1176	700	-.546	-51	1679	117	-37.1	-1270	-50	-22.4	-1.6	.	.	.
sd	576	998	318	58.6	109	1066	500	219	501	273	20.7	23	.	.	.
assets_na	-785	-1191	46.9	40.8	-154	533	-685	-1273	147	843	16	-2.07	-275	1582	.
sd	923	1516	459	87.4	173	1604	761	342	813	443	29.2	34	673	642	.
assets_ag	5833	4683	2398	-357	270	4687	-4758	2725	-11479	-2183	-165	-5.15	6306	-1901	1735
sd	4847	12459	2331	458	742	7666	3137	1786	5015	2114	151	170	3285	4891	14763

Table A.7: All coefficients, wage employment profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_market_s																	
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	educ_yrs_max	lightintensity	rpw_ag	rpw_ind	rpw_ser	partshare_ag	partshare_ind	partshare_ser
Level_effect	3277	-7270	2579	-501	375	5175	-2568	149	-410	-4738	1286	-1792	299	-9.55	-19.1	5206	14013	-14406
sd	6055	10841	2900	567	888	6587	3876	2018	3601	5000	4718	1590	235	176	130	9071	9050	8464
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	466	1722	-399	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	947	2555	228	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	-89.4	-356	76.4	-2.45	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	140	381	56.8	8.37	.	.	.	.	.	.	.	.	.	.	.	.	.	.
dist_road	170	-387	113	11.3	5.08	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	224	543	97.7	20	27.6	.	.	.	.	.	.	.	.	.	.	.	.	.
popdensity_ln	-1655	3174	-1358	73.6	-159	129	.	.	.	.	.	.	.	.	.	.	.	.
sd	2149	3884	935	180	264	854	.	.	.	.	.	.	.	.	.	.	.	.
pop_indep	785	2158	170	-14	-171	226	193	.	.	.	.	.	.	.	.	.	.	.
sd	822	1298	362	75.2	138	1272	377	.	.	.	.	.	.	.	.	.	.	.
age_head	-242	437	272	-5.59	-4.53	142	-5.31	-121	.	.	.	.	.	.	.	.	.	.
sd	361	572	160	31.2	63.6	585	278	97.5	.	.	.	.	.	.	.	.	.	.
female_head	162	542	247	68.6	-45.4	304	-43.8	-170	.	.	.	.	.	.	.	.	.	.
sd	768	1358	348	68.8	133	1183	584	275	.	.	.	.	.	.	.	.	.	.
educ_yrs_ave	1172	3749	158	-79.3	-40	-241	293	-235	753	52	.	.	.	.	.	.	.	.
sd	1082	1532	459	92.5	173	1652	854	400	871	889	.	.	.	.	.	.	.	.
educ_yrs_max	-352	-1899	-83.5	50.9	177	-109	201	266	-654	2450	-1436	.	.	.	.	.	.	.
sd	1065	1497	459	86.9	167	1561	816	366	832	1830	1036	.	.	.	.	.	.	.
lightintensity	142	-1112	104	33.3	12.9	774	-117	-65.2	360	-498	496	-46.7	.	.	.	.	.	.
sd	367	786	199	34.2	64	590	260	119	258	356	343	97.3	.	.	.	.	.	.
rpw_ag	71.3	-54.7	-36.9	-.808	2.9	-75	-42.2	-5.23	34.5	-11.1	19.4	-14.9	1.16	.	.	.	.	.
sd	66.4	126	51	5.64	9.68	75.2	31	14.9	33.3	39	38.5	15.2	1.73	.	.	.	.	.
rpw_ind	-49.6	-50.6	-9.99	4.78	-18.7	-30.2	32.7	-1.38	-4.66	-25.2	3.83	-6.76	-2.18	3.47	.	.	.	.
sd	39.6	54.9	16.3	4.34	6.97	61.7	24.1	10.6	23.7	30.1	30.3	12.4	1.59	.944	.	.	.	.
rpw_ser	-29.9	-18.5	8.05	.786	-3.34	-46	44.1	7	-14.4	30.6	-30.8	11.3	2.58	-1.5	.345	.	.	.
sd	29.5	63	14.7	2.8	4.78	45.4	14.6	6.4	14.8	20	20.8	8.63	1.2	.969	.239	.	.	.
partshare_ag	-2050	-9372	-1316	56.3	173	-1367	-1080	301	533	-542	870	615	-8.5	5.51	-48.6	-3551	.	.
sd	2091	3420	1088	173	300	3041	1420	635	1466	1817	1775	688	77.7	72.5	44.1	3054	.	.
partshare_ind	2624	-669	-861	-95.7	46.5	-5387	1059	-600	-1628	-1375	334	71.5	-52	78	31.3	-1258	4429	.
sd	1966	3140	834	159	342	3308	1380	630	1272	1757	1707	658	86.4	60.7	40.2	3227	3855	.
partshare_ser	-1890	1140	-198	116	-139	3568	-2910	1518	-3149	1150	-942	190	-32.4	-30.5	24.1	5851	1167	737
sd	2160	2987	825	165	310	3051	1427	658	1450	1749	1667	674	87.8	59.8	45.8	3630	3186	2298