Occupational Choice and Agricultural Labor Exits in Sub-Saharan Africa

Ellen B. McCullough
Department of Agricultural & Applied Economics
University of Georgia

December 15, 2018

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 occupational choice that underlies the structural transformation process. I model households' decisions to participate in different economic activities – farming, wage employment, and self employment. I estimate a structural model of occupational choice using nationally representative household survey datasets from Tanzania, matched geospatially to additional datasets. I allow household preferences to be correlated across choices, relaxing strict independence assumptions. Then, I simulate the response of occupational choice to stylized productivity shocks. I find that farming productivity shocks mostly benefit the many households that already farm, while inducing a few to enter farming and slightly slowing the rate at which farming households participate in nonfarm self and wage employment. Wage and self employment participation do respond to wage and self employment productivity shocks, respectively, where market access is good. The results suggest that households are likely to diversify into nonfarm activities before exiting agriculture, and they highlight the strong welfare effects of smallholder farmer productivity gains, especially in places with low population density and poor market access.

Acknowledgments

I am grateful to Chris Barrett, Victoria Prowse, and David Just for their research guidance. Funding support came from the African Development Bank through the Structural Transformation of African Agriculture and Rural Spaces (STAARS) project. This research was made possible through the use of Cornell University's Economics Compute Cluster Organization, which was partially funded through NSF Grant #0922005. Siobhan Murray, of the World Bank, 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 Barrett's research group, Karen Brooks, Luc Christiaensen, Hans Binswanger, Paul Christian, Erik Thorbecke, Nancy Chau, and Prabhu Pingali for helpful comments. Any errors are my responsibility solely.

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). 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). Because farming is the most common occupation for poor workers, agriculture is central to the structural change process (Christiaensen, Demery, and Kuhl 2011).

African countries are mostly in the early stages of structural transformation, with large labor shares still in agriculture and large productivity differentials between agriculture and other sectors (Gollin, Lagakos, and Waugh 2014b). Recent evidence suggests that structural change could be picking up in sub-Saharan Africa, with rising annual output per worker which is at least partly explained by labor exits from agriculture to other sectors (McMillan and Harttgen 2014). At the same time, African policy-makers are concerned with how to employ a bulging demographic of youth who are seeking economic returns commensurate with their skills (Filmer and Fox 2014).

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).

Despite widespread regularity in the associations between agricultural productivity, agricultural labor shares, and poverty reduction, many question the scope for achieving structural change in sub-Saharan Africa through smallholder-focused interventions. 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 smallholder-focused agricultural interventions for poverty reduction in the region and look to alternate interventions that raise labor productivity more generally and in other sectors of the economy.

The debate about agriculture's role in overall economic growth in sub-Saharan Africa hinges on the potential for raising labor productivity in agriculture and other sectors, and on the implications of rising agricultural labor productivity for occupational choice. This paper focuses on the second issue – the implications for occupational choice of achieving agricultural productivity growth targets. The occupational choice decisions that underlie the structural transformation process play out among many households and farms in heterogeneous settings, with variation in skill levels, agroclimatic potential, and geographic determinants of prices and wages. 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.

While there is strong 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). The large literature addressing agriculture's role in structural change is mostly macro-economic in scale (e.g., Christiaensen, Demery, and Kuhl 2011; McMillan, Rodrik, and Verduzco-Gallo 2014; Rodrik 2016; Gollin, Lagakos, and Waugh 2014a,b). Several literatures address occupational choice in developing countries, and several literatures address technology impacts in developing countries. However, few studies have addressed the impacts of agricultural technologies on occupational choice through a structural change framework, with the exception of Bustos, Caprettini, and Ponticelli (2016) and Foster and Rosenzweig (2007). One major reason for research scarcity on this topic has been, until very recently, lack of datasets that cover all relevant sources of labor earnings, including wage market participation, self employment in farming, and self employment in household-managed non-farm enterprise.

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, male-headedness of the household, land ownership, irrigation, slope, and soil characteristics. Self employment latent labor productivity is closely related to male headedness and productive asset ownership. Wage employment latent labor productivity is closely related to urban status, market access, male-headedness, education within the household, and local wage rates.

I then 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. Households that do not shift occupational choice experience the majority of welfare effects from productivity simulation. 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 African workers are likely to increase supply to non-farm activities before they exit farming altogether.

2 Model

In sub-Saharan Africa, workers outside of agriculture tend to have higher returns per worker per year (McCullough 2017; Gollin, Lagakos, and Waugh 2014b; 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 2017).

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 occupations.

While self employment is not as commonly included in occupational choice models as is wage employment, I include it here because participation rates are so high 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: farming (F), wage employment (WE), and self employment (SE). Farming includes self emloyment through a farm enterprise as well as wage employment as a farm laborer. Wage employment includes all types of wage employment except for working as a farm laborer. Self employment includes all non-farming

[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.

self employment. Agricultural wage labor is incuded in the farm self employment category so that the other two categories - wage employment and self employment - can be considered as non-agricultural. 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 a household is 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:

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

s.t.

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

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

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

[2]Modeling individual occupational choices is 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.

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 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 disutility 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

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

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.

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 over-predict participation in that activity (van Soest, Das, and Gong 2002). Occupation-specific preference shifters (δ_j) can pick up these fixed costs that 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 and between the second and third survey rounds are shown in Figures 1 (farming), 2 (self employment) and 3 (wage employment). Overall, a plurality of households never participate in wage labor markets. Self employment outside of agriculture is the more common form of non-agricultural labor supply. While entry into and exit from all of the activities is common, the largest categories are households who do not change participation in a given activity across survey rounds. This is particularly true for farming, where a large majority of households farmed in all three 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 across three survey rounds, 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 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. I pool observations across three roundns of panel data and estimate the random coefficients 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 h refers to the household identifier. 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).

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

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

Following estimation, I predict choice probabilities for each household and choice by simulating R draws, drawing values of δ_{hj} 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. Based on 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 three, pooled rounds of data. Round 1 is from 2008-09; round 2 is from 2010-11; and round 3 is from 2012-13.

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 Tables 1 (round 1), 2 (round 2), and 3 (round 3).

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 power parity conversion factor for private consumption from the World Bank's 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

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

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 part of the Global Agro-Ecological Zones dataset. I use the crop area weights described above 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 A.1 contains summaries of all of the variables used to model labor productivity and occupational choice, by survey round.

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 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.

Wage labor profit variables include locally smoothed median annual returns to wage employment and participation rates for the industry and service sectors. These are meant to proxy for the demand for wage employment. Additionally, the maximum educational

[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/

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.

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.

5 Results

I estimate the model using three pooled rounds of Tanzania survey data. Table 4 depicts the marginal effects of each variable on annual household returns to participation in farm employment, self employment, and wage employment. The selection variable coefficients are shown in Table 5. The first stage model fit is fairly pretty good, with pseudo-R^2 values of 0.23 for farming, 0.19 for self employment, and 0.30 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: female headed (neative), household size (positive), land ownership (positive), slope (positive), irrigation access (positive), soil nutrients (positive), and soil workability (negative). For self employment, the marginal effects that are significantly different from zero include: female headed household (negative), non-agricultural productive assets index (positive), and agricultural productive assets index (negative). For wage employment, the marginal effects that are significant include: urban (negative), female-headed (negative), travel time to the hearest town of at least 500,000 (negative), age of household head (positive), average education level of the househoold (positive), and average service sector wage rates (positive).

The first stage labor productivity model results suggest that households experience higher returns per worker per year outside of agriculture then within agriculture (Figure 4). A comparison between households' actual labor productivity in their chosen occupation with their predicted productivity across non-chosen options suggests that households positively sort into activities that involve both wage and self emloyment. That is, their actual productivity is higher than it would have been for non-selected occupations. The oppositve is true for households that only farm or that do not supply labor. These households face higher predicted labor productivity outside of their chosen occupations, suggesting some barriers to suplying labor to wage and self employment.

The parameter estimates for the second stage occupational choice model are depicted in Table 6. The income parameter (α), the alternative-specific coefficients for the selection variables, and the alternative-specific random coefficient parameters. The parameter estimates are consistent with scale heterogeneity and correlation of preferences across choices. Receipt of external transfers tend to decrease labor supply overall, especially for activities involving farming. Participation in farming increases in household size, on the other hand. Households with a large share of dependents are more likely to farm and less likely to supply other types of labor. Education tends to increase the probability of participating in wage employment, as does the household head's father's education. Farming participatio decreases with the length of the growing season.

In Table 7, 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 each x variable. This expression is evaluated at each data point, drawing simulated δ coefficients at the houseold level using the estimated parameters of the multivariate normal distribution. Similarly in Table 8, 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 5 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. Figure 6 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 subsample groups for which fit was checked as well. Next, I performed a validation exercise by estimating the occupational choice model on a subset of data, randomly dropping one fifth of the sample enumeration areas. Figure 7 depicts a comparison between predicted probabilities and actual choice shares for the enumeration areas held out of the model estimation sample, 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 8). 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 9). 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 10). 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

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. The simuations are meant to parallel classes of interventions that would target different types of economic activities. Farm interventions would aim to increase the output per farm worker through improved farm technologies, such as seeds. Self employment interventions might include access to micro credit or entrepreneurial training. Wage-employment interventions would arise from industrialization-focused policies focused on generating employment in the industry and service sectors of the economy.

For each simulation, I generate a new income variable for each occupational choice. These are generated by altering the first stage prediction of labor productivity for each household, which are conditioned on the variables that are arguments in the profit function for farming, self employment, and wage employment. Using the newly simulated imputed income for each occupational choice, I predict new occupational 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, expressing the predicted utility change in terms of its income equivalent as a share of the household's baseline income.

For each of the three simulations, Figure 11 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, in the sense that none of the produtivity shocks induces entry into or exit from farming. Households, 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 2 percentage points when self employment income doubles, and it decreases by half of a percentage point when wage labor productivity doubles. Similarly, when wage labor productivity doubles, wage labor participation increases by about 1.5 percentage points, 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 12). 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 adding wage employment by by households that only farm and households that participate in farming and self employment. Households respond to the farm productivity shock, with a slightly higher probability of specializing in farming and a slightly lower probability of participating in any other activity. The farming-only occupation is the one with the highest mean δ, implying it is the one households are most likely to choose holding income constant.The majority of households do not respond to the simulations by changing occupation, implying that the occupational choice predictors and choice intercepts are strongly predictive of occupational choice (Figure 13).

When the probability of changing occupation is examined across different segments of the population, several clear patterns emerge. First, households at the higher end of the income distribution are more likely to respond to the self and wage employment productivity simulations by entering into the activity whose productivity is shocked than are households at the low end of the income distribution (Figure 14). Wealthy households are more likely also to exit from wage employment when entering into self employment, and vice versa, at the high end of the income distribution. The same patterns hold in areas with good market access and high population density – households in high population density areas are about twice as likely to enter into self or wage employment when the productivity is shocked than are housheolds in low population density areas (Figures 15 and 16). Households respond similarly to the farming productivity simulation across wealth levels, market access, and population density contexts.

It is also possible to examine the welfare effects of labor productivity shocks, both 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). Figure 17 decomposes the expected welfare effects of the policy simulations between households that switch occupations in response to the simulation and those that do not. Households that always farm, and households that exit self employment or wage employment in order to enter farming, experience the highest welfare gains from the farm productivity simulations. The self employment simulations benefit households that always participate in self employment the most, followed by households that exit wage employment and farming in order to participate in self employment. The wage employment simulations benefit wage labor participants the most, with a welfare gain that equates to 30% of baseline income. Households that exit farming and self employment in order to enter into wage employment also experience large welfare gains.

The farm simulation achieves the largest welfare effects on poor households. The expected welfare gain from improving farm productivity are large across the expenditure distribution, though they are larger for poor households than for wealthy households. Because most households, and especially most poor households, participate in farming, the largest expected utility gains from the farm productivity simulation accrue to households that always farm (Figure 18). Self employment and wage employment simulations have the largest expected welfare gains for households at the top of the welfare distribution, but self employment also generates welfare gains for poorer households who also participate in farming. The farm productivity simulations largely benefit households with poor market access, and generate very low welfare gains for households who live close to major population centers (Figure 19). The self employment and wage employment simulations largely benefit households who have good market access, and achieve lower average effects on households with poor market access.

7 Discussion

Understanding the sensitivity of occupational choice to labor productivity growth in different sectors is central to understanding how structural changes are likely to play out in an economy. It also has important implications for prioritizing 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 find that latent labor productivity is indeed predictive of occupational choice for a nationally representative sample of Tanzanian households surveyed three times over a five-year period. I am able to explain occupational choice patterns by estimating an occupational choice model with a preference structure that allows for correlation of preferences across occupations.

This analysis suggests that improving agricultural productivity could slow the rates at which households supply their labor to nonfarm wage and self employment, a finding that is consistent with many theoretical predictions (Foster and Rosenzweig 2007). It also suggests that farming participation, at least at the household level, is somewhat sticky. Households tend to diversify into noonfarm wage and self employment when these activities become more remunerative without ceasing their participation in farming. Diversification over specialization seems to be the norm. A possible explanation is that households, and especially farming households, are generally underemployed and are able to participate in additional activities without hitting a labor supply constraint (McCullough 2017). 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.

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. From a distributional standpoint, improving farm productivity is the best way to improve the welfare of poor households.

References

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

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.

Bustos, P., B. Caprettini, and J. Ponticelli. 2016. "Agricultural Productivity and Structural Transformation: Evidence from Brazil." *American Economic Review* 106:1320–1365.

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.

Filmer, D., and L. Fox. 2014. *Youth Employment in Sub-Saharan Africa*. Africa Development Forum, World Bank Publications.

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, pp. 3051–3083.

Gollin, D., D. Lagakos, and M.E. Waugh. 2014a. "Agricultural Productivity Differences across Countries." *American Economic Review* 104:165–170.

—. 2014b. "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.B. 2017. "Labor productivity and employment gaps in Sub-Saharan Africa." *Food Policy* 67:133–152.

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

McMillan, M., D. Rodrik, and Ì. Verduzco-Gallo. 2014. "Globalization, Structural Change and Productivity Growth, with an Update on Africa." *World Development* 63:11–32.

Ranis, G. 2004. "The Evolution of Development Thinking: Theory and Policy." *Yale University Economic Growth Center Discussion Paper 886*, pp. 40pp.

Rodrik, D. 2016. "An African Growth Miracle?" *Journal of African Economies* 2016:1–18.

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*, 2nd ed. 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*. Washington, D.C.: World Bank.

Figures

Figure 1: Transition Matrix, farming

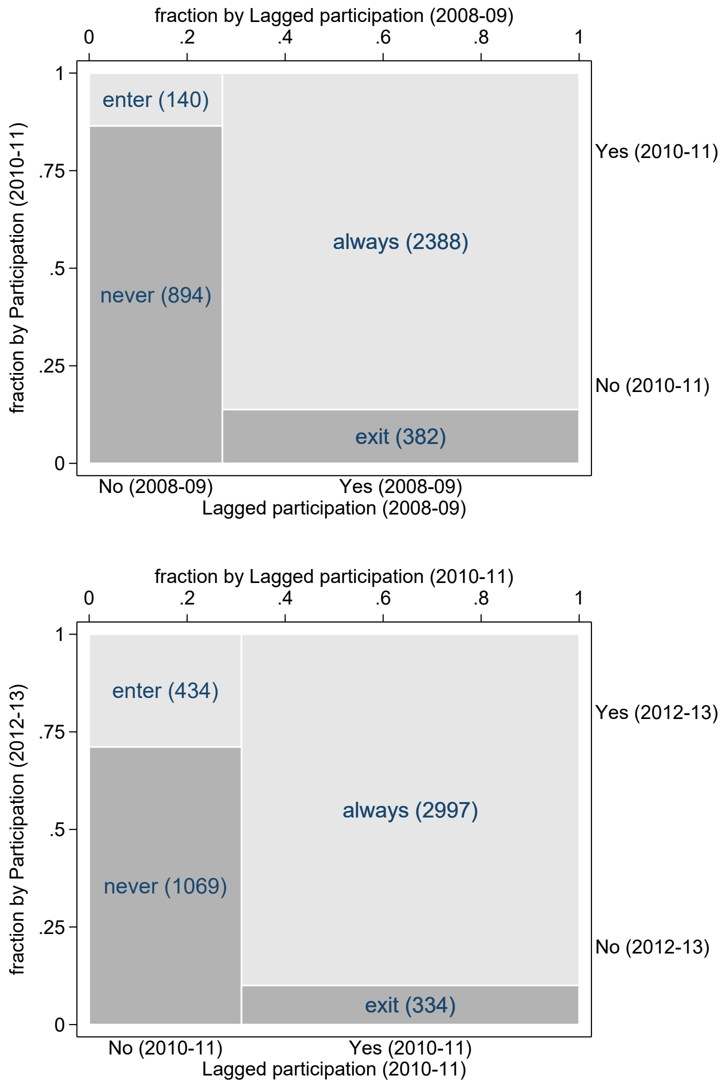


Figure 2: Transition Matrix, self employment

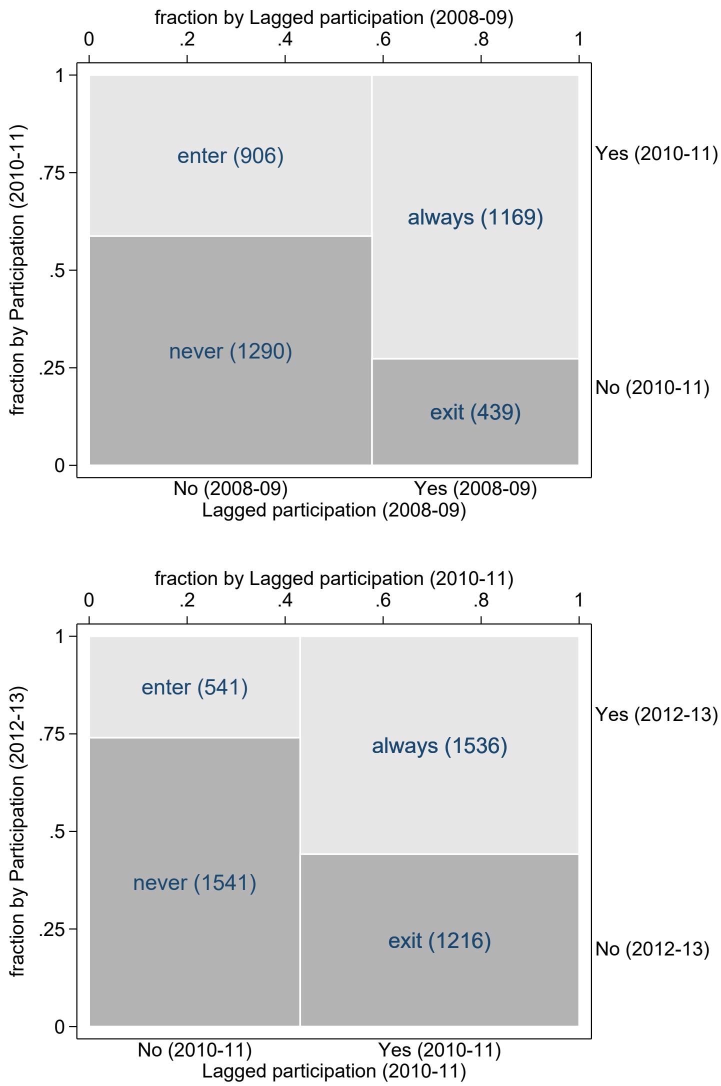


Figure 3: Transition Matrix, wage employment

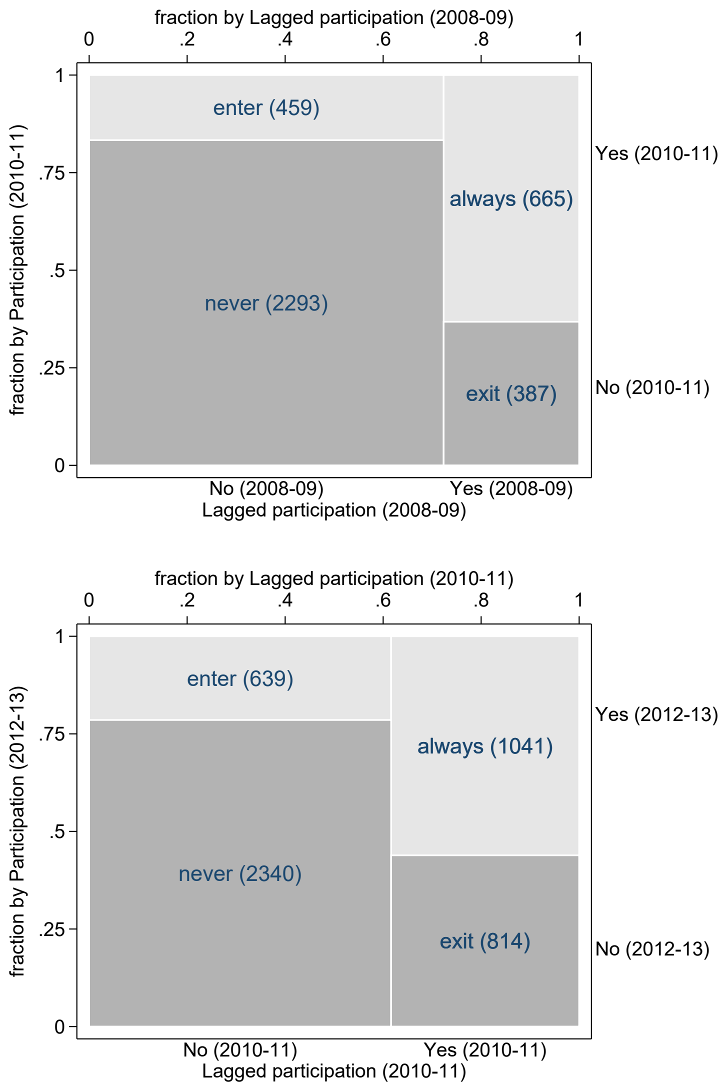


Figure 4: The top figure depicts average actual labor productivity (annual net returns per worker) across occupational choices. The box plots show mean productivity as a horizontal line, the 25th to 75th percentiles within a box, and the data range excluding outside values in the whisker lines. The bottom figure compares labor productivity of households who have chosen an activity with that of predicted labor productivity for the same households in their non-chosen activities.

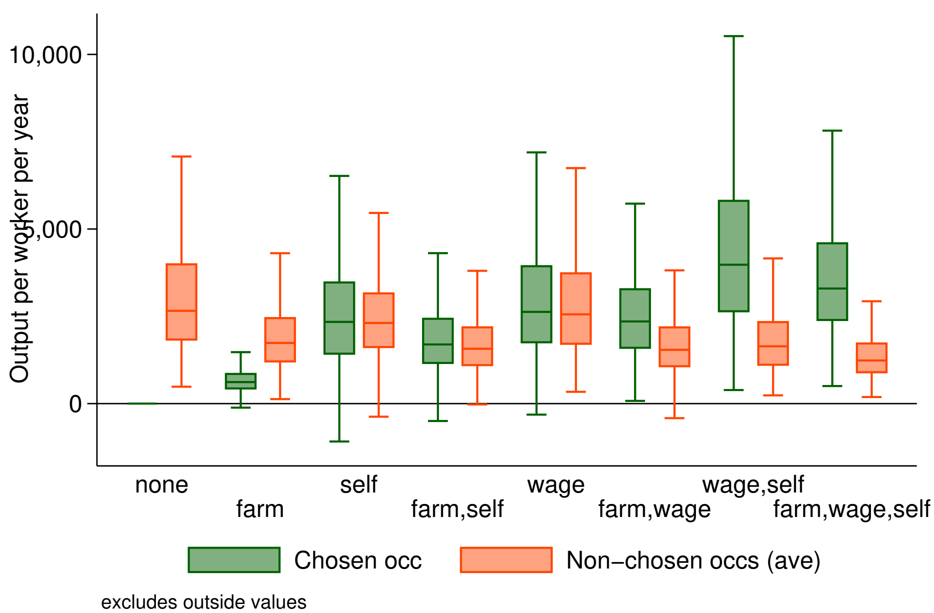


Figure 5: Comparison of predicted choice probabilities and actual choice shares.

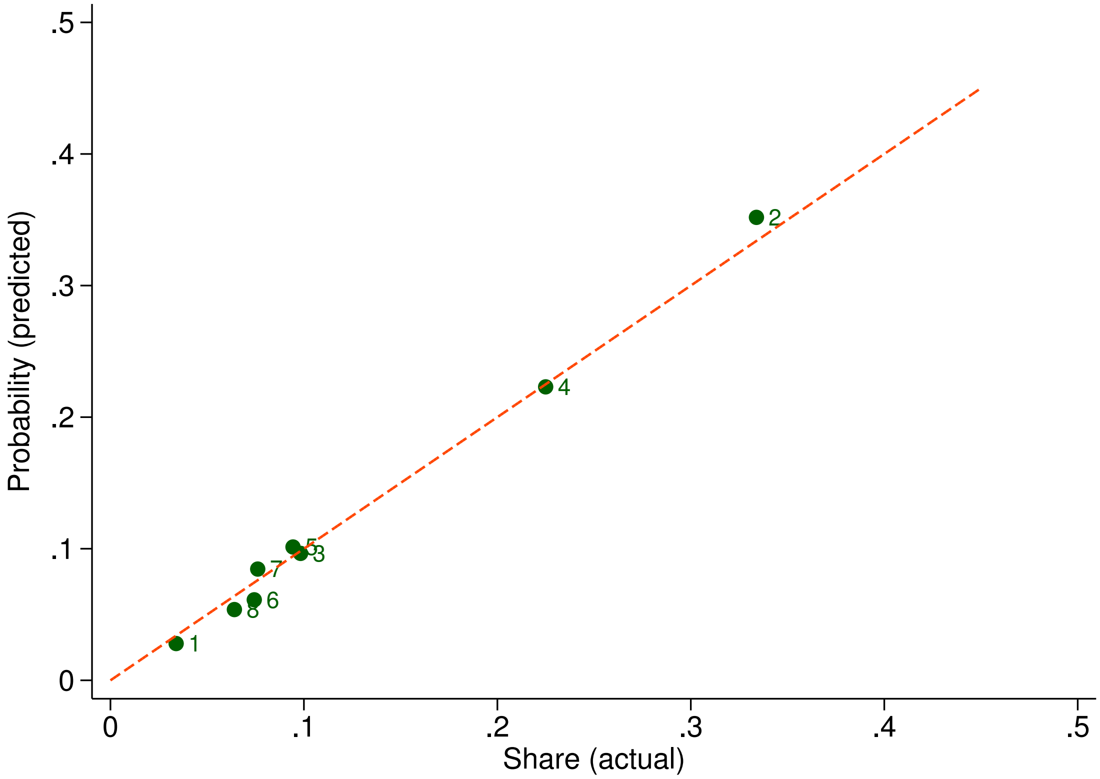


Figure 6: Comparison of predicted choice probabilities and actual choice shares by population sub-groups.

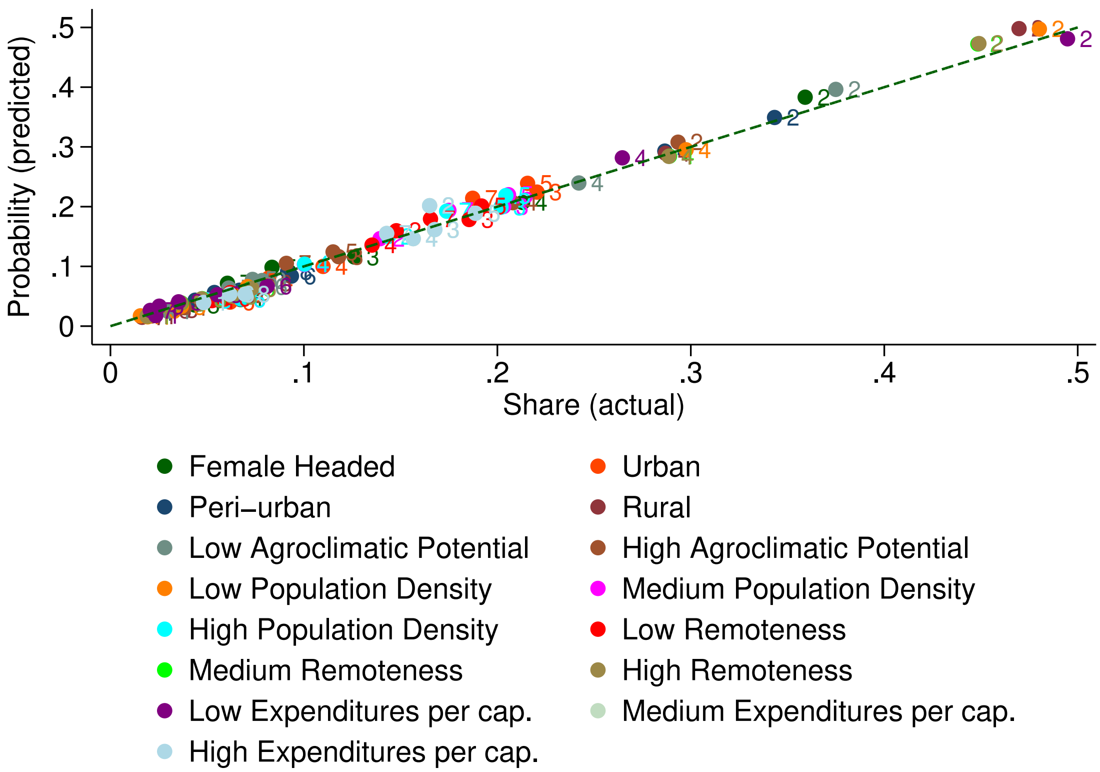


Figure 7: Validation: comparison of predicted choice probabilities with actual choice shares for 20% of sample that was randomly omitted from the estimation sample.

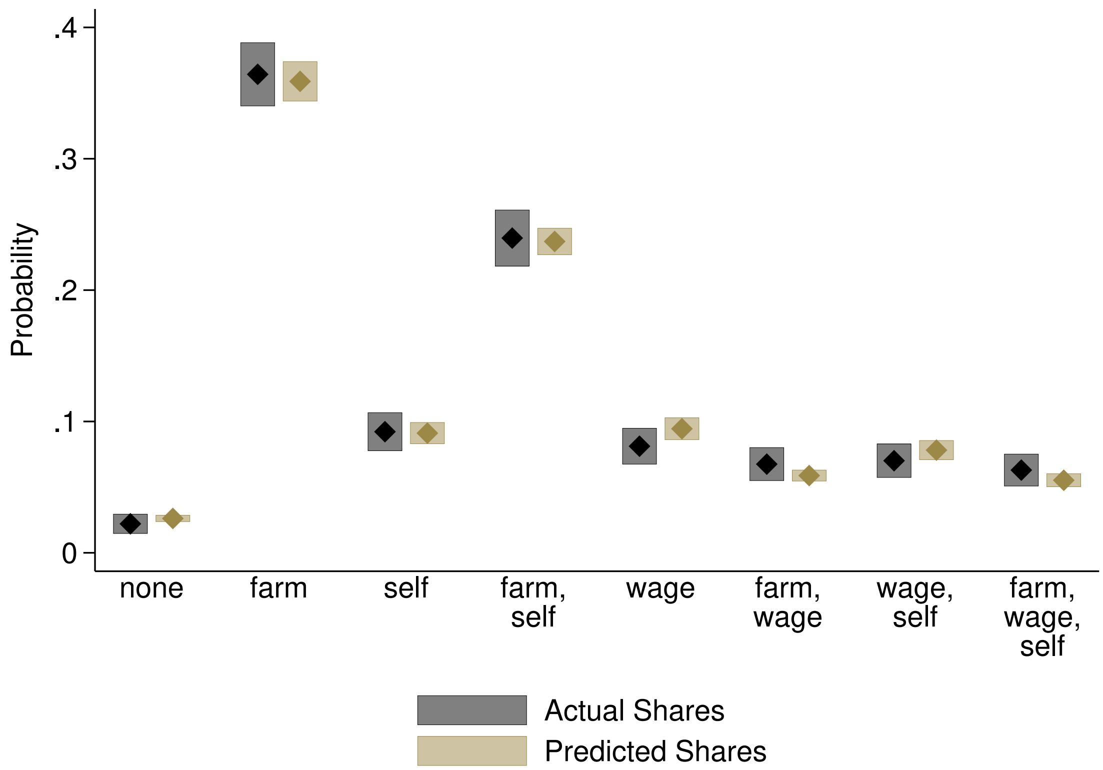


Figure 8: Conditional probability of participation in each occupation over agroclimatic potential. The probability of each choice corresponding with a specific level of agroclimatic potential is the vertical distance between the line above and the line below the area labeled with that choice.

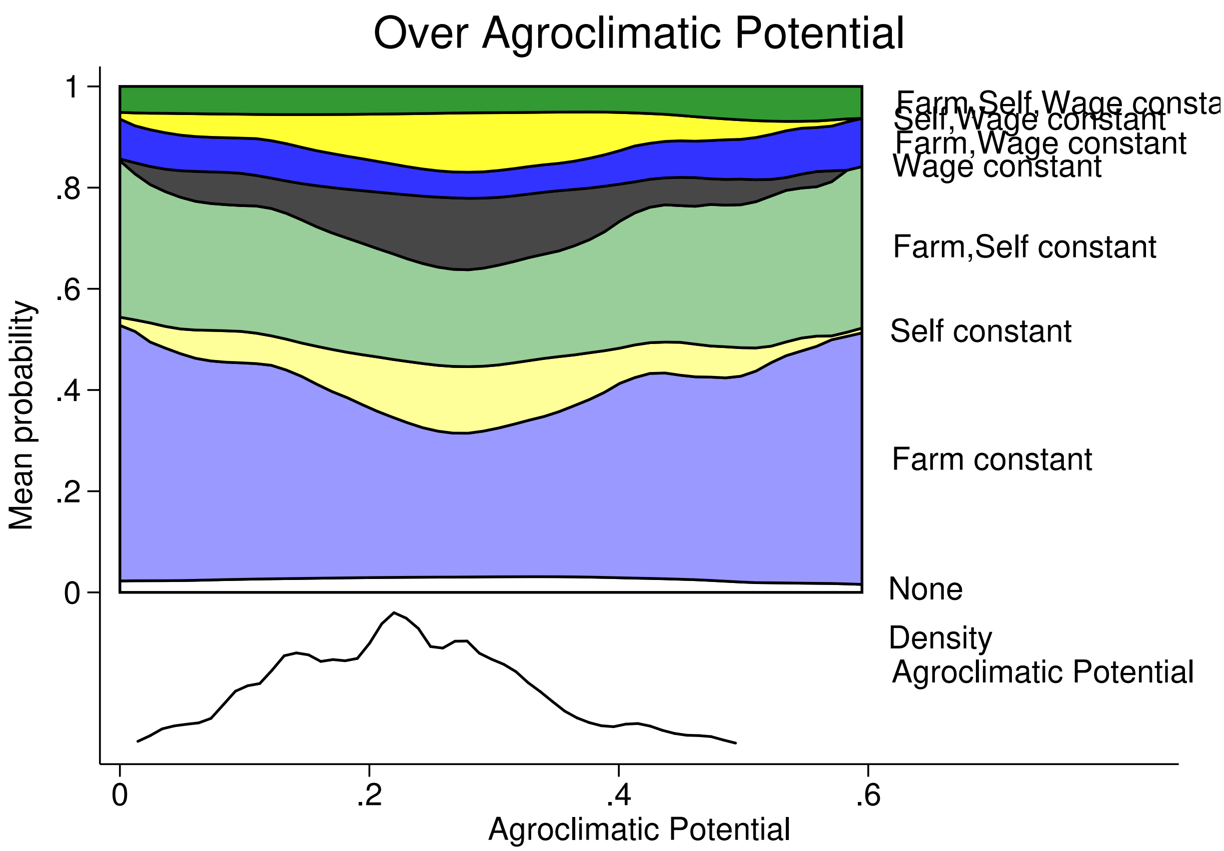


Figure 9: Conditional probability of participation in each occupation over population density. The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.

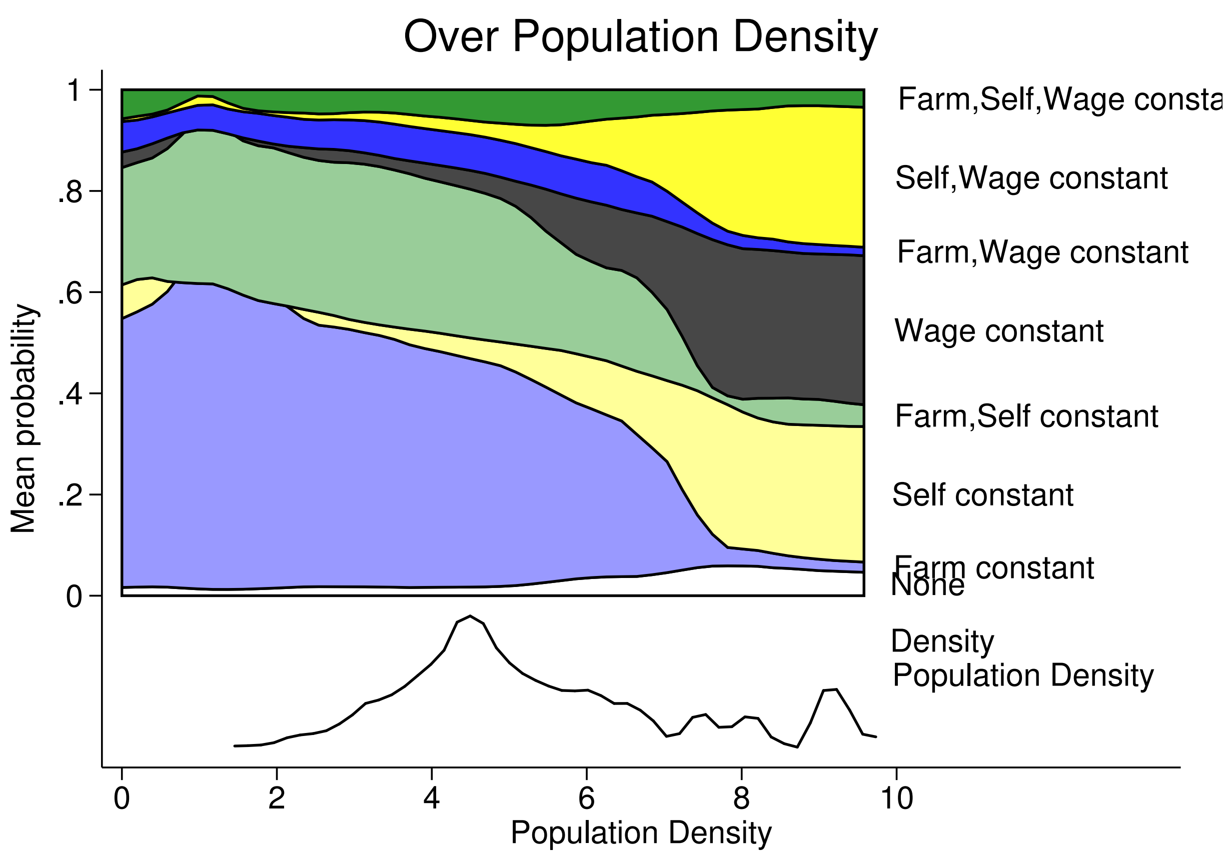


Figure 10: Conditional probability of participation in each occupation over remoteness (log of distance in km to the nearest population center of >500k). The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.

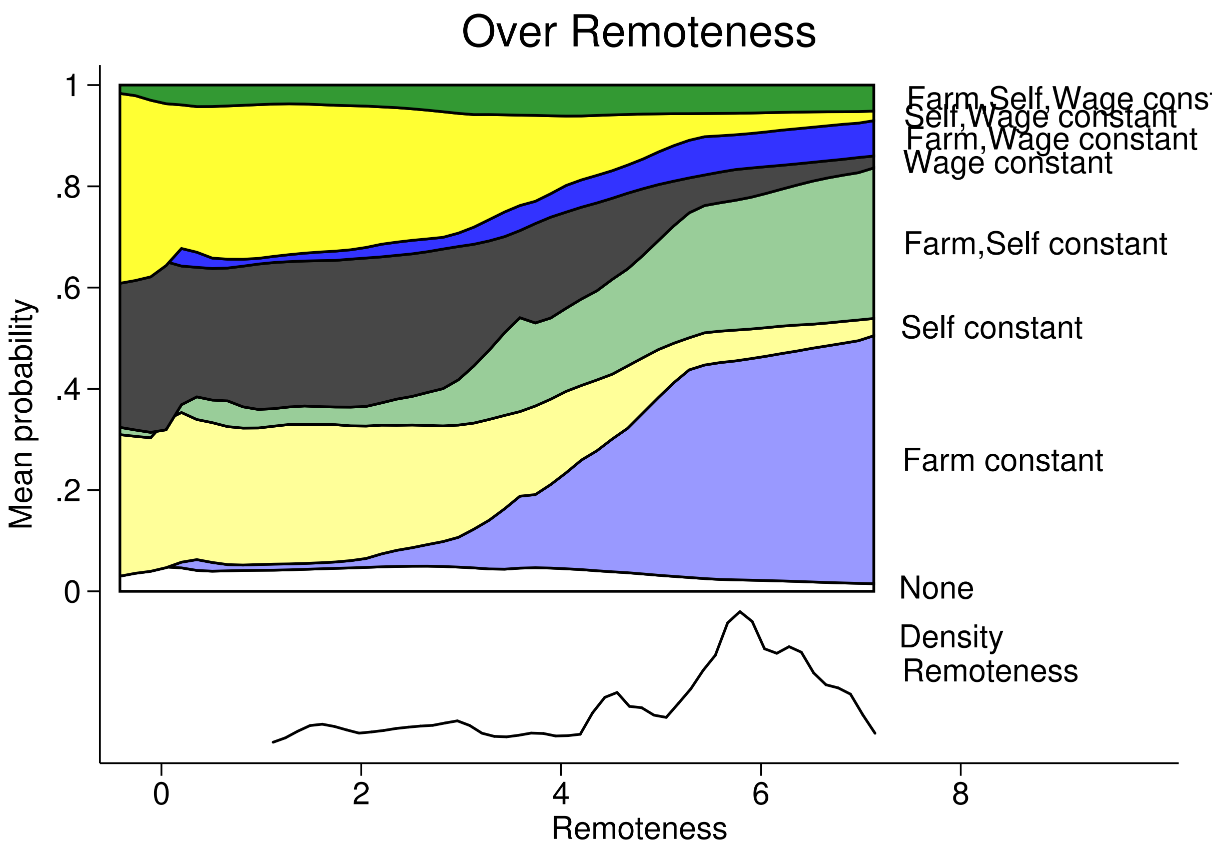


Figure 11: Difference in probability of participation in farming, self employment and wage employment across labor productivity simulations. In the farm simulation (depicted in green), farm labor productivity was doubled. In the self employment simulation (depicted in blue), self employment labor productivity 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.

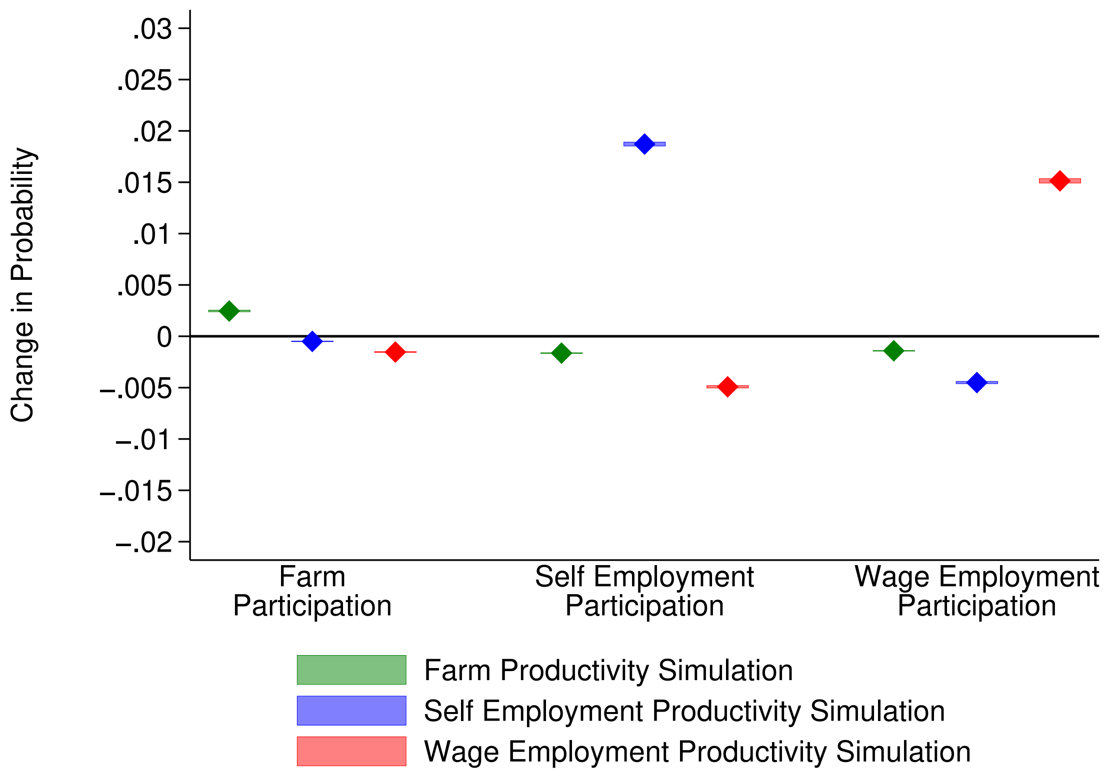


Figure 12: Difference in probability of participating in each occupational choice (combination of farming, self employment, and wage employment) across labor productivity simulations described in Figure 11.

Figure 13: 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 11.

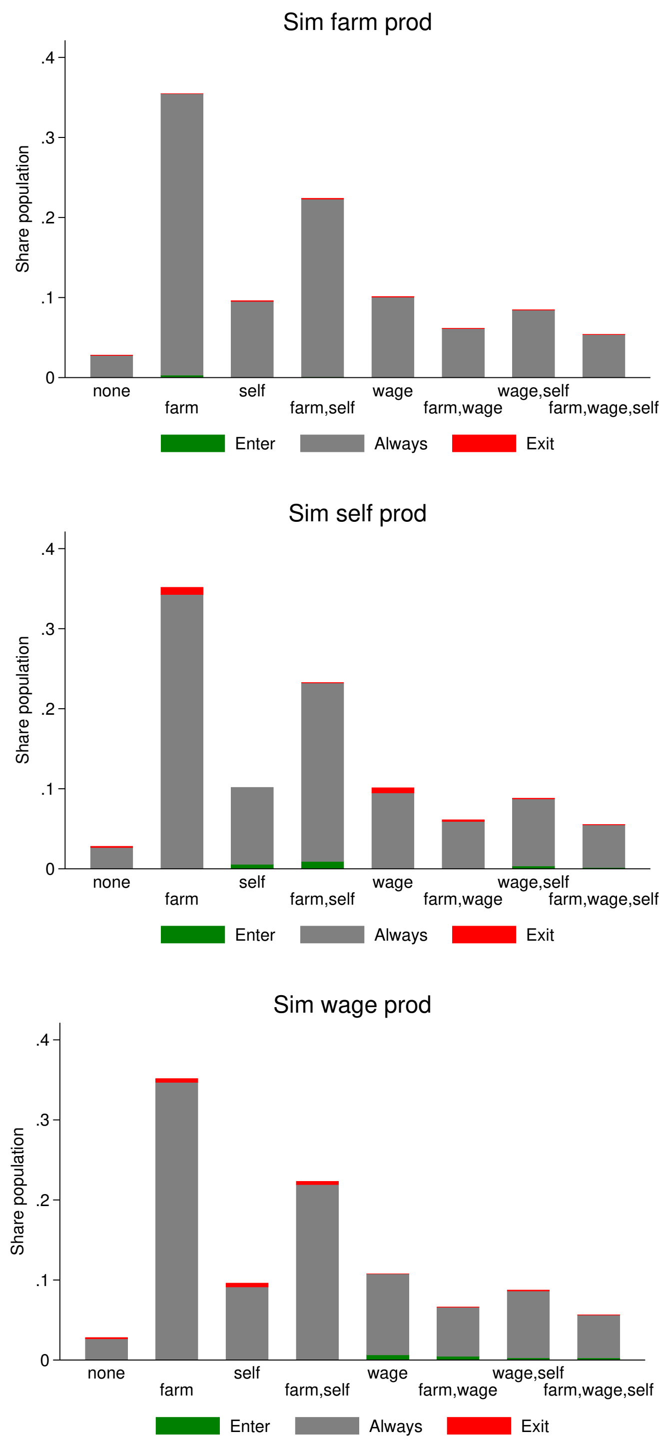
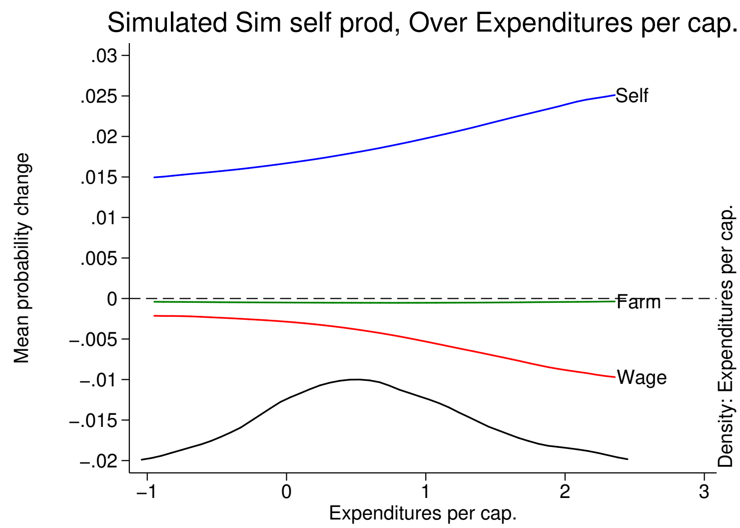


Figure 14: Simulated effect of doubling self employment (top panel) and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on household expenditures. The density of the household expenditures variable is shown underneath each regression.



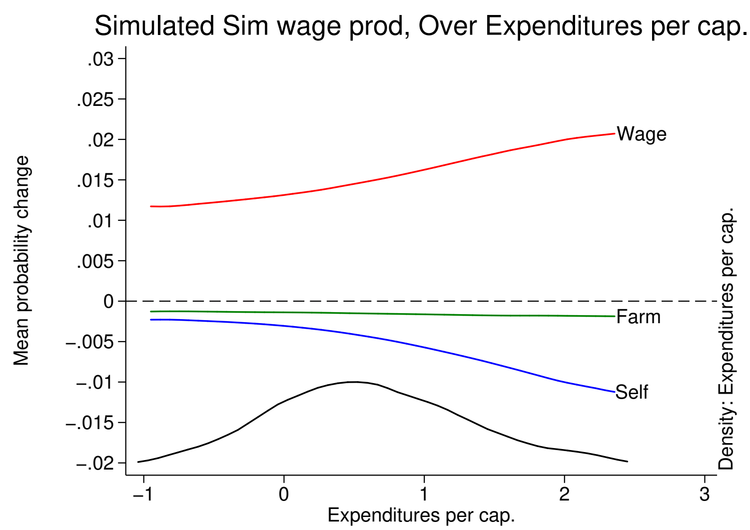


Figure 15: Simulated effect of doubling self employment (top panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on population density. The density of population density is shown underneath each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

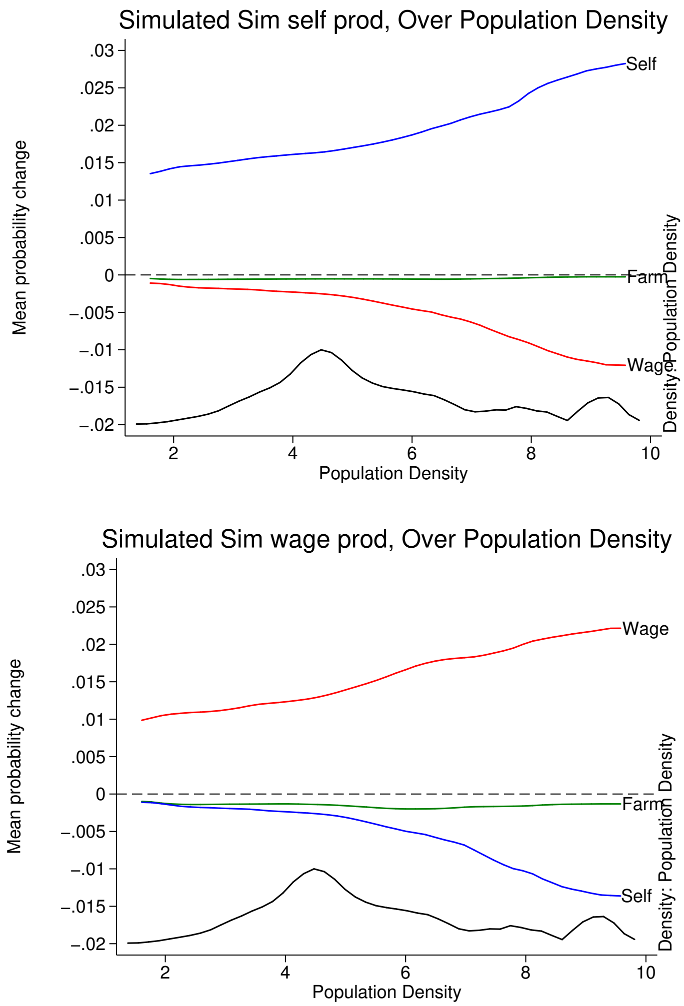


Figure 16: Simulated effect of doubling self employment (top panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on remoteness. The density of remoteness is shown underneath each regression.

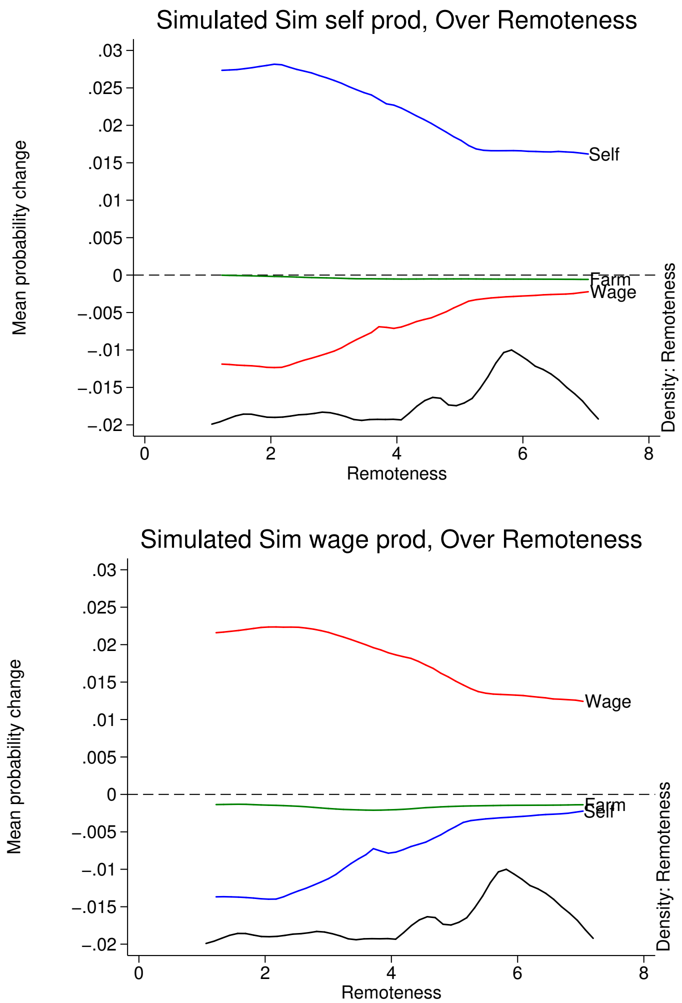


Figure 17: Welfare effects (expressed as an equivalent income change relative to baseline income) of each policy simulation. For each of the three activities – farming, self employment and wage employment – the average welfare gains are depicted for households based on whether they enter into the activity (when non-participating households shift to participation), exit from the activity (when participating households cease participation), or always participate in the activity (households participate before and after the intervention).

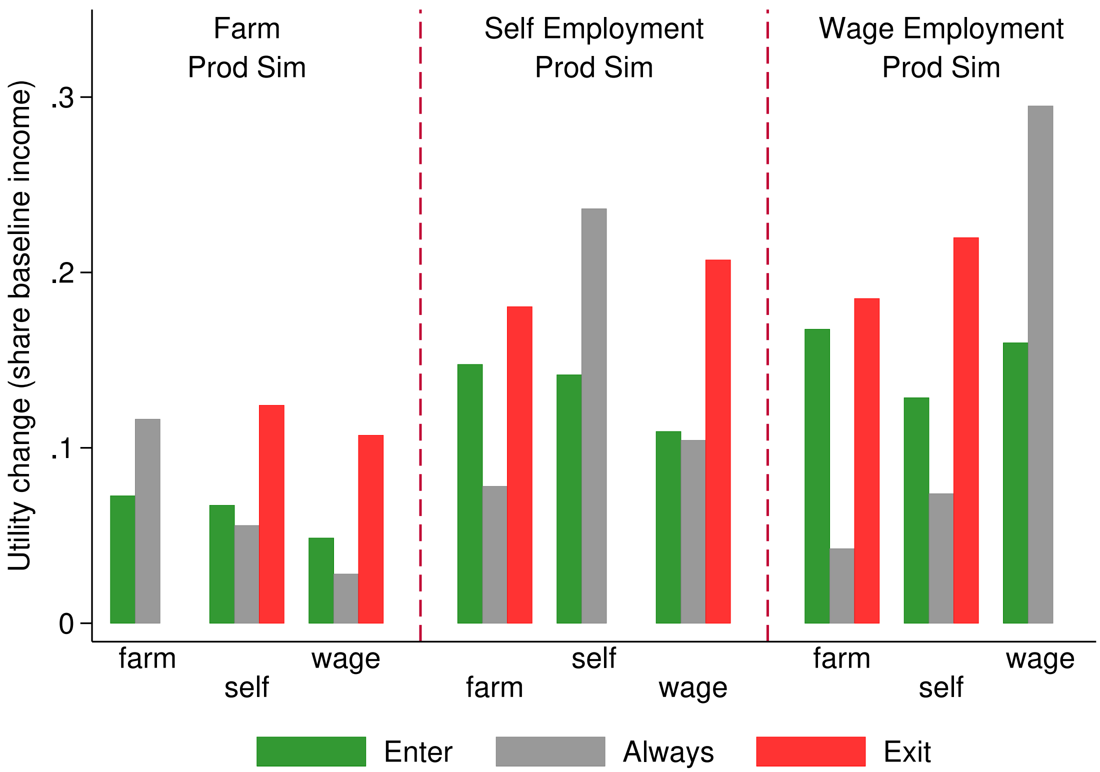


Figure 18: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over household expenditures.

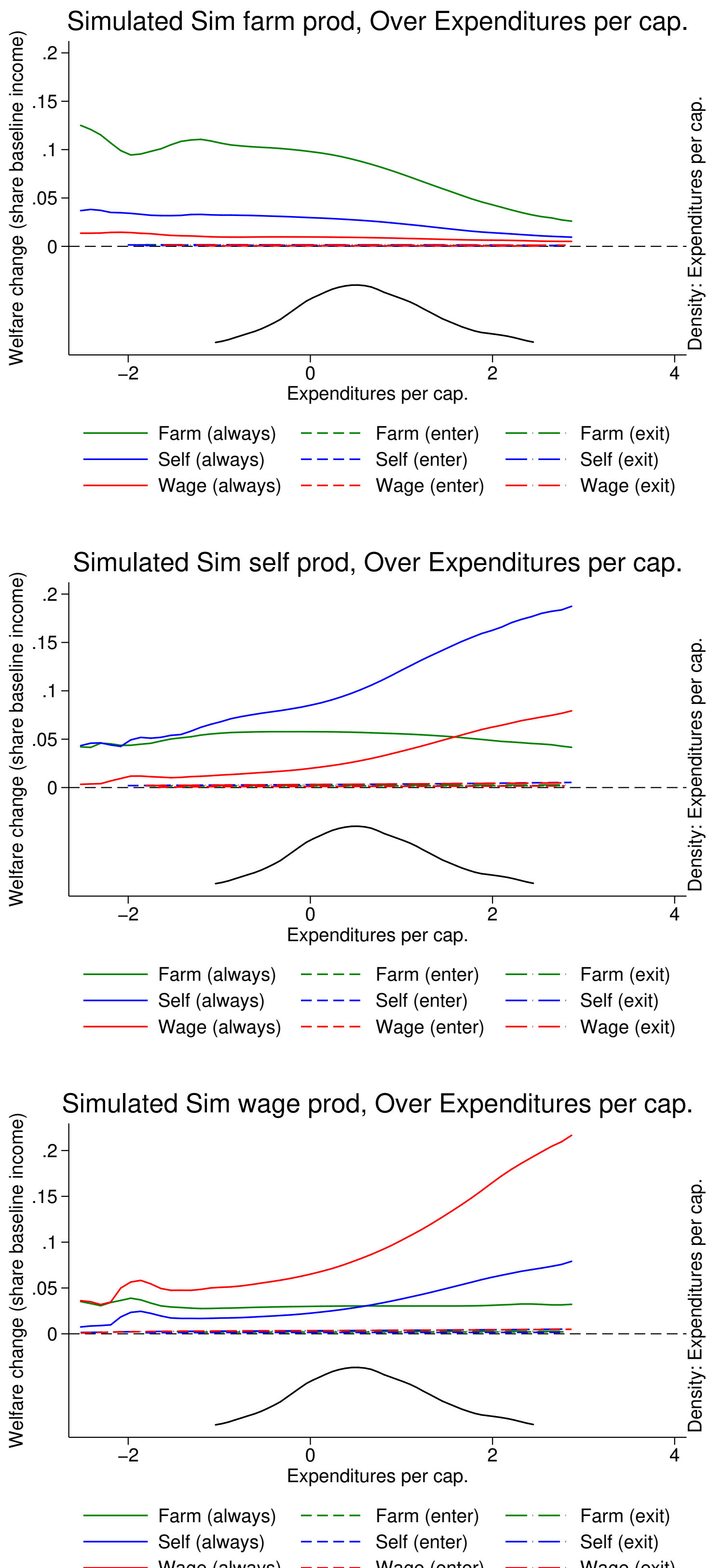
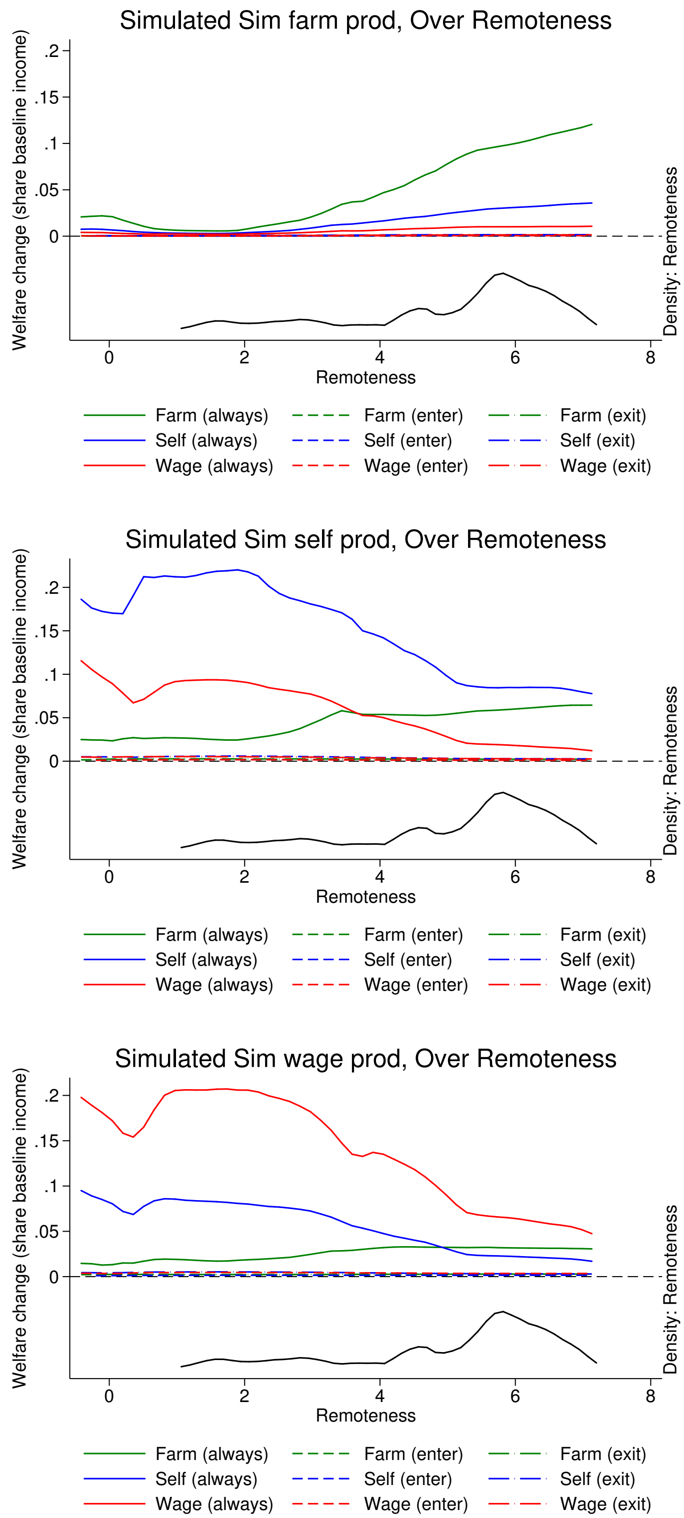


Figure 19: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over remoteness.



Tables

Table 1: Participation Table, Tanzania, 2008-09

	f0e0m0	f0e0m1	f0e1m0	f0e1m1	f1e0m0	f1e0m1	f1e1m0	f1e1m1
Rural:								
Number HHs	35	24	37	9	1124	151	539	95
Share HHs	0.0142	0.00738	0.0111	0.00305	0.566	0.0656	0.283	0.0491
Per capita consumption, TSH	993.7	1457.8	1885.8	2119.3	753.6	1076.5	860.1	984.8
(sd)	662.7	830.6	1782.2	1370.6	496.2	787.5	622.6	627.4
Urban								
Number HHs	87	249	276	176	103	68	110	68
Share HHs	0.0722	0.168	0.259	0.137	0.108	0.0614	0.124	0.0709
Per capita consumption, TSH	1745.2	2467.9	1949.5	2078.4	968.6	1300.1	1351.8	1553.0
(sd)	1610.0	1818.1	1499.3	1378.8	1005.8	931.3	919.9	1237.4

Table 2: Participation Table, Tanzania, 2010-11

	f0e0m0	f0e0m1	f0e1m0	f0e1m1	f1e0m0	f1e0m1	f1e1m0	f1e1m1
Rural:								
Number HHs	77	62	103	58	982	136	878	160
Share HHs	0.0291	0.0199	0.0288	0.0160	0.417	0.0502	0.381	0.0582
Per capita consumption, TSH	1178.3	2390.8	1468.9	1669.3	713.8	791.1	787.8	1044.8
(sd)	1216.9	1811.5	1387.3	1130.0	507.1	485.6	527.6	741.5
Urban								
Number HHs	82	260	326	308	82	48	150	92
Share HHs	0.0465	0.164	0.270	0.201	0.0704	0.0293	0.142	0.0767
Per capita consumption, TSH	2613.1	2536.6	1793.1	1999.4	872.6	1339.8	1104.7	1432.1
(sd)	2118.1	1658.1	1552.6	1356.6	534.4	1059.4	710.8	830.3

Table 3: Participation Table, Tanzania, 2012-13

	f0e0m0	f0e0m1	f0e1m0	f0e1m1	f1e0m0	f1e0m1	f1e1m0	f1e1m1
Rural:								
Number HHs	54	139	109	59	1492	334	781	227
Share HHs	0.0151	0.0362	0.0227	0.0132	0.496	0.106	0.243	0.0676
Per capita consumption, TSH	1789.9	2331.1	1793.1	1769.9	771.0	977.4	907.1	1008.9
(sd)	1499.9	1677.8	1713.8	1082.0	564.8	695.8	626.6	764.0
Urban								
Number HHs	66	381	307	288	152	139	193	113
Share HHs	0.0370	0.221	0.208	0.157	0.0967	0.0767	0.140	0.0640
Per capita consumption, TSH	2239.1	2677.9	2322.1	2240.3	1083.5	1615.0	1393.2	1995.1
(sd)	2001.9	1636.9	1812.7	1460.1	938.8	1181.5	996.0	1268.4

Table 4: First Stage GL with Selection, all profit variables

	Farm Margins at means	(SE)	Self Emplm't Margins at means	(SE)	Wage Emplm't Margins at means	(SE)
Urban (yes)	-120.71	82.93	-516.56	401.91	-1,207.66*	487.09
Peri-urban dummy (yes)	370.63	216.78	-1,228.55	1,077.52	-1,760.56	1,067.32
Female head (yes)	-105.73**	39.80	-648.81**	195.08	-715.91*	311.39
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	164.78	100.68	-533.77	474.51	-1,049.66*	494.00
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	5.04	14.69	15.54	63.79	-28.38	71.00
Dist nrst major rd (km, LSMS-ISA) (sq rt)	33.78	21.70	130.10	101.65	-55.55	123.28
People per square km, 2005 (ln, HC) (sq rt)	-20.20	178.06	347.21	940.33	803.78	1,046.37
Number hh members 16-65 (sq rt)	260.83**	70.75	341.45	415.13	378.85	560.96
Age of head (sq rt)	-23.82	33.37	-245.33	170.14	509.65*	226.59
Yrs educ, adults (ave) (sq rt)	38.71	40.78	329.29	218.60	1,651.94**	418.96
Yield Potential low (cross-crop ind) (sq rt)	-674.73	1,051.81				
Yield Potential high (cross-crop ind) (sq rt)	-4.65	558.61				
Rate improved maize seed use (mean smth) (sq rt)	-83.18	156.64				
Cost hired lbr (med smth, USD/day) (sq rt)	-130.42	89.83				
Rate of tractor use (mean smth) (sq rt)	131.74	223.68				
Land owned (ha, RIGA) (sq rt)	309.24**	51.31				
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	12.11	19.03				
Slope (pct, USGS) (sq rt)	145.83**	54.00				
Share land irrigated (percent, FAO) (sq rt)	912.84**	325.00				
Soil nutrient constraints (FAO) (moderate)	-47.14	51.92				
Soil nutrient constraints (FAO) (severe)	18.33	97.96				
Soil nutrient constraints (FAO) (other)	3,346.57*	1,303.41				
Soil workability constraints (FAO) (moderate)	-55.94	54.71				
Soil workability constraints (FAO) (severe)	-123.17	101.46				
Soil workability constraints (FAO) (other)	-2,763.79*	1,270.86				
Cost/hired worker (med smth, USD) (sq rt)			20.53	19.50		
Nighttime light ave coverage(DMSP F16) (sq rt)			30.08	22.48		
Productive non-ag asset ind, fact 1 (sq rt)			3,071.75**	525.81		
Productive ag-related asset ind, fact 1 (sq rt)			-4,106.91**	1,209.10		
Financial service available (yes)			277.18	167.19		
Max educ in hh (yrs) (sq rt)					198.73	524.35
Nighttime light intensity(DMSP F16) (sq rt)					-367.30	189.33
Returns/ind worker (med smth, USD) (sq rt)					11.04	14.56
Returns/ser worker (med smth, USD) (sq rt)					50.88**	12.90
Partpn in ind emplmt (med smth sh) (sq rt)					-398.99	976.37
Partpn in ser emplmt (med smth sh) (sq rt)					574.60	1,124.43
R2 (adj)	0.235		0.194		0.298	
Mean profits (USD/yr)	1377.36		3552.16		4496.09	
N (non-censored)	8217		5462		3644	
N	11789		11789		11789	

Table 5: First Stage GL with Selection, all selection variables

	Farm Select	(SE)	Self Emplm't Select	(SE)	Wage Emplm't Select	(SE)
Urban (yes)	-0.42**	0.01	0.18**	0.01	0.26**	0.01
Peri-urban dummy (yes)	-0.14**	0.02	0.10**	0.02	0.11**	0.02
Transfers recvd (USD)	-0.00**	0.00	-0.00**	0.00	-0.00	0.00
Household size	0.04**	0.00	0.02**	0.00	0.02**	0.00
HH dependent share	0.13**	0.02	-0.15**	0.02	-0.36**	0.02
Years educ, head	-0.02**	0.00	0.00*	0.00	0.02**	0.00
Ave length of season (days, MOD12Q2)	-0.00**	0.00	-0.00**	0.00	0.00**	0.00
Head's father attended school (yes)	-0.06**	0.01	0.08**	0.01	0.05**	0.01
Lambda	-289.98		-960.66		-1402.40	
Sigma	1141.12		5345.79		4430.79	
P value comparison test	0.00		0.00		0.00	
N	11789		11789		11789	
N (censored)	3572		6327		8145	
N (non-censored)	8217		5462		3644	

Table 6: Second stage occupational choice model parameter coefficients. The base case is non-participation in all activities. The first row shows the coefficient in the utility function for the natural log of income, which is held constant across occupational choices. The next rows show alternate-specific coefficients for each of the selection variables, and the mean and standard deviation of the alternative-specific random coefficients (the diagonal of the variance-covariance matrix describing the multivariate normal distribution of δ.

	farm	self	farm,self	wage	farm,wage	wage,self	farm,wage,self
Ln Income (pc)	0.4312**	0.4312**	0.4312**	0.4312**	0.4312**	0.4312**	0.4312**
	0.1170	0.1170	0.1170	0.1170	0.1170	0.1170	0.1170
Transfers recvd (USD)	-0.5451*	-0.4043**	-0.9494**	-0.3196	-0.8646**	-0.7239**	-1.2690**
	0.2366	0.1476	0.2847	0.1822	0.3139	0.2469	0.3601
Household size	0.4299**	0.1632**	0.5931**	0.1782**	0.6081**	0.3414**	0.7713**
	0.0259	0.0113	0.0292	0.0138	0.0314	0.0198	0.0353
HH dependent share	0.5733**	-0.9182**	-0.3449	-2.1013**	-1.5280**	-3.0194**	-2.4462**
	0.2058	0.1211	0.2428	0.1595	0.2693	0.2121	0.3066
Years educ, head	-0.1253**	0.0173*	-0.1080**	0.1314**	0.0061	0.1487**	0.0234
	0.0144	0.0074	0.0156	0.0098	0.0164	0.0132	0.0182
Ave length of season (days, MOD12Q2)	-0.0241**	-0.0035*	-0.0276**	0.0023	-0.0218**	-0.0011	-0.0253**
	0.0030	0.0014	0.0033	0.0018	0.0035	0.0024	0.0038
Head's father attended school	-0.5043**	0.3733**	-0.1310	0.2896**	-0.2147	0.6629**	0.1586
	0.1016	0.0568	0.1166	0.0684	0.1267	0.0935	0.1420
Urban	-4.9929**	0.9429**	-4.0500**	1.3372**	-3.6556**	2.2801**	-2.7127**
	0.2339	0.0936	0.2437	0.1035	0.2422	0.1490	0.2571
Peri-urban dummy	-1.2026**	0.5522**	-0.6504*	0.7097**	-0.4929	1.2619**	0.0593
	0.2442	0.1208	0.2726	0.1385	0.2827	0.1902	0.3114
δ (mean)	8.1382	0.2169	7.2690	-2.0053	4.1949	-3.5308	3.1188
δ (sd)	6.1823	2.5025	7.4720	1.0479	6.9144	2.1925	7.8066
N	11784						
Log-likelihood	-16895.29						
Pseudo R2	0.22						
AIC	1538.17						
BIC	1988.01						

Table 7: 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	wage,self	farm,wage,self
Traveltime 500K	-0.178	-0.735	0.904	1.560	-1.401	-0.467	0.243	0.073
	1.014	1.952	2.778	2.522	3.728	1.135	1.596	0.467
Netdist 100K	-0.002	-0.003	-0.009	-0.023	0.010	0.017	0.003	0.007
	0.011	0.028	0.041	0.035	0.044	0.035	0.015	0.015
Dist Road	0.028	0.220	-0.002	-0.145	0.006	-0.054	0.003	-0.057
	0.304	0.754	0.233	0.411	0.183	0.251	0.390	0.307
Popdensity Ln	1.948	1.474	6.304	8.895	-6.403	-7.587	-1.792	-2.837
	8.269	17.319	20.474	19.763	23.176	14.834	8.062	6.068
Pop Indep	-0.606	4.746	-5.447	-15.076	5.491	7.400	1.581	1.912
	2.120	13.600	14.166	22.214	15.079	14.401	5.742	4.748
Age Head	0.637	3.097	-0.424	-2.523	0.828	0.100	-1.008	-0.706
	1.647	3.511	1.505	3.162	1.836	0.566	1.937	1.006
Educ Yrs Ave	-1.914	-2.789	-5.461	-3.656	4.643	3.289	3.868	2.020
	5.218	6.398	13.454	6.363	12.175	6.360	8.246	3.548
Rypot3 Low	-1,207.219	2,923.688	-2,073.930	1,582.382	-1,675.221	685.420	-810.548	575.280
	5,616.907	8,892.180	6,167.135	4,909.888	5,428.440	2,716.614	2,076.062	1,963.662
Rypot3 High	-29.993	53.176	-27.776	19.197	-18.794	7.177	-10.107	7.120
	384.891	459.592	142.098	100.984	108.373	49.632	41.965	41.196
Tech Use Maize	4.299	-10.568	7.776	-6.034	6.169	-2.534	3.117	-2.226
	43.028	68.153	47.336	38.963	39.531	17.332	15.801	13.874
Cpd Farmhire	-0.207	0.505	-0.449	0.376	-0.351	0.163	-0.195	0.159
	1.240	2.061	1.695	1.311	1.404	0.744	0.577	0.600
Tractor Share	8.272	-19.701	13.286	-10.038	12.079	-5.460	5.967	-4.405
	78.420	137.282	103.677	91.117	97.505	55.393	50.066	48.710
Landowned	0.302	-0.544	0.305	-0.226	0.245	-0.105	0.124	-0.101
	4.646	6.257	3.714	2.828	2.936	1.700	1.481	1.433
Rainfall Mean	0.012	-0.029	0.020	-0.016	0.016	-0.007	0.008	-0.006
	0.055	0.077	0.048	0.038	0.042	0.023	0.016	0.016
Slope	-1.383	3.328	-2.339	1.823	-1.991	0.831	-0.947	0.678
	6.748	9.873	6.322	5.042	6.047	3.346	2.154	2.041
Irrigation Sh	0.114	-0.236	0.221	-0.176	0.112	-0.049	0.080	-0.065
	3.307	4.211	3.148	2.491	1.577	0.838	0.965	0.946
Cpw Enthire	0.000	0.003	-0.000	-0.003	-0.000	0.000	0.000	-0.000
	0.001	0.004	0.003	0.004	0.003	0.001	0.002	0.001
Lightsum	0.001	0.000	-0.002	-0.000	0.003	0.000	-0.002	-0.000
	0.003	0.002	0.007	0.002	0.010	0.001	0.007	0.001
Assets Na	9.566	6.243	-28.867	-7.180	38.777	8.859	-20.688	-6.711
	253.070	769.841	487.055	756.121	665.412	143.658	400.970	152.595
Assets Ag	-85.675	-604.731	140.080	596.157	138.282	-269.809	-119.706	205.415
	5,083.631	9,261.977	11,737.672	8,662.127	16,905.722	2,432.506	11,356.282	2,150.950
Educ Yrs Max	-1.764	-4.772	-5.324	-3.862	5.276	4.475	3.170	2.801
	4.899	8.248	11.065	5.446	11.092	7.740	6.117	4.261
Lightintensity	-0.036	-0.009	-0.120	-0.015	0.095	0.007	0.069	0.009
	0.273	0.156	0.641	0.172	0.584	0.136	0.347	0.129
Rpw Ind	-0.000	-0.001	-0.001	-0.001	0.001	0.001	0.001	0.000
	0.001	0.002	0.003	0.001	0.003	0.002	0.002	0.001
Rpw Ser	-0.001	-0.001	-0.002	-0.001	0.002	0.001	0.001	0.001
	0.002	0.002	0.005	0.002	0.004	0.002	0.003	0.001
Partshare Ind	-487.351	-1,036.959	-1,550.628	-845.953	1,536.995	950.748	849.499	583.504
	4,055.827	7,301.098	11,073.778	5,344.932	10,636.342	6,704.935	5,987.446	3,844.634
Partshare Ser	-386.138	-735.180	-1,169.979	-623.124	1,142.156	662.873	680.151	429.183
	1,862.616	3,425.056	4,487.386	2,528.784	4,539.063	3,255.502	2,514.613	1,862.790
Urban (=1)	-0.009	-0.002	-0.014	-0.020	0.006	0.010	0.024	0.004
	0.046	0.061	0.093	0.048	0.114	0.051	0.090	0.029
Periurban (=1)	0.017	-0.003	-0.016	-0.006	0.050	-0.000	-0.037	-0.004
	0.060	0.065	0.120	0.045	0.175	0.050	0.133	0.035
Female Head (=1)	0.019	0.020	0.010	0.018	0.008	-0.015	-0.048	-0.012
	0.058	0.050	0.064	0.032	0.103	0.039	0.117	0.027
Nutrientavail (=2)	0.001	-0.002	0.002	-0.001	0.001	-0.001	0.001	-0.000
	0.003	0.006	0.005	0.004	0.004	0.002	0.002	0.002
Nutrientavail (=3)	-0.002	0.005	-0.004	0.003	-0.003	0.001	-0.001	0.001
	0.012	0.016	0.009	0.007	0.008	0.005	0.003	0.003
Nutrientavail (=4)	-0.006	0.015	-0.011	0.009	-0.009	0.004	-0.005	0.003
	0.031	0.046	0.029	0.023	0.028	0.015	0.010	0.010
Soilworkability (=2)	-0.002	0.004	-0.003	0.002	-0.003	0.001	-0.001	0.001
	0.008	0.012	0.007	0.006	0.007	0.004	0.003	0.002
Soilworkability (=3)	-0.003	0.008	-0.006	0.004	-0.005	0.002	-0.002	0.002
	0.016	0.022	0.014	0.011	0.012	0.007	0.005	0.004
Soilworkability (=4)	0.002	-0.005	0.004	-0.003	0.003	-0.001	0.002	-0.001
	0.014	0.023	0.015	0.012	0.013	0.007	0.005	0.005
Fs Available (=1)	-0.000	-0.011	0.000	0.011	0.002	-0.002	-0.001	0.001
	0.006	0.014	0.014	0.013	0.021	0.004	0.014	0.003

Table 8: 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	wage,self	farm,wage,self
Inctransfers	0.0132	0.0463	0.0114	-0.0439	0.0185	-0.0070	-0.0149	-0.0237
	0.0199	0.0413	0.0191	0.0313	0.0269	0.0115	0.0287	0.0285
Hhsize	-0.0078	-0.0166	-0.0109	0.0202	-0.0095	0.0066	0.0048	0.0131
	0.0122	0.0212	0.0158	0.0155	0.0143	0.0082	0.0135	0.0163
Hh Depshare	0.0295	0.1955	0.0388	-0.0127	-0.0336	-0.0525	-0.0909	-0.0741
	0.0556	0.1278	0.0968	0.1117	0.0714	0.0722	0.1280	0.0986
Educ Yrs Head	-0.0006	-0.0096	-0.0017	-0.0054	0.0056	0.0035	0.0055	0.0028
	0.0034	0.0077	0.0078	0.0075	0.0076	0.0054	0.0076	0.0055
Seasonlength	0.0002	-0.0002	0.0003	-0.0010	0.0007	-0.0000	0.0002	-0.0003
	0.0005	0.0009	0.0008	0.0008	0.0009	0.0004	0.0005	0.0006
Anyschool Headfather	-0.0032	-0.0622	0.0113	0.0203	-0.0000	-0.0027	0.0254	0.0112
	0.0130	0.0380	0.0234	0.0387	0.0230	0.0148	0.0351	0.0203
Urban	0.0183	-0.2539	0.0852	-0.0465	0.0833	-0.0168	0.1237	0.0066
	0.0962	0.1976	0.1855	0.2053	0.1609	0.1002	0.1745	0.1149
Periurban	-0.0038	-0.1122	0.0158	0.0104	0.0169	0.0027	0.0503	0.0198
	0.0280	0.0713	0.0518	0.0719	0.0468	0.0324	0.0686	0.0419

A Appendix

Table A.1: Summary statistics of regressors

	2008-09		2011-12		2013-14	
Hrs travel to nrst town >500k (LSMS-ISA)	6.138	(4.282)	6.104	(4.348)	5.977	(4.381)
Network dist to nrst town >100k (km, LSMS-ISA)	120.4	(98.26)	116.0	(96.78)	115.6	(97.75)
Dist nrst major rd (km, LSMS-ISA)	16.61	(20.35)	16.44	(20.41)	16.49	(20.61)
People per square km, 2005 (ln, HC)	5.000	(1.672)	5.066	(1.702)	5.087	(1.777)
Number hh members 16-65	2.527	(1.518)	2.550	(1.566)	2.472	(1.510)
Age of head	45.97	(15.70)	45.99	(15.91)	45.43	(16.05)
Yrs educ, adults (ave)	5.853	(3.582)	6.069	(3.734)	6.502	(3.840)
Urban	0.258	(0.438)	0.283	(0.450)	0.276	(0.447)
Peri-urban dummy	0.101	(0.301)	0.101	(0.302)	0.107	(0.309)
Female head	0.251	(0.434)	0.264	(0.441)	0.266	(0.442)
Yield Potential low (cross-crop ind)	0.127	(0.0495)	0.127	(0.0496)	0.127	(0.0484)
Yield Potential high (cross-crop ind)	0.529	(0.200)	0.515	(0.201)	0.519	(0.202)
Rate improved maize seed use (mean smth)	0.166	(0.189)	0.123	(0.170)	0.235	(0.229)
Cost hired lbr (med smth, USD/day)	2.020	(1.608)	2.486	(1.974)	2.517	(2.102)
Rate of tractor use (mean smth)	0.0243	(0.0731)	0.0340	(0.0930)	0.0816	(0.152)
Land owned (ha, RIGA)	1.291	(1.743)	1.320	(1.902)	1.332	(2.232)
Mean precip wettest qrtr (mm, NOAA CPC)	542.0	(168.7)	540.5	(168.8)	538.4	(165.2)
Slope (pct, USGS)	5.525	(5.479)	5.412	(5.326)	5.485	(5.484)
Share land irrigated (percent, FAO)	0.809	(3.397)	0.804	(3.370)	0.828	(3.445)
Soil nutrient constraints (FAO)	1.891	(0.836)	1.897	(0.840)	1.920	(0.864)
Soil workability constraints (FAO)	1.848	(0.944)	1.864	(0.950)	1.875	(0.966)
Cost/hired worker (med smth, USD)	523.8	(519.8)	581.9	(599.9)	705.6	(718.9)
Nighttime light ave coverage(DMSP F16)	405.8	(1096.2)	436.2	(1139.0)	493.6	(1231.5)
Productive non-ag asset ind, fact 1	0.792	(0.726)	0.799	(0.685)	0.832	(0.715)
Productive ag-related asset ind, fact 1	0.0299	(0.0820)	0.0334	(0.0449)	0.0277	(0.668)
Financial service available	0.329	(0.470)	0.377	(0.485)	0.378	(0.485)
Max educ in hh (yrs)	8.310	(3.439)	8.708	(3.644)	9.004	(3.743)
Nighttime light intensity(DMSP F16)	5.448	(11.07)	5.803	(11.47)	6.302	(12.32)
Returns/ind worker (med smth, USD)	2916.5	(1509.9)	1284.5	(991.8)	1342.0	(1531.7)
Returns/ser worker (med smth, USD)	2851.6	(2125.7)	1638.4	(1922.3)	1941.9	(1746.9)
Partpn in ind emplmt (med smth sh)	0.0757	(0.104)	0.0574	(0.0904)	0.124	(0.124)
Partpn in ser emplmt (med smth sh)	0.164	(0.193)	0.193	(0.191)	0.227	(0.228)
Transfers recvd (USD)	99.13	(184.2)	64.65	(124.8)	76.43	(151.5)
Household size	5.207	(2.792)	5.142	(2.909)	4.927	(2.846)
HH dependent share	0.380	(0.250)	0.373	(0.253)	0.359	(0.258)
Years educ, head	6.000	(4.062)	6.133	(4.198)	6.509	(4.290)
Ave length of season (days, MOD12Q2)	177.7	(26.20)	177.5	(25.87)	178.6	(25.84)
Head's father attended school	0.427	(0.495)	0.474	(0.499)	0.501	(0.500)
Urban	0.258	(0.438)	0.283	(0.450)	0.276	(0.447)
Peri-urban dummy	0.101	(0.301)	0.101	(0.302)	0.107	(0.309)
Observations	3151		3804		4834	

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 Selection at means	(SE)	Heckman coefs	
Urban (yes)	-423.42**	77.83	-120.71	82.93	-0.42**	0.01
Peri-urban dummy (yes)	418.95	243.24	370.63	216.78	-0.14**	0.02
Female head (yes)	-150.48**	36.05	-105.73**	39.80		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	158.33	111.03	164.78	100.68		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	7.96	15.78	5.04	14.69		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	27.40	21.97	33.78	21.70		
People per square km, 2005 (ln, HC) (sq rt)	73.42	195.47	-20.20	178.06		
Number hh members 16-65 (sq rt)	315.04**	74.68	260.83**	70.75		
Age of head (sq rt)	-22.85	36.13	-23.82	33.37		
Yrs educ, adults (ave) (sq rt)	24.62	42.23	38.71	40.78		
Yield Potential low (cross-crop ind) (sq rt)	-661.66	1,217.96	-674.73	1,051.81		
Yield Potential high (cross-crop ind) (sq rt)	-47.28	633.33	-4.65	558.61		
Rate improved maize seed use (mean smth) (sq rt)	-1.38	173.48	-83.18	156.64		
Cost hired lbr (med smth, USD/day) (sq rt)	-208.59*	98.83	-130.42	89.83		
Rate of tractor use (mean smth) (sq rt)	189.54	232.98	131.74	223.68		
Land owned (ha, RIGA) (sq rt)	259.09**	55.43	309.24**	51.31		
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	7.38	19.74	12.11	19.03		
Slope (pct, USGS) (sq rt)	110.95	57.35	145.83**	54.00		
Share land irrigated (percent, FAO) (sq rt)	926.86**	333.98	912.84**	325.00		
Soil nutrient constraints (FAO) (moderate)	-12.57	45.85	-47.14	51.92		
Soil nutrient constraints (FAO) (severe)	54.57	82.77	18.33	97.96		
Soil nutrient constraints (FAO) (other)	3,220.18*	1,323.71	3,346.57*	1,303.41		
Soil workability constraints (FAO) (moderate)	-49.50	44.63	-55.94	54.71		
Soil workability constraints (FAO) (severe)	-76.74	80.09	-123.17	101.46		
Soil workability constraints (FAO) (other)	-2,608.30*	1,280.23	-2,763.79*	1,270.86		
Transfers recvd (USD)					-0.00**	0.00
Household size					0.04**	0.00
HH dependent share					0.13**	0.02
Years educ, head					-0.02**	0.00
Ave length of season (days, MOD12Q2)					-0.00**	0.00
Head's father attended school (yes)					-0.06**	0.01
R2 (adj)	0.240		0.235			
Lambda			-289.98			
Sigma			1141.12			
P value comparison test			0.00			
N (non-censored)	8217		8217			
N (censored)					3572	
N	11789		11789		11789	

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 Selection at means	(SE)	Heckman coefs	
Urban (yes)	-278.90	457.46	-516.56	401.91	0.18**	0.01
Peri-urban dummy (yes)	-873.06	1,047.93	-1,228.55	1,077.52	0.10**	0.02
Female head (yes)	-840.10**	201.94	-648.81**	195.08		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-581.58	473.48	-533.77	474.51		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	26.07	65.44	15.54	63.79		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	144.30	110.07	130.10	101.65		
People per square km, 2005 (ln, HC) (sq rt)	521.31	977.67	347.21	940.33		
Number hh members 16-65 (sq rt)	354.59	406.46	341.45	415.13		
Age of head (sq rt)	-321.94	172.20	-245.33	170.14		
Yrs educ, adults (ave) (sq rt)	408.39	226.36	329.29	218.60		
Cost/hired worker (med smth, USD) (sq rt)	22.33	19.06	20.53	19.50		
Nighttime light ave coverage(DMSP F16) (sq rt)	26.46	22.11	30.08	22.48		
Productive non-ag asset ind, fact 1 (sq rt)	3,755.46**	515.77	3,071.75**	525.81		
Productive ag-related asset ind, fact 1 (sq rt)	-4,310.76**	1,210.73	-4,106.91**	1,209.10		
Financial service available (yes)	274.92	157.63	277.18	167.19		
Transfers recvd (USD)					-0.00**	0.00
Household size					0.02**	0.00
HH dependent share					-0.15**	0.02
Years educ, head					0.00*	0.00
Ave length of season (days, MOD12Q2)					-0.00**	0.00
Head's father attended school (yes)					0.08**	0.01
R2 (adj)	0.194		0.194			
Lambda			-960.66			
Sigma			5345.79			
P value comparison test			0.00			
N (non-censored)	5462		5462			
N (censored)					6327	
N	11789		11789		11789	

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 Selection at means	(SE)	Heckman coefs	
Urban (yes)	-194.39	472.86	-1,207.66*	487.09	0.26**	0.01
Peri-urban dummy (yes)	-617.18	1,031.72	-1,760.56	1,067.32	0.11**	0.02
Female head (yes)	-858.73**	212.78	-715.91*	311.39		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-1,189.79*	480.98	-1,049.66*	494.00		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	3.40	69.45	-28.38	71.00		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	-77.29	129.67	-55.55	123.28		
People per square km, 2005 (ln, HC) (sq rt)	1,392.69	1,034.03	803.78	1,046.37		
Number hh members 16-65 (sq rt)	1,410.60**	477.14	378.85	560.96		
Age of head (sq rt)	387.65	203.57	509.65*	226.59		
Yrs educ, adults (ave) (sq rt)	2,609.90**	390.66	1,651.94**	418.96		
Max educ in hh (yrs) (sq rt)	672.40	494.06	198.73	524.35		
Nighttime light intensity(DMSP F16) (sq rt)	-231.12	206.27	-367.30	189.33		
Returns/ind worker (med smth, USD) (sq rt)	-4.16	13.04	11.04	14.56		
Returns/ser worker (med smth, USD) (sq rt)	61.79**	11.95	50.88**	12.90		
Partpn in ind emplmt (med smth sh) (sq rt)	177.32	826.27	-398.99	976.37		
Partpn in ser emplmt (med smth sh) (sq rt)	-192.75	988.80	574.60	1,124.43		
Transfers recvd (USD)					-0.00	0.00
Household size					0.02**	0.00
HH dependent share					-0.36**	0.02
Years educ, head					0.02**	0.00
Ave length of season (days, MOD12Q2)					0.00**	0.00
Head's father attended school (yes)					0.05**	0.01
R2 (adj)	0.307		0.298			
Lambda			-1402.40			
Sigma			4430.79			
P value comparison test			0.00			
N (non-censored)	3644		3644			
N (censored)					8145	
N	11789		11789		11789	