Preliminary and incomplete draft. This version: 29/02/2012

SUPPLY RESPONSE, MARKET PARTICIPATION AND TRANSACTION COSTS IN FOOD MARKETS: EVIDENCE FROM A TANZANIAN PANEL

Antonio Martuscelli[1]

Abstract

This paper analyses the role barriers to trade in the form of high transaction costs have on market participation and supply response in rural food markets in developing countries. We present an empirical application on a sample of households in the Tanzanian region of Kagera. Exploiting the availability of a long term panel dataset to develop an error components switching regression model we can take into account individual unobserved heterogeneity. The results show the importance of households' heterogeneous market participation for the estimation of supply response and confirm the role of transaction costs in orienting households toward a lower degree of specialization.

JEL classification: O13, Q12, Q18
Keywords: Food markets; Supply response; Transaction costs

[1] University of Sussex (UK). Contact: a.martuscelli@sussex.ac.uk

1. Introduction

We analyze the role of transaction costs in the food market of Tanzania's Kagera region using a longitudinal household survey that permits taking into account households' unobserved heterogeneity. The results show that high transaction costs increase households' reliance on domestic production at the expenses of market exchange and thus hinder specialization.

A switching regression model which endogenizes the three possible households' choices of being a food buyer, seller or self-sufficient shows that measures of transaction costs are both economically and statistically significant in explaining an increase in food domestic production for buyers and a reduction for sellers implying a reduction in the role of market exchange and a lower degree of specialization that would otherwise arise in absence of transaction costs.

High transaction costs have been considered an important feature characterizing the functioning of rural markets in developing countries. A general definition of transaction costs identifies them as all those costs that agents have to incur in order to conclude a market transaction beyond the price of the good or service which are the immediate object of the transaction.

The literature on the argument (Goetz 1992, Heltberg and Tarp 2002, Key et al. 2000) has usually distinguished between two different types of transaction costs: proportional transaction costs (*PTC*) and fixed transaction costs (*FTC*). The former vary in function of the quantity traded and the latter instead are a lumpy sum representing the one off cost of entering a market irrespective of the quantity traded.

PTC include the costs of transferring the good traded such as its transport and the time spent in between to reach the market. *FTC* include the costs of: a) searching for a market or a trading partner; b) negotiating and bargaining; c) screening, enforcing and supervising contracts.

The main challenge we face when analyzing the impact of transaction costs is that they are usually not observed and measured directly in most surveys. What we often observe are some factors which are thought to affect them. For example, *FTCs* such as searching, screening and so on are very difficult to quantify directly but we can instead observe other variables which indicate agents' degree of information and network relationship available which are expected to determine their *FTCs*.

The effect that transaction costs have on farmers is twofold: on the one side they directly affect the effective price received by sellers and paid by buyers and thus the quantities exchanged. On the other side, transaction costs also generates heterogeneity in the way households relate to markets and can potentially explain why some farmers decide to take part to market exchange while others remain self-sufficient. In particular fixed transaction costs can be seen as barrier to trade which need to be overcame if the farmer has to take part to the market.

The implication for the estimation of agricultural supply response is that both the heterogeneity in market participation and the role of non-price factors have to be taken into account to obtain correct estimates of supply response. In particular, accounting for the heterogeneity in market participation implies the estimation of separate supply functions for each market regime as they are likely to behave differently and to respond to different incentives.

In the following sections after a review of the literature on supply response and market participation we review the theoretical framework which incorporates transaction costs, heterogeneity in market participation and supply response. Then we account for the problem we have to face when using a panel data for this analysis and our strategy to address it. We then report the results of the empirical analysis and finally draw some conclusions.

2. Literature review

Several studies have addressed under a wide spectrum of angles and methodologies the role that transport costs, infrastructures, isolation and transaction costs play in shaping economic decisions and the process of economic development. In recent years a number of studies have dedicated an increasing attention to the role of rural infrastructures in developing countries identifying several reasons through which rural infrastructures and transaction costs can influence agricultural production, income and poverty in developing countries' rural areas.

Transportation and transaction costs can affect directly agricultural productivity and output. Productivity can be affected through input adoption as the price of imported inputs rises the higher are transportation costs but also through increased price volatility or differing specialization patterns and crop mix. Stifel and Minten (2008) present evidence that transportation costs reduce rice yields and input use in Madagascar. Dorosh et al. (2010) also show a significant effect of road infrastructure on agricultural output and input adoption using a more aggregated cross-sectional spatial approach for Sub-Saharan Africa.

One factor explaining the inverse relationship between productivity and transportation costs runs through the effect they have on the level of specialization and choice of the crop mix. The evidence on the link between transaction costs and specialization is mixed. Qin and Zhang (2011) directly link the Herfindal specialization index to road access in a Chinese rural province and find a higher degree of specialization among better connected households. Stifel et al (2003) found for a sample of households in Madagascar a lower level of concentration of agricultural production in more isolated areas and a shift towards staple food production at the expenses of more valuable crops. On the other hand, Gibson and Rozelle (2003) find that increased isolation reduces the number of income generation activities pursued by households and thus increases specialization. Omamo (1998) uses simulation techniques to show that households facing higher transaction costs tend to alter the crop mix and increase the share of food crops at the expenses of cash crops.

A somewhat different but interrelated aspect implied by high transaction costs concerns the impact they have on the degree of commercialization and on farmers' market participation decisions. The literature on market participation and transaction costs has shown that high transaction costs can drive households out of the market as an optimal strategy to avoid high fixed and proportional transaction costs (De Janvry et al. 1991).

The main implication of this finding is that transaction costs will generate heterogeneity in how households relate to the market as some will optimally choice not to take part to market transactions. This in turn might imply heterogeneous behavior that needs to be taken into account in empirical applications and requires an appropriate econometric strategy.

Goetz (1992) in his pioneer study of Senegalese grain market analyses the marketing behavior of buyers and sellers respectively. He models the discrete market participation decision of buyers and sellers and then estimates separate market surplus equations accounting for selectivity into the corresponding regime. The evidence is not totally conclusive but it does show that information is a significant driver of market participation decisions. The drawback of this study is that it looks only at how fixed transaction costs influence households' decision to enter the market while not allowing for the role of transaction costs in influencing the quantity transacted.

With a similar approach, Heltberg and Tarp (2002) in their study of supply response in Mozambique model selling farmers' marketing behavior. They found evidence that ownership of a means of transport and proximity to a railway increases both the likelihood of entering the market as a seller and the quantity sold. The break-down of the marginal effect into the entry/exit and quantity components shows that the first effect is substantially larger. Implying that promotion of market access can solicit a greater volume of additional supply from peasants entering the market for the first time than for existing participants. The study suffers from two shortcomings. First, they use cross-sectional data and are not able to identify any price effect which limits their ability to compare the effectiveness of price versus non-price factors in increasing market participation and sales. Second, the proxy used for fixed transaction costs as population density and information dummy are not statistically significant casting some doubts on the identification of the model.

Key, Sadoulet and de Janvry (2000) follow a different approach in their study of Mexican farmers' participation in the maize market. They estimate a structural model of market participation and supply decisions taking into account the distinct role of proportional and fixed transaction costs. They jointly estimate the supply functions and the production thresholds using a censoring model with unobserved censoring thresholds. Here the focus is not on marketing behavior but on production behavior given the heterogeneity in market participation. They found that both proportional and fixed transaction costs do matter for market entry and output decisions.

Bellemare and Barrett (2006) look at the pastoralists' participation in livestock markets in Ethiopia and Kenya and estimate the determinants of marketing surplus for the different regimes of net-buyers, net-sellers and self-sufficient. Ouma et al. (2010) analyze smallholders' participation in banana markets in Central Africa adopting an approach very similar to the one presented by Goetz (1992) showing that farmers located one hour further from the nearest urban market reduce the transacted quantities by 17% for sellers and 12% for buyers. Alene et al. (2008) present a study of maize supply and fertilizer demand in Kenya and find that farmers located far from the market reduce transacted quantities by 62%.

Some of the above studies look at the market surplus as opposed to supply. Their aim in that case is to analyze marketing behavior estimating the determinants of the quantities bought or/and sold while controlling for the endogenous selection into the respective market participation regime. Other studies look instead at the production side only and estimate the determinants of quantity produced controlling for regime selection.

Other studies have focused more on the role of transaction costs in the labour and land market. Carter and Yao (2002) estimates regime specific equations for the households' labor intensity taking into account their participation regime in the land rental market. The transaction cost in their study is a measure of legal limitations which encumber transactions in the land market in China. They use an ordered probit as selection equation and estimate both selection and outcome equations using simulated maximum likelihood. They use a panel dataset of Chinese households and make use of a correction to control for households' fixed effects. Sadoulet, de Janvry and Benjamin (1998) model households labor intensity distinguishing across regimes of participation in the labor market for a sample of Mexican households. They apply a two-step procedure à la Heckman where however the first step selection equation is an ordered probit.

The above studies while all pertaining to the same stream of the literature on market participation and transaction costs differ in some important aspects. The econometric strategy varies. Some studies make use of the ordered structure of the regime choice while others do not. As we will show in the theoretical model in the next section an ordered structure for the market participation decision is in principle preferable as it exploits the ordered structure of the relationship but it also complicates the estimation.

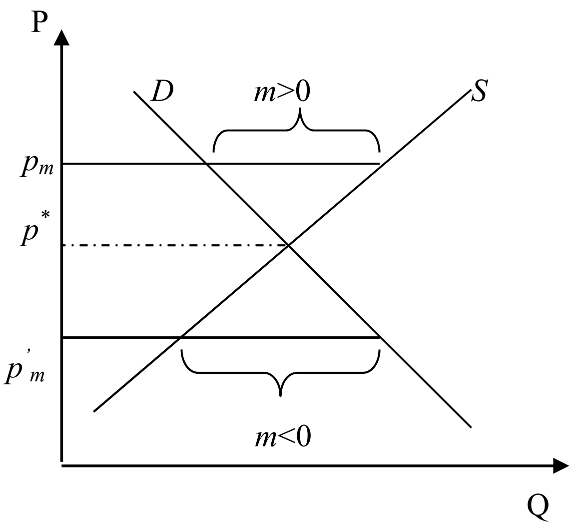
The type of data used is generally cross-sections, only few studies make use of longitudinal data in this literature and when they do, apart from Carter and Yao (2002), they treat them as cross-sections even if the econometric literature has established the potentially serious biases this procedure involves when individual unobserved heterogeneity is present. The availability of longitudinal data potentially also opens up the possibility to explore the dynamic pattern of market participation. The main aim of this paper is to address these shortcomings and use the longitudinal data at hand to incorporate unobserved heterogeneity into the analysis.

3. Theoretical framework

Before introducing a more formal treatment of the implication of transaction costs for market participation and the estimation of the supply function it will be useful to look at the intuition behind just comparing the scenarios in presence and absence of transaction costs.

Consider a farm household which both produces and consumes food. The household will be characterized by a demand and a supply function (*figure 1*). In absence of transaction costs the household is a net-seller if the market price (p_m) is higher than the shadow price (p^*) defined as the price that would equate demand and supply of the household. In a similar way the household would be a net-buyer if the shadow price is higher than the market price (p'_m). The passage from being a net-buyer to being a net seller is a continuous one and only in the limited case when the shadow price equals exactly the market price non-participation to the market would be a utility maximizing choice.

Figure 1

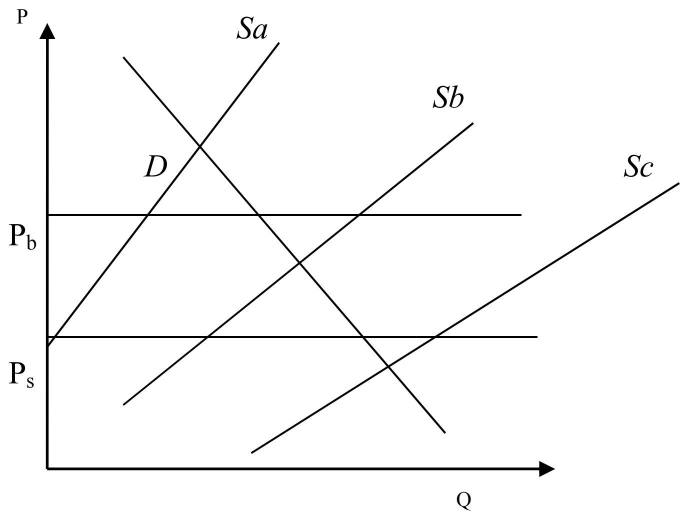


Thus, in absence of transaction costs households face a single price for buying and selling and all will take part to the market either as buyers or sellers. The set of autarchic households should tend towards emptiness.

The introduction of proportional transaction costs, as said, would increase the effective price paid by buyers while reducing the effective price received by sellers and thus generates a price band between the two prices. As in the case where there are no transaction costs, the choice of the regime of market participation is based on comparing the utility obtained in the different regimes for a particular commodity. The difference is that the presence of a price band is likely to generate a set of households for which utility is maximized by being self-sufficient and not taking part to the market exchange. A third regime, autarchy or self-sufficiency, is now possible and market participation becomes a discrete choice for effect of the price band generated by transaction costs.

The utility obtained in different regimes and thus the decision on whether to take part to the market depends on a comparison between the shadow price of those goods with the buying and selling market prices. If the shadow price is higher than the buying price then the household will maximize its utility level by being a net-buyer. If the household shadow price is below the market selling price then the household will be better off as a net-seller. Finally, if the shadow price follows within the band non-participation is the utility maximizing strategy for the household. *Figure 2* represents three different hypothetical households differing only in term of the supply (*Sa, Sb* and *Sc*). The different point where supply crosses demand determines household's market position.

Figure 2



The role of transaction costs can be analyzed in a formal model which endogenises market participation decisions. The model presented below follows closely the one proposed by Key et al (2000), one of the most complete exposition of the role of transaction costs in rural food markets. Consider a utility maximizing household which consumes a set of goods c_i and produces q_i agricultural products. To simplify notation let's define m_i as the marketed surplus for good i and express the cash constraint in term of marketed surplus taking into account transaction costs. Market prices are corrected for proportional transaction costs which add to the market price in case of goods purchased ($m_i < 0$) and reduce the effective sale price ($m_i > 0$). Transaction costs (τ) are differentiated between proportional (τ_{pi}) and fixed (τ_{fi}) transaction costs and between buyers (τ^b) and sellers (τ^s). They are not directly observed but some factors affecting them (z_t) are observed. z^c and z^p are respectively households' consumption and production shifters. The household's problem is to choose quantities of consumption goods in order to maximize utility subject to a set of constraints:

$$\max_c U(c_i; z^c)$$

s.t

$$g(q_i; z^p) = 0$$

$$q_i - m_i - c_i = 0$$

$$\sum_{i=1}^{N} \left[(p^m - \tau^s_{pi}(z^s_t))\delta^s_i + (p^m + \tau^b_{pi}(z^b_t))\delta^b_i \right] m_i - \tau^s_{fi}(z^s_t)\delta^s_i - \tau^b_{fi}(z^b_t)\delta^b_i + E = 0$$

The first constraint is a standard well-behaved production function. The second is a resource constraint imposing that quantity consumed are equal to quantity produced deducted (added) sales (purchases). The third constraint is a cash constraint imposing that expenditures need to be equal to revenues from the sale of farm products plus the exogenous income E taking into account proportional and fixed transaction costs (δ^j_i takes value of 1 if the household is class j and zero otherwise). The lagrangean for this problem is:

$$\begin{aligned} L &= U(c_i; z^c) \\ &+ \lambda\left[\sum_{i=1}^{N} \left[(p^m - \tau^s_{pi}(z^s_t))\delta^s_i + (p^m + \tau^b_{pi}(z^b_t))\delta^b_i \right] m_i - \tau^s_{fi}(z^s_t)\delta^s_i - \tau^b_{fi}(z^b_t)\delta^b_i + E\right] \\ &+ \phi g(q_i; z^p) + \eta(q_i - m_i - c_i) \end{aligned}$$

Given the presence of fixed transaction costs which generates discontinuities in the Lagrangean function Key et al. (2000) show how the solution can be decomposed into two steps, first finding the optimal solution conditional on the market participation regime and then choosing the market participation regime which maximizes utility.

The first order conditions for consumption goods, outputs and traded goods assuming interior solutions are respectively:

$$\frac{\delta U}{\delta c_i} - \eta = 0$$

$$\phi \frac{\delta g}{\delta q_i} + \eta = 0$$

$$\lambda \left[(p^m - \tau^s_{pi}(z^s_t))\delta^s_i + (p^m + \tau^b_{pi}(z^b_t))\delta^b_i \right] - \eta = 0$$

We can define the decision price as:

$$p_i = \begin{cases} p_i^m - \tau^s_{pi} & \text{if } m_i > 0, \text{ seller} \\ p_i^m + \tau^b_{pi} & \text{if } m_i < 0, \text{ buyer} \\ \mathrm{p_i^*} = \dfrac{\eta}{\lambda} & \text{if } m_i = 0, \text{ self-sufficient} \end{cases}$$

The problem is now to choose the utility maximizing regime by comparing utilities under different regimes. Using the above defined decision price we can define the maximum utility attained in each regime using the same indirect utility function $V(p_i, y, z^c)$. Define $y_0(p_i)$ as the household income before incurring any fixed transaction costs:

$$y_0(p_i) = \sum_{i=1}^{N} p_i q_i + E$$

Then the utility levels for different regimes can be written as:

$$V^s = V(p^m - \tau^s_p, y_0(p^m - \tau^s_p) - \tau^s_f, z^c) \quad \text{if seller}$$
$$V^b = V(p^m + \tau^b_p, y_0(p^m + \tau^b_p) - \tau^b_f, z^c) \quad \text{if buyer}$$
$$V^a = V(p^*, y_0(p^*), z^c) \quad \text{if autarkic}$$

These expressions show that in absence of fixed transaction costs the household would be indifferent between selling and being self-sufficient if $p^* = p^m - \tau^s_p$. From the FOC it can be shown that utility is increasing in the decision price for sellers and decreasing in the decision price for buyers. Thus, if $p^m - \tau^s_p > p^*$ a household facing no fixed transaction costs would be better off selling. In a similar way, the household will be indifferent between buying and being self-sufficient if $p^* = p^m + \tau^b_p$ and it would be better-off buying on the market if $p^m + \tau^b_p < p^*$.

Figure 3

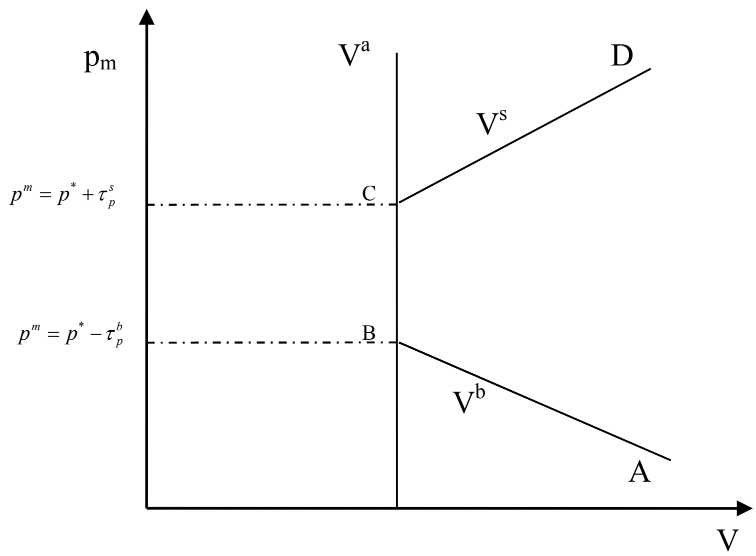
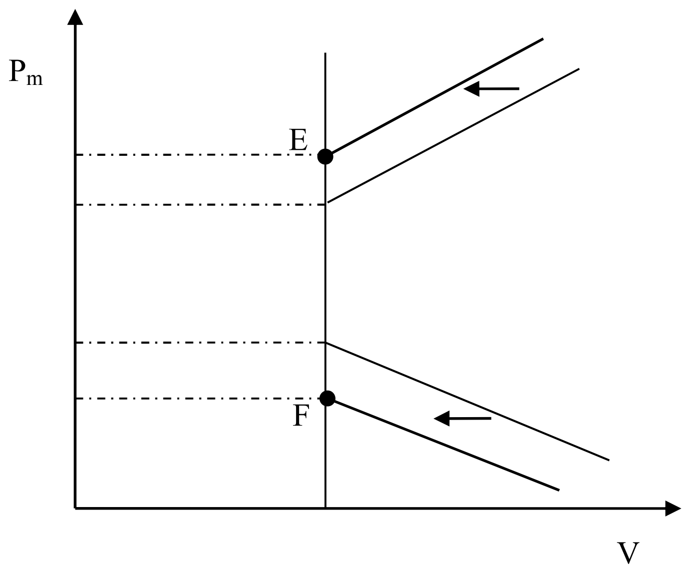


Figure 3 shows the indirect utility function *V* as a function of the market price. The vertical line shows the utility attainable by autarkic households which is independent of the market price. At point C the household will be indifferent between selling and being autarkic and for prices above C it will be a seller with utility increasing with price. In point B the household will be indifferent between buying and being autarkic and for prices below B it will be better-off buying with utility decreasing with the price. If the market price is between B and C the household will be better-off staying in autarky. The optimal market participation strategy is ABCD.

The implication of the introduction of fixed transaction costs as well as the proportional ones can be shown looking at the fact that fixed transaction costs lower household income and thus utility for each level of price. This will shift the utility curves to the left as shown in *figure 4*. Fixed transaction costs will thus refrain households to enter the market until the price is sufficiently high (low) to cover fixed transaction costs for sellers (buyers), points E and F in *figure 4*.

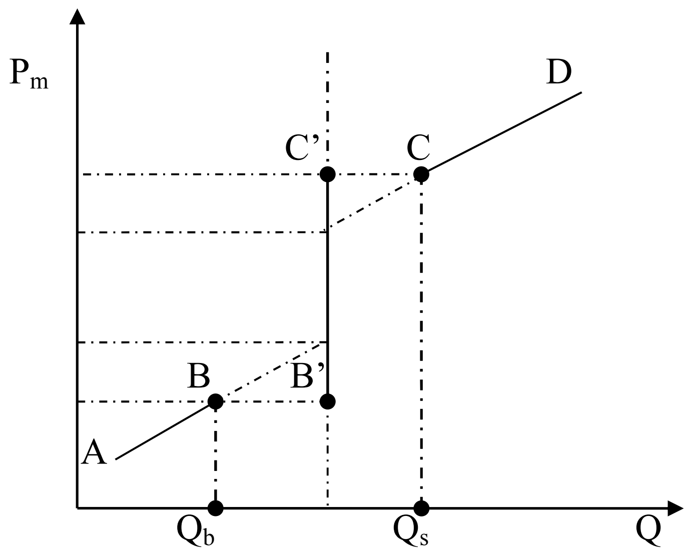
Figure 4



However, fixed transaction costs do not enter household's supply curve because only the marginal return to production affects production decisions. The first implication of this finding is that once entering the market either as a buyer or seller there will be a discrete jump in the quantity produced as the decision price will change discretely. The supply function derived has three distinct regimes for different price levels and is shown in *figure 5*. The vertical part of the supply function corresponds to the autarkic regime which is unresponsive to prices. The lower segment AB is the buying region and finally the upper segment CD corresponds to the selling region. The discrete change in the quantities produced when entering the market either as buyer or seller can also be seen in *figure 5* at the points Q_s and Q_b. These two quantities are the quantity thresholds below which it is not optimal for the household to enter the market.

The second important implication of this distinct role of fixed transaction costs is that it gives a way of econometrically identifying the parameters of both the market participation and volume decisions.

Figure 5



The above analysis has several implications for the estimation of supply response. First, we need to take into account the unresponsiveness of autarchic households. Unless this is accounted for it is likely that pooled estimates of the true underlying supply response will be downward biased. Second, changes in prices and other non-price factors which affect market participation decisions should be accounted for in the estimation of the overall supply response. Third, transaction costs have a different impact on buyers and sellers and estimation of regime specific supply functions can help identify and test the effective importance of transaction costs on productive behavior. For a seller an increase in transaction costs is expected to have a negative effect on production. For a buyer instead an increase in transaction costs is expected to have a positive impact on production.

In the following sections, after discussing the econometrics challenges we face when using longitudinal data in this contest, we will try to estimate empirically this model and analyze in a coherent framework market participation and supply decisions.

4. Supply response and market participation: An error components switching regression model

The analysis on how transaction costs affect the market participation decision and how this decision in turn affects supply response implies the need for models able to consider simultaneously the discrete choice on market participation and the continuous one regarding output. The literature on discrete choice models and selection models is very extensive and several different techniques have been developed[2]. The extension of these techniques to longitudinal panel data is however not completely straightforward and developments in this area for panel data have been slower. .

We adopt a Full Information Maximum Likelihood approach instead of the two step procedure. The model was first proposed by Zabel (1992) in the context of a binary probit selection rule and as an alternative to other two step procedure developed since then. As shown above, in our case an ordered selection rule would be more appropriate. The main drawback of Zabel approach was its computational intensity but as noted in Greene (2006) the development of simulation methods have to some extent made it easier to estimate such models. Carter and Yao (2002) present one application of the same family with an ordered probit selection rule.

We sort individuals (or households) into J+1 classes according to an ordered probit selection rule:

$$d_{it}^{*} = \gamma^{'} Z_{it} + r_i + u_{it}$$

$$d_{it} = \begin{cases} 0 \text{ if } -\infty < \mathrm{d}_{it}^{*} \leq \mu_0 \\ 1 \text{ if } \mu_0 \leq \mathrm{d}_{it}^{*} \geq \mu_1 \\ 2 \text{ if } \mu_1 \leq \mathrm{d}_{it}^{*} \geq \mu_2 \\ .... \\ J \text{ if } \mu_{j-1} < \mathrm{d}_{it}^{*} \geq +\infty \end{cases}$$

where d_{it}^{*} is a latent unobserved variable assumed to be dependent on a vector of explanatory variables Z , a vector of unknown parameters γ, an individual specific heterogeneity component r_i assumed to be random and an idiosyncratic random term u_{it} . What we actually observe is the discrete variable d_{it} which takes values from 0 to J for each of the J+1 classes.

For each of the J+1 classes we observe an outcome variable:

[2] See Greene (2008) for a review.

$$Y_{itj} = \begin{cases} \beta_0' X_{it0} + v_{i0} + \varepsilon_{it0} & \text{if } d_{it} = 0 \\ \beta_1' X_{it1} + v_{i1} + \varepsilon_{it1} & \text{if } d_{it} = 1 \\ \beta_2' X_{it2} + v_{i2} + \varepsilon_{it2} & \text{if } d_{it} = 2 \\ \ldots\ldots \\ \beta_j' X_{itj} + v_{ij} + \varepsilon_{itj} & \text{if } d_{it} = J \end{cases}$$

Y_{itj} is the outcome variable, at time *t* for household *i* belonging to class *j* and is a linear function of a vector of explanatory variables X_{itj}, an individual specific time invariant random effect v_{ij} and an idiosyncratic error term ε_{itj}.

Both the outcome and selection equations allow for the presence of individual random effects. There are J+1 random effects in the outcome equations and an additional one in the ordered probit equation for a total of J+2 random effects. The assumptions needed for the random components and the idiosyncratic error are:

$$v_{ij} \sim N(0, \sigma_{vj}^2)$$
$$\varepsilon_{itj} \sim N(0, \sigma_{\varepsilon j}^2)$$

for the outcome equations and

$$r_i \sim N(0, \sigma_r^2)$$
$$u_{it} \sim N(0,1)$$

for the ordered probit selection equation. The variance of *u* is normalized to unit for identification purposes.

To simplify the estimation we assume the individual-specific error terms are uncorrelated with each other, an assumption which could in principle be relaxed. Each idiosyncratic error terms in the outcome equations ε_{jtj} and the one in the selection equation u_{it} have instead correlation coefficient ρ_j. Thus, ε_{itj} and u_{it} follow a bivariate normal distribution:

$$(\varepsilon_{itj}, u_{it}) \sim N_2(0, 0; \sigma_{\varepsilon j}^2, 1; \rho_j \sigma_{\varepsilon j}^2)$$

The likelihood for individual *i* has the following form:

$$L_i = \prod_t f(y_{itj}) \times \Pr(d_{it} = j \mid y_{itj})$$

$$= \iint \prod_t \frac{1}{\sigma_{\varepsilon j}} \phi\left(t_{itj}\right) \left[\Phi\left(\frac{\gamma Z + r_i + \rho_j t_{itj} - \mu_{j-1}}{\sqrt{1-\rho_j^2}} \right) - \Phi\left(\frac{\gamma Z + r_i + \rho_j t_{itj} - \mu_j}{\sqrt{1-\rho_j^2}} \right) \right] \phi\left(v_{ij}\right) \phi\left(r_i\right) dv_{ij} dr_i$$

Where $t_{itj} = \frac{y_{itj} - \beta_j^{'} X_{itj} - v_{ij}}{\sigma_{\varepsilon j}}$, ϕ is the standard normal density function and Φ is the standard normal cumulative function. The derivation uses the fact that given the joint distribution of $(\varepsilon_{itj}, u_{it}) \sim N_2(0,0;\sigma_{\varepsilon j}^2, 1; \rho_j \sigma_{\varepsilon j}^2)$ the conditional distribution of u given ε_j is:

$$(u_{it} \mid \varepsilon_{itj}) \sim N(\frac{\rho_j \varepsilon_j}{\sigma_{\varepsilon j}}; 1-\rho_j^2) = N(\rho_j t_{itj}; 1-\rho_j^2)$$

Then the log-likelihood is $\ln L = \sum_i \ln L_i$.

The likelihood requires a double integration for any observation and is computationally intensive. However, recent developments in simulation methods provide a way to evaluate the above likelihood function. Simulation techniques are methods to numerically approximate otherwise intractable integrals. There are various forms of simulation used for different kind of problems. In our case it is enough to notice that the above integrals are expectations over the random individual effects:

$$E_r \left[E_v \left[\phi(.)[\Phi(.) - \Phi(.)] / \sigma \right] \right]$$

and these expectations can be approximated with the average of a sufficient number of draws from the standard normal distribution generating r_i and v_{ij}.

The above likelihood function can then be approximated with the following simulated likelihood function:

$$L_i^s = \frac{1}{G} \sum_{g=1}^{G} \prod_t \frac{1}{\sigma_{\varepsilon j}} \phi\left(t_{itj,g}\right) \left[\Phi\left(\frac{\gamma Z + r_{i,g} + \rho_j t_{itj,g} - \mu_{j-1}}{\sqrt{1-\rho_j^2}} \right) - \Phi\left(\frac{\gamma Z + r_{i,g} + \rho_j t_{itj,g} - \mu_j}{\sqrt{1-\rho_j^2}} \right) \right]$$

where $t_{itj,g} = \frac{y_{itj} - \beta_j^{'} X_{itj} - v_{ij,g}}{\sigma_{\varepsilon j}}$. $r_{i,g}$ and $v_{ij,g}$ are random draws from the respective normal distributions and G is the total number of draws taken from the distributions of v and u.

To generate the random draws we follow Train (2009)[3] and use draws derived from Halton sequences which improve accuracy with a reduced number of draws with respect to pseudo-random number generators. This is due to the fact that draws derived from Halton sequences increase the coverage of the domain of integration and induce a negative correlation between the draws thus reducing the variance of the simulation.

The simulated likelihood is then maximized numerically with respect to the parameters $\beta_j, \gamma, \sigma_{\varepsilon j}, \sigma_{vj}, \sigma_r, \rho_j, \mu_1..\mu_{j-1}$. This method is called Maximum Simulated Likelihood (MSL)[4].

In our case the classes of market participation are three, in order: buyers, autarchic and sellers.

(1)
$$d_{it}^{*} = \gamma^{'} Z_{it} + r_i + u_{it}$$
$$d_{it} = \begin{cases} 0 \text{ if } -\infty < d_{it}^{*} \leq \mu_0 \text{ if buyer} \\ 1 \text{ if } \mu_0 \leq d_{it}^{*} \geq \mu_1 \text{ if autarchic} \\ 2 \text{ if } \mu_1 < d_{it}^{*} \geq +\infty \text{ if seller} \end{cases}$$

For each regime of market participation we observe food supply:

(2)
$$\begin{cases} Y_{it}^{B} = \beta_B^{'} X_{itB} + v_{iB} + \varepsilon_{itB} & \text{if } d_{it} = 0 \\ Y_{it}^{A} = \beta_A^{'} X_{itA} + v_{iA} + \varepsilon_{itA} & \text{if } d_{it} = 1 \\ Y_{it}^{S} = \beta_S^{'} X_{itS} + v_{iS} + \varepsilon_{itS} & \text{if } d_{it} = 2 \end{cases}$$

The likelihood of a household being respectively a buyer, autarchic and seller at time t is then given by the above formulas and the full log-likelihood is obtained taking the within-group product over time of the single observation likelihoods, integrating this products over the random effects, taking logs and summing over groups we obtain the full log-likelihood:

$$\ln L = \sum_i \ln \iint \prod_t \left\{L_{it}^{B}\right\}^{b} \left\{L_{it}^{A}\right\}^{a} \left\{L_{it}^{S}\right\}^{s} \phi(v_{iA}) \phi(r_i) dv_{iA} dr_i$$

[3] See Cappellari and Jenkins (2006) for a brief review of the different methods available to generate random draws for MSL and their relative advantages. Also, they develop the Stata routine *mdraws* used for this application to generate Halton sequences.

[4] See Train (2009) and Gourieroux and Monfort (1993).

Where the exponents *b*, *a* and *s* take value of 1 if the household *i* at time t is respectively a buyer, autarchic or seller and zero otherwise.

The simulated likelihood is derived as follow:

$$\ln {}^{s}L = \sum_{i} \ln \frac{1}{G} \sum_{g} \prod_{t} \left\{ {}^{s}L_{it}^{B} \right\}^{b} \left\{ {}^{s}L_{it}^{A} \right\}^{a} \left\{ {}^{s}L_{it}^{S} \right\}^{s}$$

In our estimation we use 50 random draws to approximate the likelihood function. Gourieroux and Monfort (1993) show that with this method a moderate number of replications is enough to obtain a good approximation of the likelihood function.

5. The data

We use a long term panel of households living in Tanzania's region of Kagera for our empirical estimation. It is a remote region in the northern part of Tanzania bordering Uganda in the north, Rwanda and Burundi in the west and the Lake Victoria on the Eastern part. It is a rural region where agriculture represents the main source of income. Coffee produced for export and food crops, in particular banana the main staple in the region but also maize, cassava and beans are the prevalent products produced in the region.

The dataset used for the empirical analysis is the Kagera Health and Development Survey (KHDS), a panel of households in the Kagera region. The sample is composed of 733 households interviewed during four waves on a six month interval between 1991 and 1993 and then re-interviewed during a fifth wave in 2004.

The agricultural system is characterized by smallholder producers that tend to produce both coffee and food crops, often intercropped. In the sample the average amount of land cultivated is of four acres with a small decrease in 2004 from 1991. A high percentage of the households, around 65%, produce both food and coffee. Kagera is one of the main coffee producer regions of Tanzania, where coffee is one of the country's main exports. In terms of inputs the data shows that around one quarter of farmers hires labor while the percentage applying fertilizers or pesticide is quite low; less than five and ten percent respectively.

Attrition during the five waves was quite low especially considering the gap of ten years between the fourth and fifth wave. In fact, 93% of the baseline households has been re-contacted and re-interviewed in 2004.

This dataset represents one of the few examples of long term longitudinal data in developing countries. This is a potential advantage of the data in what it permits looking at long term changes in households behavior and also permits to analyze the role of factors which, being quasi-fixed, are usually washed out in standard panel analysis due to lack of time variation. At the same time this characteristic of the data presents several challenges as the own concept of household becomes blurry in a ten year long period. In fact, tracking each individual in the original sample of households interviewed in the first round of the survey gives rise to a much higher number of households after ten years. We use a balanced version of the panel looking only at the households which have been interviewed in all the five waves of the survey for a total of 733 households[5].

The survey collected a number of important pieces of information on demographic characteristics, consumption, farming and non-farming activities among other. It also complements the households' specific information with a community survey which collects information on the infrastructure and public services for each of the 51 communities covered by the survey and a price survey which collects data on local market prices of food and non-food products. This information permits us to obtain a wide range of variables of interest.

Two variables in particular are of special importance for our analysis: food output and prices. Concerning the output measure, the survey collects information on the value of output which has to be deflated by an appropriate price to obtain a measure of output. We use the cluster average of the households' producer price to deflate the output value and obtain a measure of output.

As the food price will also be one of the main covariates of interest the deflation might create an econometric problem in presence of measurement error. To avoid this problem we exploit the presence of a price survey collecting local market prices for food. Thus, we use food market prices as regressor.

We construct our measure of food output from the aggregation of four food crops which represent both the main food crops produced in the region and the main staple food consumed: bananas, maize, beans and cassava. The food price index is calculated as the average of the market prices of the four crops.

All the other variables of interest are easily obtained from the survey. The community questionnaire collected information on cluster wages for agricultural workers[6]. The total value of assets is the value of equipment, buildings, land, durables and livestock and the net value of financial assets reported by household members in the survey questionnaire. The total land area is the sum of all *shambas* owned or cultivated by the household and is expressed in acres. The education variable is the number of year of

[5] To link households in 2004 with the original household we follow in order the head of the household, the spouse of the head, the oldest son or daughter.

[6] In wave one to four the wage for a day of work (length of which is not specified) is recorded while in the fifth wave the hourly wage is recorded. To make them consistent we assume that a standard day of work is of eight hours and transform the hourly wage into the daily counterpart for wave five.

education of the household head. Rainfall is the total amount of rainfall as recorded in the closest weather station in the growing season. Distance from a motorable road is a community variable which expresses the community distance from a motorable road in kilometres[7].

The dataset offers several interesting insights into the evolution of agriculture in the region and the different behaviour of households pertaining to different food market regimes. Table 1 shows some descriptive statistics on households' production structure for 1991 and 2004 decomposed by regime of market participation.

The first interesting thing to notice is the evolution of the share of agricultural production on total expenditures which drops quite dramatically during the period across all categories. Net-buyers are the group where agriculture accounts for the lower share of expenditures. These figures show the increasing diversification of rural households into different income generating activities also documented in other studies on the recent evolution of agriculture in Sub-Saharan Africa. Staple food maintains a high and stable share of total production.

Yields have actually decreased during the period. While this might be due to the life-cycle evolution of our sample or to contingent weather-related factors it is an indication of a stagnant agricultural sector. Notably yields are on average lower for net-buyers in both years. An indirect confirmation of the stagnation of agriculture comes from input use figures. The proportion of households applying fertilizers and pesticides, if low in both periods, has actually dropped even more. Again, net-buyers show a lower inputs adoption.

[7] This information is collected only in the first and last wave and we extend the first wave distance to the other three waves of the first round of the survey. Thus, time variation in road distance is based on difference from the 91-94 value and the 2004 one.

Table 1: Descriptive statistics

	1991							
	Net-buyers		Autarchy		Net-seller		Overall	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Yield (000')	90.0	141.7	93.5	135.6	120.3	275.9	102.9	207.5
Conspc (000'tsh)	189.1	125.3	150.1	94.3	195.9	139.8	187.2	129.0
				Production shares				
Food	0.71	0.19	0.67	0.2	0.7	0.17	0.7	0.18
Coffee	0.04	0.07	0.09	0.14	0.07	0.1	0.06	0.09
Vegetables	0.03	0.06	0.02	0.04	0.03	0.04	0.03	0.05
Other cash crops	0	0.03	0.01	0.06	0	0.03	0	0.03
Fruits	0.07	0.1	0.07	0.08	0.06	0.08	0.07	0.09
Other food	0.09	0.12	0.08	0.09	0.06	0.1	0.08	0.11
Other crops	0.06	0.09	0.07	0.09	0.07	0.08	0.07	0.09
				Expenditure shares				
Agricultural production	0.54	0.34	0.73	0.53	0.77	0.59	0.66	0.49
Sales	0.05	0.09	0.09	0.15	0.14	0.13	0.09	0.12
Purchases	0.57	0.2	0.43	0.18	0.49	0.17	0.52	0.19
				Inputs				
Land size (acre)	3.64	2.76	4.48	2.95	4.95	2.97	4.29	2.94
Hired labor (% hiring)	19	40	14	35	37	48	26	44
Fertilizer (% applying)	5	21	4	21	7	26	6	23
Pesticide (% applying)	8	27	9	28	19	39	13	33
Manure (% applying)	45	50	42	50	45	50	45	50
N	329		91		295		715	

	2004							
	Net-buyer		Autarchy		Net seller		Overall	
	Mean	sd	Mean	sd	mean	sd	Mean	Sd
Yield (000')	54.0	76.5	69.4	118.8	83.9	103.6	69.3	96.4
Conspc (000'tsh)	181.9	131.6	188.5	129.7	215.5	187.9	197.7	159.4
				Production shares				
Food	0.72	0.17	0.73	0.18	0.72	0.16	0.72	0.17
Coffee	0.03	0.06	0.03	0.07	0.05	0.08	0.04	0.07
Vegetables	0.02	0.06	0.02	0.05	0.02	0.04	0.02	0.05
Other cash crops	0	0.03	0	0	0	0.04	0	0.03
Fruits	0.05	0.06	0.03	0.04	0.05	0.07	0.05	0.06
Other food	0.13	0.12	0.13	0.14	0.1	0.11	0.11	0.12
Other crops	0.05	0.07	0.06	0.1	0.06	0.08	0.06	0.08
				Expenditure shares				
Agricultural production	0.4	0.2	0.49	0.24	0.61	0.41	0.51	0.33
Sales	0.05	0.08	0.04	0.09	0.15	0.14	0.09	0.12
Purchases	0.65	0.18	0.55	0.19	0.57	0.16	0.6	0.18
				Inputs				
Land size (acre)	3.38	2.63	3.81	3	4.56	2.92	3.96	2.86
Hired labor (% hiring)	20	40	24	43	41	49	31	46
Fertilizer (% applying)	2	15	3	17	3	17	3	16
Pesticide (% applying)	6	24	7	26	7	25	7	25
Manure (% applying)	14	35	13	33	27	45	20	40
N	262		87		274		623	

6. Empirical analysis

We estimate the switching regression model for the food market participation and output decisions (eq. 1 and 2).

Households can position themselves into three classes with an ordered structure: buyers (B), autarchic (A) and sellers (S). We define autarkic or self-sufficient households as the ones that do not buy or sell any quantity of food in a given year.

We have a problem in classifying households' market position as a high percentage of households buy and sell food at the same time[8]. This is mainly an artifact of the aggregation of four food crops into a single food aggregate. In fact, several households have different marketing relationships for different crops. We decide to use the net market position to classify households as food net-buyers, net-sellers or self-sufficient[9]. An alternative option would be to exclude from the estimation households that buy and sell at the same time as the model of market participation based on transaction costs cannot properly explain this behavior. We use this as a robustness check in section 7 below.

Explanatory variables in the supply functions are, according to the theoretical model, supposed to capture both households' supply and demand factors (Z), which determine the relative position of the supply and demand curves, exogenous income (E), prices (P) and proxies for proportional transaction costs (T^p). In the market participation equation the explanatory variables are all the variables included in the supply equations plus fixed transaction costs (T^f).

We run the model as a random effects model with district (ϑ_d) and time (μ_t) fixed effects in both supply and participation relationship. The equations to estimate are:

$$d_{it} = f(\ln P_{it}, Z_{it}, E_{it}, T_{it}^p, T_{it}^f, \vartheta_d, \mu_t)$$

$$\ln Q_{jit} = f(\ln P_{it}, Z_{it}, E_{it}, T_{it}^p, \vartheta_d, \mu_t) \qquad j=\{B, A, S\}$$

The set of prices used as explanatory variables includes coffee and food prices, agricultural wages and kerosene price. All these prices are expressed in real terms by deflation with the Laspeyre price index at the regional level calculated from the consumer price survey collected in conjunction with the main survey[10]. We use the

---

[8] Out of the 3,487 observations that compose the sample 1,089 (31%) are of households that buy and sell food in a given season. The remaining sample is composed of 2,398 observations of which 944 (39%) are buyers, 524 autarkic (21%) and 930 (38%) sellers.

[9] Using this classification we have 1552 (44%) Net-buyers, 533 (15%) Self-sufficient and 1402 (40%) Net-sellers observations.

[10] See Beegle, De Weerdt and Dercon (2006).

cluster average of producer prices for coffee while for food and kerosene we use market prices collected in the price survey.

Table 2: Summary statistics main variables

	Net-buyers		Self-sufficient		Net-sellers	
	Avg	Sd	Avg	Sd	Avg	Sd
Food output (kg)	164.5	165.8	189.9	187.5	242.4	377.8
Coffee price (91' Tzsh)	46.1	16.7	47.1	18.9	49.7	19.3
Food price (91' Tzsh)	54.7	12.7	53.8	11.6	54.7	12.5
Kerosene price (91' Tzsh)	89.7	30.3	94.1	33.4	98.5	31.7
Wage (91' Tzsh)	158.9	82.3	174.1	82	166.4	77.8
Age head	49.4	17.1	54.5	16.9	50	16.7
Female head	0.29	0.46	0.3	0.46	0.27	0.44
Totarea (Acres)	3.87	2.98	4.77	3.94	5.37	3.9
Rainfall (100mm)	10.3	4.7	10.3	4.2	9.9	4.9
Assets (mln Tzsh)	0.94	1.11	1.25	1.46	1.59	2.15
HH Size	6.01	3.13	5.78	2.86	5.75	2.93
Market	0.69	0.46	0.67	0.47	0.7	0.46
Road distance (Km)	0.19	0.99	0.15	0.85	0.16	0.9
Road impassable	0.42	0.49	0.5	0.5	0.42	0.49
Education (years)	4.19	3.5	3.68	3.36	4.32	3.25
Transport ownership	0.28	0.45	0.31	0.46	0.36	0.48
Density (000' per Km^2)	0.68	1.7	0.19	0.37	0.20	0.39
Main ethnic group	0.6	0.49	0.66	0.47	0.65	0.48
Info	0.31	0.46	0.3	0.46	0.33	0.47
N	1552		533		1402	

We control for the educational status, the age and gender of the head and for land and capital endowment of farmers. Endowment of capital is measured by the logarithm of the value of the household's assets. Land endowment is measured as logarithm of the total amount of land cultivated in acres.

We use five district dummies corresponding to the six administrative sub-regions of Kagera and a urban dummy for communities in Bukoba town to control for the unobserved fertility and other characteristics of land and the availability of off-farm work. We control for weather conditions using the total amount of rainfall in the community in the last season.

The variables used to capture transaction costs are distance to a motorable road, a dichotomous variable indicating whether the road is impassable in certain periods of the year or not, ownership of a means of transport (bicycle, motorbike, and car), presence of a market in the community, population density, a dummy for ethnic minority and an information dummy for ownership of a radio, telephone or TV. The last three should capture the availability of information and network possibilities and thus affect fixed transaction costs rather than proportional ones. These variables have been proposed as measures of fixed transaction costs from different authors in studies of supply response and transaction costs (Heltberg and Tarp 2002 use population density and ownership of information means as measures of fixed transaction costs; Goetz 1992 uses also the ethnicity dummy for networking opportunities). They act as identifiers in the selection model as they affect only participation decisions and not output ones. Population density data at the district level is available only for 2002 and to use it we need to make the assumption that district density is relatively stable across time[11]. Table 2 presents summary statistics of the explanatory variables for each regime of market participation.

Table 3 illustrates the results of the estimation. Results of the market participation equation (column 4 in Table 3) show that the main factors affecting the decision are cultivated land and assets value which both increase the probability of being a seller. Family size instead increases the probability of being buyer. None of the price variables is statistically significant. Among the proportional transaction cost variables only the dummy for impassable road is marginally significant and as expected tends to reduce the probability of being a seller.

Population density and the dummy for the ethnic group, the proxy used for fixed transaction costs, are instead highly significant. Higher population density increases the probability of being a buyer while belonging to the main ethnic group increases the likelihood of being a seller. The third variable reflecting fixed transaction costs is instead not statistically significant.

[11] Population density data at the ward (administrative district) level was obtained from the International Livestock Research Institute and National Census Bureau (geo-information section) of Tanzania. Source of data used for development was census maps of population and housing census 2002.

Table 3: Results

<table>
<tr><th></th><th>(1)</th><th>(2)</th><th>(3)</th><th>(4)</th></tr>
<tr><th>VARIABLES</th><th>Net-buyers</th><th>Net-seller</th><th>Autarkic</th><th>Oprobit</th></tr>
<tr><td>Coffee price (log)</td><td colspan="2">0.162***</td><td>0.263**</td><td>0.127</td></tr>
<tr><td></td><td colspan="2">(0.045)</td><td>(0.103)</td><td>(0.085)</td></tr>
<tr><td>Food price (log)</td><td colspan="2">0.309***</td><td></td><td>-0.088</td></tr>
<tr><td></td><td colspan="2">(0.058)</td><td></td><td>(0.121)</td></tr>
<tr><td>Kerosene price (log)</td><td colspan="2">0.001</td><td>0.080</td><td>0.104</td></tr>
<tr><td></td><td colspan="2">(0.052)</td><td>(0.111)</td><td>(0.103)</td></tr>
<tr><td>Wage (log)</td><td colspan="2">0.219***</td><td>0.262***</td><td>0.006</td></tr>
<tr><td></td><td colspan="2">(0.029)</td><td>(0.072)</td><td>(0.056)</td></tr>
<tr><td>Age head</td><td colspan="2">0.014***</td><td>0.0134</td><td>0.018**</td></tr>
<tr><td></td><td colspan="2">(0.005)</td><td>(0.010)</td><td>(0.009)</td></tr>
<tr><td>Age squared</td><td colspan="2">-0.000**</td><td>-0.000</td><td>-0.000**</td></tr>
<tr><td></td><td colspan="2">(0.000)</td><td>(0.000)</td><td>(0.000)</td></tr>
<tr><td>Female</td><td colspan="2">0.044</td><td>0.067</td><td>0.089</td></tr>
<tr><td></td><td colspan="2">(0.037)</td><td>(0.082)</td><td>(0.069)</td></tr>
<tr><td>Land area (log)</td><td colspan="2">0.177***</td><td>0.062</td><td>0.265***</td></tr>
<tr><td></td><td colspan="2">(0.030)</td><td>(0.073)</td><td>(0.053)</td></tr>
<tr><td>Rainfall (00’ mm)</td><td colspan="2">-0.005</td><td>-0.002</td><td>-0.009</td></tr>
<tr><td></td><td colspan="2">(0.003)</td><td>(0.009)</td><td>(0.006)</td></tr>
<tr><td>Assets (log)</td><td colspan="2">0.053***</td><td>0.116**</td><td>0.167***</td></tr>
<tr><td></td><td colspan="2">(0.020)</td><td>(0.050)</td><td>(0.037)</td></tr>
<tr><td>Size</td><td colspan="2">0.086***</td><td>0.091***</td><td>-0.076***</td></tr>
<tr><td></td><td colspan="2">(0.006)</td><td>(0.014)</td><td>(0.010)</td></tr>
<tr><td>Education (year)</td><td colspan="2">0.024**</td><td>0.044**</td><td>0.024</td></tr>
<tr><td></td><td colspan="2">(0.010)</td><td>(0.022)</td><td>(0.019)</td></tr>
<tr><td>Education square</td><td colspan="2">-0.001</td><td>-0.001</td><td>-0.001</td></tr>
<tr><td></td><td colspan="2">(0.001)</td><td>(0.002)</td><td>(0.001)</td></tr>
<tr><td>Market</td><td>-0.125**</td><td>0.007</td><td></td><td>0.010</td></tr>
<tr><td></td><td>(0.050)</td><td>(0.042)</td><td></td><td>(0.059)</td></tr>
<tr><td>Road distance (km)</td><td>0.013</td><td>-0.019</td><td></td><td>-0.038</td></tr>
<tr><td></td><td>(0.023)</td><td>(0.020)</td><td></td><td>(0.028)</td></tr>
<tr><td>Road impassable</td><td>0.020</td><td>-0.077*</td><td></td><td>-0.098</td></tr>
<tr><td></td><td>(0.048)</td><td>(0.039)</td><td></td><td>(0.061)</td></tr>
<tr><td>Transpown</td><td>0.021</td><td>0.135***</td><td></td><td>0.042</td></tr>
<tr><td></td><td>(0.051)</td><td>(0.040)</td><td></td><td>(0.063)</td></tr>
<tr><td>Density (000’/km^2)</td><td></td><td></td><td></td><td>-0.285***</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.042)</td></tr>
<tr><td>Methnic</td><td></td><td></td><td></td><td>0.322***</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.101)</td></tr>
<tr><td>Info</td><td></td><td></td><td></td><td>-0.029</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.061)</td></tr>
</table>

Note: Sample 3,487 observations. Log-likelihood=-6830. Dependent variable in column one to three is log of food output. In column four the dependent variable is the ordered probit index (Net-buyer=0, Self-sufficient=1, Net-seller=2). Covariates not reported in the table but included in the estimation are five district dummies for the six districts in the region, two seasonal dummies for the three growing/harvesting seasons, time dummies and an urban community dummy. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The first three columns of table 3 show the results for the food supply equations respectively for the buying and selling regimes and for the autarchy regime. The buying and selling equations differ only for the proportional transaction costs variables which according to the theoretical model have a differential impact in the two regimes.

Two main results are of interest. The first concerns the own price elasticity which although small at 0.3 is highly significant. Comparing this result with the pooled one obtained neglecting the heterogeneity in market participation[12] it becomes clear that not taking into account the price unresponsiveness of self-sufficient households introduces a downward bias in the estimation of the price elasticity.

The second result concerns the role of proportional transaction costs. As highlighted above we expect transaction costs to have an opposite impact on buyers and sellers. High transaction costs should give buyers an incentive to produce more in order to reduce their reliance on expensive market goods. On the contrary, sellers facing high transaction costs have an incentive to reduce production as the price they receive for their produce will be lower. The model reproduces this theoretical result quite well. The transaction cost variables in the buying and selling equations have the expected opposite signs with the exclusion of the transport ownership in the buying equation. For buyers only the presence of a market in the community has a significant impact on food production. For sellers transport ownership and impassable road are significant. The magnitudes of the coefficients are quite significant in economic terms. The presence of a regular market reduces buyers' food production by around 13%. Ownership of a means of transport increases sellers' output by 14% while living in areas where the road becomes impassable in certain periods of the year reduces sellers' output by around 8%.

These results add some evidence to the relevance transaction costs have on rural households' behavior. In particular, transaction costs increase net-buyers reliance on "home" food production while reducing net-sellers' capacity to supply the market. High transaction costs provide thus an incentive towards the adoption of a self-sufficiency strategy and thus reduce specialization and the productivity gains associated with it. This distinct effect of transaction costs cannot be analyzed without taking into account the heterogeneous relation households have with the market.

A further prediction of the model is that higher transaction costs would increase the probability of households being in the self-sufficiency region. This is the "discrete" effect on households' market participation as opposed to the "continuous" one on output decisions highlighted before. This prediction is not supported by the data that show proportional transaction costs variables not being an important determinant of market participation choices.

---

[12] Appendix 1 contains the results of this pooled random effect estimation. The own price elasticity has negative sign and is not statistically different from zero.

An interesting feature of this switching model is that we can compute different marginal effects of interest according to the research question. In particular as shown by Huang et al. (1991) and McDonald and Moffitt (1980), we can compute the unconditional marginal effect for the full sample for the total supply regardless of the regime for which it is observed as:

$$\sum_{j=0}^{J} M_j = \frac{\delta E(y_j \mid d=j)}{\delta x} \times \Pr(d=j) + \frac{\delta \Pr(d=j)}{\delta x} \times E(y_j \mid d=j)$$
$$= \beta_j \left[\Phi\left(\gamma Z + r - \mu_{j-1}\right) - \Phi\left(\gamma Z + r - \mu_j\right) \right] +$$
$$+\gamma \left[\phi\left(\gamma Z + r - \mu_{j-1}\right) - \phi\left(\gamma Z + r - \mu_j\right) \right] \left[X_j \beta_j + v_j + (\gamma Z + r)\rho_j \right]$$

Where, as above, the subscript j identifies the regime of market participation. The above expression provides the marginal effect of variable x on total quantity supplied. The first term represents the quantity response weighted by the probability of being in regime j and the second term the marginal change in the probability of being in regime j weighted by the expected value of the quantity if in regime j. the total effect can thus be decomposed into a quantity response component and a regime switching component. Table 4 shows the unconditional marginal effects computed in this way.

The total unconditional own price elasticity of food is 0.2 taking into account the unresponsiveness of self-sufficient households and the effect of regime switching. The main effects on output come from land and asset endowments, wages and coffee price which operate through both the selection and the quantity side.

Table 4: Decomposition of unconditional marginal effects

	Quantity	Selection	Total
Coffee price (log)	0.179***	0.082	0.261***
	(0.051)	(0.071)	(0.068)
Food price (log)	0.256**	-0.057	0.199
	(0.103)	(0.07)	(0.139)
Kerosene price (log)	0.015	0.067	0.082
	(0.053)	(0.081)	(0.079)
Wage (log)	0.226***	0.004	0.230***
	(0.031)	(0.037)	(0.041)
Land size (log)	0.157***	0.171***	0.328***
	(0.048)	(0.073)	(0.110)
Assets (log)	0.064***	0.108***	0.172***
	(0.026)	(0.044)	(0.037)
Size	0.087***	-0.049***	0.038*
	(0.006)	(0.019)	(0.021)
Road distance (km)	-0.002	-0.025	-0.027
	(0.015)	(0.020)	(0.026)
Age	0.013***	0.011	0.025***
	(0.004)	(0.008)	(0.009)
Education	0.026***	0.014	0.039***
	(0.010)	(0.012)	(0.013)
Female	0.074	0.069	0.143**
	(0.054)	(0.059)	(0.074)
Market	-0.122**	0.008	-0.114*
	(0.053)	(0.047)	(0.062)
Road Impassable	-0.010	-0.077	-0.087
	(0.058)	(0.063)	(0.092)
Transport own	0.074	0.032	0.107
	(0.073)	(0.051)	(0.090)
Info		-0.021	-0.021
		(0.045)	(0.045)
Main eithnic group		0.230**	0.230**
		(0.118)	(0.118)
Density		-0.2**	-0.2**
		(0.1)	(0.1)

Note: The first column shows the unconditional marginal effect of a marginal change in the covariate on the log food output coming from adjustment in the quantity weighted by the probability of being in regime j. The second column shows the unconditional marginal effect of a marginal change in the covariate on log food output coming from regime switching and weighted by the expected output if in regime j. The third column is the sum of the quantity and selection marginal effects. Standard errors in parentheses are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

7. Robustness checks

We perform three different estimations to control if the above baseline results are robust to different specifications. The first concern is about the role of other market imperfections that could bias our estimates of the price elasticity and the transaction costs coefficients. The second concern is that using the households' net market position to distinguish among net-buyers, net-sellers and self-sufficient could also impact on the estimates. Finally a third issue relates to the different recall period used in the three intermediate waves of the survey.

The first factor that could influence our results relates to the possible influence of missing markets for insurance and credit and a failure to control adequately for land characteristics. We thus introduce some variables to control for covariate risk factors, access to credit and land quality. To control for covariate risk factors we use the five year before the survey rainfall variation coefficient, the previous year rainfall deviation from the fifteen-year median rainfall and a drought prone dummy for communities that experienced a drought in the ten years before the survey. To control for credit availability we use a dummy for the presence of a bank, money-lender or credit cooperative in the community. Finally we attempt to better control for land characteristics using the average food yield in the community expressed as kilograms per acre and a variable indicating the roughness of the terrain defined as the difference between the highest and lowest altitude. Table 5 shows the results for the estimation including these additional variables. The main results derived in the baseline estimation are all unchanged by the inclusion of these additional controls.

The second robustness check consists in excluding from the estimation households that buy and sell food at the same time to see if results are biased by the use of the net market position to characterize households. Table 6 presents the results. There are no major changes from the baseline estimation if not that transaction costs' coefficients are slightly higher and more precisely estimated as one would have expected.

Finally the third estimation checks if the use we have made of the longitudinal survey affects the estimation. If fact, the second, third and fourth wave of the survey have a six month recall period instead of a full year. We control for this recall difference in the baseline estimation with time and seasonal dummies but there could still be a bias in particular if we misclassify households' market position. An alternative way to use these data that has been proposed is to merge the second and third wave to form a comparable full year wave while dropping the fourth wave. We perform the baseline estimation using these three full-year waves only. Table 7 presents the results of the estimation. The main results are robust to this alternative use of the data and they actually suggest that the baseline transaction costs coefficients could be underestimated.

Table 5: Controlling for risk, credit access and land quality

VARIABLES	(1) Net-buyers	(2) Net-sellers	(3) Autarkic	(4) Oprobit
Coffee price (log)	0.112**		0.166*	0.099
	(0.045)		(0.101)	(0.086)
Food price (log)	0.273***			-0.049
	(0.059)			(0.124)
Kerosene price (log)	0.032		0.228**	0.056
	(0.053)		(0.114)	(0.109)
Wage (log)	0.163***		0.183**	0.003
	(0.029)		(0.074)	(0.058)
Age head	0.014***		0.016*	0.0166*
	(0.005)		(0.009)	(0.009)
Age squared	-0.000**		-0.000	-0.000*
	(0.000)		(0.000)	(0.000)
Female	0.044		0.108	0.093
	(0.036)		(0.082)	(0.069)
Land area (log)	0.183***		0.095	0.288***
	(0.030)		(0.071)	(0.053)
Rainfall (00' mm)	0.016**		0.005	-0.045***
	(0.008)		(0.016)	(0.014)
Assets (log)	0.059***		0.098**	0.165***
	(0.020)		(0.048)	(0.037)
Size	0.084***		0.087***	-0.077***
	(0.006)		(0.014)	(0.010)
Education (year)	0.021**		0.041*	0.025
	(0.010)		(0.022)	(0.019)
Education square	-0.001		-0.001	-0.001
	(0.001)		(0.002)	(0.001)
Rainfall CV	0.035	0.314***	-0.603**	0.442**
	(0.171)	(0.111)	(0.244)	(0.181)
Rainfall deviation (100mm)	-0.274***	-0.212***	0.039	0.387***
	(0.091)	(0.080)	(0.172)	(0.134)
Drought	-0.009	0.022	-0.033	-0.088*
	(0.045)	(0.037)	(0.069)	(0.053)
Credit	-0.114***	0.015	-0.120*	-0.086*
	(0.044)	(0.037)	(0.066)	(0.051)
Terrain roughness	-0.002	0.006	-0.009	-0.004
	(0.006)	(0.006)	(0.008)	(0.009)
Food yields (kg/acre)	0.019***	0.007***	0.027***	0.005*
	(0.003)	(0.002)	(0.005)	(0.003)
Market	-0.087*	-0.003		0.002
	(0.049)	(0.042)		(0.059)
Road distance (km)	0.010	-0.011		-0.048*
	(0.022)	(0.020)		(0.028)
Road impassable	0.035	-0.072*		-0.083
	(0.048)	(0.041)		(0.063)
Transpown	0.026	0.136***		0.033
	(0.049)	(0.039)		(0.063)
Density (000'/km^2)				-0.264***
				(0.042)
Methnic				0.360***
				(0.102)
Info				-0.026
				(0.062)
Observations	3,487			
Ll	-6734			

Note: Dependent variable in column one to three is log of food output. In column four the dependent variable is the ordered probit index (Net-buyer=0, Self-sufficient=1, Net-seller=2). Covariates not reported in the table but included in the estimation are five district dummies for the six districts in the region, two seasonal dummies for the three growing/harvesting seasons, time dummies and a urban community dummy.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Excluding buying and selling households

<table>
<tr><td></td><td>(1)</td><td>(2)</td><td>(3)</td><td>(4)</td></tr>
<tr><td>VARIABLES</td><td>Buyers</td><td>Sellers</td><td>Autarkic</td><td>Oprobit</td></tr>
<tr><td>Coffee price (log)</td><td colspan="2">0.198***</td><td>0.239**</td><td>0.141</td></tr>
<tr><td></td><td colspan="2">(0.062)</td><td>(0.107)</td><td>(0.109)</td></tr>
<tr><td>Food price (log)</td><td colspan="2">0.315***</td><td></td><td>-0.113</td></tr>
<tr><td></td><td colspan="2">(0.075)</td><td></td><td>(0.150)</td></tr>
<tr><td>Kerosene price (log)</td><td colspan="2">-0.073</td><td>0.064</td><td>0.139</td></tr>
<tr><td></td><td colspan="2">(0.069)</td><td>(0.115)</td><td>(0.129)</td></tr>
<tr><td>Wage (log)</td><td colspan="2">0.222***</td><td>0.300***</td><td>-0.020</td></tr>
<tr><td></td><td colspan="2">(0.038)</td><td>(0.075)</td><td>(0.069)</td></tr>
<tr><td>Age head</td><td colspan="2">0.016***</td><td>0.019**</td><td>0.016</td></tr>
<tr><td></td><td colspan="2">(0.006)</td><td>(0.010)</td><td>(0.011)</td></tr>
<tr><td>Age squared</td><td colspan="2">-0.000*</td><td>-0.000*</td><td>-0.000</td></tr>
<tr><td></td><td colspan="2">(0.000)</td><td>(0.000)</td><td>(0.000)</td></tr>
<tr><td>Female</td><td colspan="2">0.050</td><td>0.074</td><td>0.114</td></tr>
<tr><td></td><td colspan="2">(0.048)</td><td>(0.086)</td><td>(0.090)</td></tr>
<tr><td>Land area (log)</td><td colspan="2">0.175***</td><td>0.033</td><td>0.360***</td></tr>
<tr><td></td><td colspan="2">(0.039)</td><td>(0.074)</td><td>(0.066)</td></tr>
<tr><td>Rainfall (00’ mm)</td><td colspan="2">-0.000*</td><td>0.000</td><td>-0.000</td></tr>
<tr><td></td><td colspan="2">(0.000)</td><td>(0.000)</td><td>(0.000)</td></tr>
<tr><td>Assets (log)</td><td colspan="2">0.054**</td><td>0.110**</td><td>0.144***</td></tr>
<tr><td></td><td colspan="2">(0.025)</td><td>(0.049)</td><td>(0.047)</td></tr>
<tr><td>Size</td><td colspan="2">0.098***</td><td>0.109***</td><td>-0.092***</td></tr>
<tr><td></td><td colspan="2">(0.007)</td><td>(0.014)</td><td>(0.013)</td></tr>
<tr><td>Education (year)</td><td colspan="2">0.020</td><td>0.043*</td><td>0.011</td></tr>
<tr><td></td><td colspan="2">(0.014)</td><td>(0.023)</td><td>(0.024)</td></tr>
<tr><td>Education square</td><td colspan="2">0.000</td><td>-0.001</td><td>-0.001</td></tr>
<tr><td></td><td colspan="2">(0.001)</td><td>(0.002)</td><td>(0.002)</td></tr>
<tr><td>Market</td><td>-0.127*</td><td>0.035</td><td></td><td>-0.085</td></tr>
<tr><td></td><td>(0.071)</td><td>(0.052)</td><td></td><td>(0.074)</td></tr>
<tr><td>Road distance (km)</td><td>0.002</td><td>-0.022</td><td></td><td>-0.027</td></tr>
<tr><td></td><td>(0.034)</td><td>(0.025)</td><td></td><td>(0.036)</td></tr>
<tr><td>Road impassable</td><td>0.054</td><td>-0.065</td><td></td><td>-0.181**</td></tr>
<tr><td></td><td>(0.066)</td><td>(0.048)</td><td></td><td>(0.080)</td></tr>
<tr><td>Transpown</td><td>0.051</td><td>0.120**</td><td></td><td>0.064</td></tr>
<tr><td></td><td>(0.071)</td><td>(0.049)</td><td></td><td>(0.082)</td></tr>
<tr><td>Density (000’/km^2)</td><td></td><td></td><td></td><td>-0.314***</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.055)</td></tr>
<tr><td>Methnic</td><td></td><td></td><td></td><td>0.461***</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.128)</td></tr>
<tr><td>Info</td><td></td><td></td><td></td><td>0.005</td></tr>
<tr><td></td><td></td><td></td><td></td><td>(0.081)</td></tr>
<tr><td>Observations</td><td>2,398</td><td></td><td></td><td></td></tr>
<tr><td>Ll</td><td>-4892</td><td></td><td></td><td></td></tr>
</table>

Note: Dependent variable in column one to three is log of food output. In column four the dependent variable is the ordered probit index (Buyer=0, Self-sufficient=1, Seller=2). Covariates not reported in the table but included in the estimation are five district dummies for the six districts in the region, two seasonal dummies for the three growing/harvesting seasons, time dummies and a urban community dummy.
Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Three waves estimation

VARIABLES	(1) Net-Buyers	(2) Net-Sellers	(3) Autarkic	(4) OProbit
Coffee price (log)	0.204***		0.159	0.137
	(0.057)		(0.148)	(0.109)
Food price (log)	0.277***			-0.006
	(0.071)			(0.164)
Kerosene price (log)	0.021		0.205	0.280**
	(0.062)		(0.154)	(0.141)
Wage (log)	0.205***		0.321***	0.075
	(0.037)		(0.121)	(0.083)
Age head	0.015***		0.005	0.0165*
	(0.005)		(0.014)	(0.009)
Age squared	-0.000**		0.000	-0.000
	(0.000)		(0.000)	(0.000)
Female	0.003		0.196	0.039
	(0.041)		(0.120)	(0.078)
Land area (log)	0.188***		0.017	0.268***
	(0.039)		(0.110)	(0.066)
Rainfall (00' mm)	-0.003		-0.034**	0.005
	(0.004)		(0.016)	(0.008)
Assets (log)	0.084***		0.247***	0.177***
	(0.026)		(0.072)	(0.048)
Size	0.072***		0.077***	-0.072***
	(0.008)		(0.023)	(0.012)
Education (year)	0.031***		0.046	0.030
	(0.012)		(0.030)	(0.021)
Education square	-0.001		-0.001	-0.001
	(0.001)		(0.002)	(0.002)
Market	-0.216***	-0.002		0.014
	(0.056)	(0.052)		(0.077)
Road distance (km)	0.010	-0.010		-0.032
	(0.022)	(0.022)		(0.029)
Road impassable	-0.063	-0.075		-0.177**
	(0.054)	(0.050)		(0.071)
Transpown	-0.050	0.174***		-0.000
	(0.055)	(0.049)		(0.073)
Density (000'/km^2)				-0.309***
				(0.007)
Methnic				0.074
				(0.129)
Info				-0.015
				(0.071)
Observations	2,063			
Ll	-3792			

Note: Dependent variable in column one to three is log of food output. In column four the dependent variable is the ordered probit index (Net-buyer=0, self-sufficient=1, Net-seller=2). Covariates not reported in the table but included in the estimation are five district dummies for the six districts in the region, two seasonal dummies for the three growing/harvesting seasons, time dummies and a urban community dummy.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

8. Conclusions

Starting from the theoretical model proposed by Key et al. (2000) we estimate a model of food supply response which incorporates the effect of transaction costs and households' heterogeneity in market participation using a 91-2004 household panel for Tanzania's Kagera region. We adopt simulated maximum likelihood methods to estimate a random component switching regression with an ordered probit as selection rule.

The results confirm the importance of taking into account the unresponsiveness of self-sufficient producers when estimating supply response in rural contexts characterized by a high degree of self-sufficiency. The estimated price elasticity of 0.3 for households taking part to the market although low shows some degree of responsiveness to price incentives in rural food markets.

The results also provide evidence of the importance of transaction costs in developing countries' rural areas. The asymmetric effect of transaction costs on surplus and deficit households shows that policies able to reduce these costs can promote higher specialization and release unexploited productivity gains.

Appendix 1

Food supply pooled estimation

VARIABLES	Food Output (log)
Coffee price (log)	0.082*
	(0.046)
Food price (log)	-0.024
	(0.065)
Kerosene price (log)	-0.173***
	(0.055)
Wage (log)	0.136***
	(0.029)
Age	0.008*
	(0.005)
Age square	-0.000
	(0.000)
Female	0.068*
	(0.038)
Land (log)	0.217***
	(0.027)
Rainfall(100mm)	-0.007**
	(0.003)
Assets (log)	0.095***
	(0.019)
Size	0.071***
	(0.005)
Education	0.024**
	(0.010)
Education square	-0.001
	(0.001)
Market	-0.111***
	(0.031)
Road distance	-0.011
	(0.015)
Road impassable	0.002
	(0.033)
Transport ownership	0.072**
	(0.033)
Observations	3,487
Number of HH	732
Sigma	0.718
sigma_e	0.655
sigma_u	0.295
r2_w	0.292
r2_b	0.600
r2_o	0.443

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

References

Alene, A.D., Manyong, V.M., Omanya, G., Mignouna, H.D., Bokanga, M., Odhiambo, G., (2008), Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. Food Policy 33(4), 318–328.

Beegle K., De Weerdt J. and Dercon S., (2006) Kagera Health and Development Survey 2004, Basic Information Document, mimeo

Bellemare, M.F., Barrett, C. B., (2006), An ordered Tobit model of market participation: Evidence from Kenya and Ethiopia. Am. J. Agric. Econ. 88(2), 324–337.

Cappellari L. and Jenkins S., (2006), Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation The Stata Journal 6, Number 2, pp. 156–189

Carter M. and Yao Y., (2002), Local versus global separability in agricultural household models: the factor price equalization effect of land transfer rights, *Amer. J. Agr. Econ.* 84(3) (August 2002): 702–715

Chiburis R. and Lokshin M., (2007), Maximum likelihood and two-step estimation of an ordered-probit selection model, The Stata Journal 7, Number 2, pp. 167-182

Contoyannis P., Jones A. and Rice N., (2004) The dynamics of health in the British Household Panel Survey, Journal of Applied Econometrics 19: 473–503

Cragg J. (1971), Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, Econometrica, Vol. 39, No. 5 (Sep., 1971), pp. 829-844

Dorosh P., Wang H., You L. and Schmidt E., (2010), "Crop Production and Road Connectivity in Sub-Saharan Africa A Spatial Analysis" World Bank PRWP 5385

Gibson J. and Rozelle S. (2003) "Poverty and Access to Roads in Papua New Guinea, Economic Development and Cultural Change, Vol. 52, No. 1 (October 2003), pp. 159-185

Goetz, S.J. (1992) "A Selectivity Model of Household Food Marketing Behavior in Sub-Sahara Africa." *Amer. J. Agr. Econ*. 74 (May 1992):444–52.

Gourieroux C. and Monfort A., (1993), Simulation-based inference: Asurvey with special reference to panel data models, Journal of Eonometrics 59(1993) 5-33

Greene, W. and Hensher, David A. (2008), Modeling Ordered Choices: A Primer and Recent Developments. Available at SSRN: http://ssrn.com/abstract=1213093

Greene, W., Fixed and Random Effects in Nonlinear Models, (2001) *Department of Economics Working Paper, Stern School of Business, New York University,* Greene W., (2006) Censored data and truncated distributions, in *Palgrave Handbook of Econometrics*,

Greene W. (2008) Discrete Choice Modeling, in *The Handbook of Econometrics: Vol. 2, Applied Econometrics*, Part 4.2., ed. T. Mills and K. Patterson, Palgrave, London.

Haan P., and Uhlendorff A. (2006), Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood, The Stata Journal (2006) 6, Number 2, pp. 229–245

Heckman, J. (1979) Sample selection bias as a specification error. Econometrica 47, 153–61.

Heltberg R. and Tarp F., 2002, Agricultural supply response and poverty in Mozambique Food Policy 27 (2002) 103–124

Huang C, Raunikar R. and Misra, S. 1991, The application and economic interpretation of selectivity models, American Journal of Agricultural economics, Vol. 73 (2) 496-501

Key N., Sadoulet E., and de Janvry A., 2000, Transactions Costs and Agricultural Household Supply Response, *Amer. J. Agr. Econ*. 82 (May 2000): 245–259

McDonald J. and Moffitt R., 1980, The use of tobit analysis, the Review of economics and Statistics, Vol 62, (2) 318-321

Mundlak, Y. (1978) On the pooling of time series and cross sectional data. Econometrica 56, 69–86.

Nelson F., (1977), Censored Regression Models with Unobserved, Stochastic Censoring Thresholds, Journal of Econometrics 6 (1977) 309-327.

Ouma E., Jagwe J., Obare G., Abele S. (2010), Determinants of smallholder farmers' participation in banana markets in Central Africa: the role of transaction costs, Agricultural Economics (41) 111–122

Qin Y. and Zhang X., (2011) "Road to Specialization in Agricultural Production: Tales of 18 Natural Villages in China" Paper presented at the AAEA 2011 Conference

Sadoulet E., de Janvry A., and Benjamin C., (1998), Household Behavior with Imperfect Labor Markets, Industrial Relations, vol. 37, no. 1 (January 1998).

Stifel D. and Minten B. (2008), "Isolation and agricultural productivity". Agricultural Economics 39 (2008) 1–15.

Train K., (2009) Discrete Choice Methods with Simulation, Cambridge University Press

Vella, F. (1998), Estimating models with sample selection bias: a survey. Journal of Human Resources 33(1), 127–69.

Vella, F. and M. Verbeek (1999), Two-step estimation of panel data models with censored endogenous variables and selection bias. Journal of Econometrics 90, 239–63.

Wooldridge, J. (1995), Selection corrections for panel data models under conditional mean independence assumptions. Journal of Econometrics 68(1), 115–32.

Wooldridge, J. (2005), Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20(1), 39–54.

Wooldridge, J. (2010), Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Zabel, J. (1992) Estimating fixed and random effects models with selectivity. Economics Letters 40, 269–72.