Behavior of subsistence producers in response to technological change- The elasticity of cassava production and home consumption in Benin *

Hiroyuki Takeshima

University of Illinois, Urbana-Champaign

email:htakesh2@uiuc.edu

Selected Poster Prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008

Abstract

The welfare effects of GM (genetic modification)-led productivity growth for cassava producers are partly affected by the characteristics of individual cassava producing households. Those household characteristics include the elasticity of production and home consumption of cassava. Some studies assume the inelastic home consumption when conducting ex-ante welfare effects analysis for subsistence crops. This study modifies the estimation methods used in the past literature to estimate both elasticities using the dataset from Benin. Several assumptions are also tested regarding the heterogeneity of cassava producers. On estimation of elasticities, the paper tests the hypothesis that on-farm sellers are characteristically different from off-farm sellers by employing the double hurdle model. The findings contribute to the literatures analyzing the distributional effects of welfare effects from GM-led productivity growth for cassava, which are gaining importance in the context of the policy impacts on poverty reduction.
Key word: cassava, subsistence, double-hurdle, non-nested test
JEL classifications: Q11, Q12

1 Background story and research questions

1.1 Background story and motivation of study

The recent advancement of biotechnology such as genetic engineering provides better prospects for many African countries to increase the agricultural productivity. Many orphan crops like cassava which has been underinvested for variety improvement research may also benefit from the application of genetic engineering.

The actual return from cassava productivity growth in an African country and how the return will be shared by different population groups, however, depends on many socio-economic factors including the characteristics of cassava producers and how they respond to cassava productivity growth.

Cassava producers can respond in several ways to the opportunity of lower production costs by transgenic cassava. Adopting transgenic cassava, producers can either produce more, probably sell more and consume more or less. Or, a producer may produce, consume and sell the same amount of cassava at lower costs. These two behaviors of producers lead to different effects on the market price and farmgate price of cassava, and thus different effects on welfare of cassava producers and consumers. Which of these are more likely has, however, been less studied by the past literature, although the latter seems to be implicitly assumed in several studies.

Aforementioned characteristics of subsistence cassava producers are partly observed by the elasticities of productions, sales and own consumptions with respect to the farmgate cassava prices.

Few studies estimate the elasticity of home consumption of agricultural commodity produced by subsistence producers with respect to the farmgate price of those commodities. Toquero et al. (1975) find that the rice consumption by subsistence rice producers in the Philippines is rather price inelastic. Some studies assume the inelastic home consumption to conduct the ex-ante welfare analysis of GM-led productivity growth (Qaim, 2001). How the findings from Toquero et al. (1975) is relevant to African countries is, however, unclear since many African households consume several staple crops other than cassava unlike in the Philippines where the rice is the predominant staple crops. Many studies estimate the elasticity of market supply of agricultural commodities by subsistence producers, which can be estimated without the inference on the elasticity of home consumption. As is discussed in section 3.1, however, the home consumption of cassava can be elastic to the farmgate price under certain conditions, which may affect the welfare effects as was discussed in the previous chapter. The Benin dataset used in this study contains the information required to estimate the elasticity of home consumption of cassava.

1.2 Research question

One important factor that influences the elasticities is the high transactions cost in many developing countries. While proportional transactions costs influence the quantity sold, produced and consumed, the fixed transactions costs affect whether to sell cassava or not, and possibly where to sell. These transactions costs oftentimes vary across different producers and are unobserved or only partially observed.

Some complications arise when estimating aforementioned elasticities to incorporate the unobserved transactions costs. First, the estimation needs to be corrected for the sample selection bias since the elasticities of production, sales or consumption with respect to cassava price can be estimated only from the behavior of those who participate in the market. Second, sellers may need to be differentiated based on where they sell cassava. In the case of Benin dataset used in this study, there are two types of sellers regarding where they sell cassava. While those who sell cassava at the farmgate (*on farm sellers* hereafter) report the farmgate price of cassava, those who travel to the market (*off farm sellers* hereafter) report the market price of cassava which includes unobserved proportional transaction costs.

A question of interest is whether off-farm sellers respond to the cassava price differently from on-farm sellers do, or both types of sellers respond in the same way, when deciding on how much to produce, consume and sell given the expected cassava price. The underlying hypothesis in the former case is that if a high fixed transactions costs exist to change the sales location, a cassava producer first makes a decision of sales location for cassava, and allocated the resource to maximize his utility based on the conditions in that particular market. Some cassava producers may prefer to travel to nearby market incurring some transportation costs, if he thinks it too costly to find traders who come at his farmgate and buy cassava from him, and vice versa.

The answer to the question above has an important bearing on how we can estimate the elasticities that are representative of all cassava producers in Benin. If both types of cassava sellers respond to cassava price in the same way, one can obtain the elasticities using the data for on-farm sellers alone, which is beneficial since no estimation is required for unobserved transaction costs for off-farm sellers. If off-farm sellers respond to the cassava price in the different way, a separate estimation is required for off-farm sellers, which also requires the estimation of unobserved transaction costs.

Several approaches can be used to gain insight into the answer to the above questions as discussed in section 4. This study employs the model similar to the "double hurdle" model proposed by Cragg (1971). More explicitly, this study assumes that each type of sellers, independently from cassava price, make decisions on whether to sell cassava, and where to sell cassava before deciding on production, sales and consumption. The crucial assumption of aforementioned independence to cassava price is inspired by the findings from Bellemare and Barrett (2006) which analyze similar issues for livestock sellers and purchasers in Kenya and Ethiopia. The assumption is important in our dataset since only the sellers and buyers report the cassava price, and it is thus difficult to analyze how the cassava price affects the cassava producers decisions on whether to sell cassava or not[1]. The more detailed description of estimation procedure is laid out in section 4.

1.3 Contribution of this paper

The contributions of this chapter can be summarized in the following ways. The findings of the elasticities of home consumption and production of cassava together provide empirical guidance to whether the ex-ante welfare effects analysis using EDM can be more informative and accurate. The test for the structural difference between on-farm cassava sellers and off-

[1]Some studies include the price as an explanatory variable into the first stage selection equation (Goetz, 1992)

farm cassava-sellers provides an insight into how welfare gains from GM cassava will be shared among cassava producers.

2 Literature reviews for the estimation of supply response of subsistence crops

2.1 Literature on subsistence producer's market participation decisions

Many studies estimate the supply response of subsistence crops in the market. Several estimation methodologies are employed to incorporate the economic issues inherent to the decision-making on the market supply of crops by subsistence farmers (de Janvry and Sadoulet, 2005). Strauss (1984) lays the groundwork for the estimation of production and consumption decision-makings for subsistence agricultural households. Goetz (1992) focuses on the mechanism as to how subsistence farmers in Africa make market participation decisions given the transactions costs each farmer must incur, and how that mechanism may affect the estimation of market sales response. Goetz (1992) employs the Heckman's sample correction methods to consistently estimate the price elasticity of sales for rice in Senegal. Several studies follow Goetz (1992) by employing various methods to incorporate the unobserved transactions costs in estimating the supply response of subsistence farmers. Conventional Heckman's sample correction methods are widely used (Heltberg and Tarp, 2002).

Bellemare and Barrett (2006) extends conventional Heckman's sample selection methods to test whether a livestock traders make market participation decision and sales or purchase quantity decision sequentially. The findings by Bellemare and Barrett (2006) are in favor of sequential decision-making, and suggest that some factors such as price are not considered when livestock traders decide whether to sell or not.

2.2 Approaches in the literature to incorporate unobserved transactions costs into elasticity estimations

One challenge in estimating the aforementioned elasticities is how to factor the unobserved fixed and proportional transactions costs into the estimation model.

One advantage of Bellemare and Barrett (2006) is the explicitly reported fixed costs and variable costs of market participation for each livestock trader. The dataset in this paper, however, contains only the transportation costs for some off farm sellers, which do not distinguish between fixed and variable transactions. Several studies propose an alternative approach to incorporate transactions costs when no direct measures of transactions costs are reported, or reported transactions costs presumably fail to capture the real transactions costs.

The unobserved fixed transactions costs are often controlled by Heckman's sample selections model as employed in many studies Goetz (1992); Heltberg and Tarp (2002). Renkow, Hallstrom, and Karanja (2004) estimates fixed transactions costs as functions of explanatory variables using a model different from Heckman's model. The findings by Renkow, Hallstrom, and Karanja (2004) provides this study with insight into which explanatory variables should be used to explain the market participation decisions by cassava producers.

Alternative methods to analyze the unobserved FTC are proposed several studies. Key, Sadoulet, and de Janvry (2000) relaxes the restriction of the common threshold for sample selection assumed in Heckman's approach and allows thresholds to vary across each household. Holloway, Barrett, and Ehui (2005) employ Bayesian econometrics model to obtain robust estimates of the structural equation for milk sales in Ethiopia, as well as a minimum sales quantity threshold that traders decide to enter the market. While the aforementioned approaches provide room to improve the elasticity estimates, this study, however, adheres to the conventional Heckman's sample selection approach due to the following reasons. The stochastic threshold approach by Key, Sadoulet, and de Janvry (2000) requires that price is reported for all observations, and inapplicable to our dataset which reports prices only for sellers and buyers. The Bayesian approach by Holloway, Barrett, and Ehui (2005) may be complicated when applied to the estimation with dual selection criteria which this focuses on.

The Key, Sadoulet, and de Janvry (2000) and Vakis, Sadoulet, and de Janvry (2003) employ models in which the unobserved proportional transactions costs are approximated as linear functions of a set of explanatory variables including reported PTC. While Key, Sadoulet, and de Janvry (2000) simply adds those PTC-related explanatory variables to the structural equation, Vakis, Sadoulet, and de Janvry (2003) regresses reported PTC on other PTC-related explanatory variables to obtain predicted PTC and adds the predicted PTC to the structural equation. Recently Henning and Henningsen (2007) followed Key, Sadoulet, and de Janvry (2000), although referring Vakis, Sadoulet, and de Janvry (2003) when selecting PTC-related variables. This study follows Key, Sadoulet, and de Janvry (2000) by arguing that the approach by Key, Sadoulet, and de Janvry (2000) is more robust to the functional form of PTC and the estimation is less complicated, as described more in Section 4.

Some of the recent papers center their focus on the estimation of the unobserved transactions costs in addition to the supply response. Key, Sadoulet, and de Janvry (2000), with the information of both farmgate price and purchase price of maize in Kenya, estimate jointly the price elasticities of supply and demand by maize producing households and transaction costs. The estimation methods by Key, Sadoulet, and de Janvry (2000) or Holloway, Barrett, and Ehui (2005) are computationally formidable, although methods are expected to provide better estimates of elasticities than when the unobserved thresholds are ignored.

2.3 Literature comparing on-farm sellers and off-farm sellers

Fafchamps and Hill (2005) analyzes factors that affect the decision-making on where to sell, and conduct empirical analysis for coffee producers in Uganda. Findings by Fafchamps and Hill (2005) suggests that the opportunity costs of traveling play a key role in deciding where to sell, and coffee producers with a larger sales quantity tend to travel to the market. This argument is supported by Fafchamps, Gabre-Madhin, and Minten (2005) which suggests no economies of scale for traders thus farmers with large sales quantity do not necessarily attract more pick-up traders with lower transactions costs.

2.4 Literature analyzing the behavior under two-selection criteria

The sample selection model proposed by Heckman (1979) have been extended to the case of multiple selection criteria by Catsiapis and Robinson (1982) and Maddala (1983). Several studies apply the model with dual selection criteria. Vijverberg (1995) analyzes the labor-wage relationship in Ivory Coast when the wage and labor is observed only for those who first migrate to the city, and then decide to work given the wage they receive.

3 Conceptual framework

This section summarizes the theoretical framework that describes the behavior by cassava producers. The definition of parameters are on Table 1.

3.1 Utility maximization by the cassava producing households

Table 1: Definition of parameters

Parameters	Definition
A_k	Endowment in good k
c_k	Consumption of good k
$G(\cdot)$	Production technology
m_k	Net supply of k to (purchase of k from) the market
p_k^m	Market price of goods k
q_k	Production of goods k
T	Exogenous transfers and other incomes
t_p^R	Proportional transaction costs for type-R producer
t_f^R	Fixed transaction costs for type-R producer
t_f^S	Fixed transaction costs for type-S seller
x_k	Input k
$U(c;z_u)$	Utility as a function of c and z_u
W_k	non-productive liquid household wealth at the beginning of period
z^u	Exogenous shifters in utility
z^c	Exogenous shifters in demand function
z^q	Exogenous shifters in production function
I_k^R	=1 if a producer is in regime R $\forall R \in$ {Buyer, Autarky, Seller}
I_k^S	=1 if a seller is of type S $\forall S \in$ {On farm seller, Off farm seller}
λ	Lagrange multiplier for profit
μ	Lagrange multiplier
ϕ	Lagrange multiplier for production technology

Extending Bellemare and Barrett (2006) and Key, Sadoulet, and de Janvry (2000), a cassava household's utility maximization problem for sequential decision making can be expressed as the following:

$$\max_{I_t^R, I_t^S, q_t^R} \quad \sum_{t=0}^{2} u_t(c_k; z_u) \tag{3.1}$$

subject to

$$\sum_{t=0}^{2}\sum_{k=1}^{K}\Big\{\Big[(p^m_{kt}-t^s_{pkt})I^{\text{off farm}}_{kt}+p^f_{kt}I^{\text{on farm}}_{kt}+(p^m_{kt}+t^b_{pkt})I^{\text{buyer}}_{kt}\Big]m_{kt}$$

$$-\sum_R\sum_{k=1}^{K}t^R_{f,kt}I^R_{kt}\Big|_{t=0}-\sum_R\sum_{k=1}^{K}t^S_{f,kt}I^S_{kt}\Big|_{t=1}+T_t\Big\}=0 \quad (3.2)$$

$$q_{kt}-x_{kt}+W_t+A_{kt}-m_{kt}-c_{kt}=0, \quad \forall k\in 1,\dots,\text{cassava},\dots,K \quad (3.3)$$

$$W_1=W_0-\sum_R I^R_{kt}t^R_f \quad (3.4)$$

$$W_2=W_1-\sum_S I^S_{kt}t^S_f \quad (3.5)$$

$$G(q,x;z_q)=0 \quad (3.6)$$

$$c_{kt},q_{kt},x_{kt}\geq 0 \quad (3.7)$$

It is important to note first that there are three time period $t=0,1,2$, in each of which period there is corresponding utility measure $u_t(c_k;z_u)$. Here we assume that a producer determines I^R_t at $t=0$, I^S_t at $t=1$, q^R_t at $t=2$. The utility maximization at $t=2$ can therefore be solved by usual first order conditions, which will be expressed in section 3.1.1.

As in Key, Sadoulet, and de Janvry (2000) and Bellemare and Barrett (2006), conditions (3.2) through (3.6) are interpreted as the following; the cash condition (3.2) states that revenues from all sales an other income transfers must cover expenditures on all purchases. the condition (3.3) requires that for each of K goods, consumption, input use, sales quantity must be equal to the production, purchase and endowment in the beginning of period t. (3.4) states the condition for the unproductive liquid wealth in the beginning of period 0 and 1, which is affected by the fixed cost of market participation. The condition (3.6) sates the production technology that determines how the inputs and the outputs are related.

In period $t=0$, a producer chooses a regime out of 3 regimes, namely a seller, autarky, a buyer based on the relevant characteristics of himself or the markets. The producer, after deciding to be a seller, decides $t=1$ whether to be a on-farm seller or an off-farm seller based on some characteristics. decides cassava production and inputs based on the expected cassava price and input costs in $t = 2$. The producer who decided to be autarkic decides in period $t = 2$ the production and inputs based on factors such as utility from cassava consumption and production costs. If the seller makes decisions sequentially, then $I^R_{\text{cassava},0} = I^R_{\text{cassava},1} = I^R_{\text{cassava},2}$ with $I^S_{\text{cassava},1} = I^S_{\text{cassava},2}$ and $q_{\text{cassava},0} = m_{\text{cassava},0} = q_{\text{cassava},1} = m_{\text{cassava},1} = 0$. $I^S_{\text{cassava},0}$ is undefined here since a producer does not decide which seller to be before he decides to be a seller at all in period $t = 0$. In other words, the regime choices are made only in $t = 0$ and $t = 1$ while cassava production, sales and consumption decisions are made only in $t = 2$.

3.1.1 Empirical strategy

The elasticities of production and consumption of cassava with respect to the farmgate price are obtained as the relationship between the solutions in utility maximization problem (3.1) and the cassava price. Although the specifications (3.1) through (3.7) consists of three period $t(t = 0, 1, 2)$, c_k is solved for only $t = 2$, meaning that the usual first order condition can be used to obtain the solutions. Modifying Key, Sadoulet, and de Janvry (2000), Lagrangian of this utility maximization problem for $t = 2$ can be expressed as the following (subscript t is suppressed);

$$\begin{aligned} L &= u(c; z_u) + \sum_{k=1}^{K} \mu_k (q_k - x_k + W + A_k - m_k - c_k) + \phi G(q, x; z_q) \\ &+ \lambda \left\{ \sum_{k=1}^{K} \left[(p^m_k - t^s_{pk}) I^{\text{Off farm}}_k + p^f_k I^{\text{On farm}}_k + (p^m_k + t^b_{pk}) I^{\text{buyer}}_k \right] m_k \right. \\ &- \left. \sum_R \sum_{k=1}^{K} t^R_{f,kt} I^R_{kt} \Big|_{t=0} - \sum_R \sum_{k=1}^{K} t^S_{f,kt} I^S_{kt} \Big|_{t=1} + T \right\} \qquad (3.8) \end{aligned}$$

and FOCs for commodity k are,

$$
\begin{aligned}
&(\text{Consumption } c_k) && \frac{\partial u}{\partial c_k} - \mu_k = 0 && (3.9)\\
&(\text{Production } q_k) && \mu_k + \phi \frac{\partial G}{\partial q_k} = 0 && (3.10)\\
&(\text{Input } x_k) && -\mu_k + \phi \frac{\partial G}{\partial x_k} = 0 && (3.11)\\
&(\text{Sales } m_k) && -\mu_k + \lambda \left[(p_k^m - t_{pk}^s) I_k^{\text{off farm}} + p_k^f I_k^{\text{on farm}} + (p_k^m + t_{pk}^b) I_k^{\text{buyer}} \right] &&\\
& && = 0 && (3.12)
\end{aligned}
$$

From (3.9) and (3.12), we have

$$
\frac{\partial u}{\partial c_k} = \lambda \left[(p_k^m - t_{pk}^s) I_k^{\text{off farm}} + p_k^f I_k^{\text{on farm}} + (p_k^m + t_{pk}^b) I_k^{\text{buyer}} \right] \tag{3.13}
$$

(3.13) indicates that, if $\lambda > 0$ and the utility is strictly concave with respect to the consumption of cassava, a higher p^m for cassava leads to a lower optimal home consumption of cassava, i.e., a higher $\frac{\partial u}{\partial c_k}$. The possibility that $\frac{\partial c_k^*}{\partial p_k^m} \neq 0$ (more specifically, > 0) motivates the estimation of home consumption elasticity, which Key, Sadoulet, and de Janvry (2000) exclude but is included by Renkow, Hallstrom, and Karanja (2004).

The conditions (3.9) through (3.12) involve endogenous variables c_k, q_k and x_k and all the relevant exogenous variables. All Lagrange multipliers μ_k, ϕ_k and λ are the functions of the combinations of variables.

The price elasticity of supply of cassava for a cassava producing household is essentially the relationship between q_k^* and p_k^m determined by (3.9) through (3.12). The estimation of elasticity of each endogenous variables with respect to p_k^m in (3.9) through (3.12) reduces to regressing each endogenous variable independently or jointly on all the exogenous variables in the system (3.9) through (3.12).

The reduced form is generally then,

$$q = \alpha_q + \beta_p p + \sum_{k=\text{output}} \beta_k p_k + \sum_{\ell=\text{input}} \gamma_{w\ell} w_\ell + \sum_{m=\text{shifter}} \gamma_z z_m + u_q \tag{3.14}$$

where P is the relevant price of cassava, and p_k, w_l are the sales price of k-th commodity other than cassava and ℓ-th inputs while z_m are other factors that affect production, consumption of cassava.

3.1.2 Comparison of simultaneous decision-making and sequential-decision making

The first order conditions in the previous section illustrates the mechanism of how we observe the relationship between household production, consumption of cassava and its farmgate price. More generally, the solutions I_t^{R*} and q_t^{R*} for utility maximization problem (3.1) are obtained as reduced forms which are the functions of all the exogenous parameters in (3.1). The expressions of I_t^{R*} and q_t^{R*} are, however, different depending on whether a cassava producer makes decisions simultaneously or sequentially as in Bellemare and Barrett (2006). Sequential decision-making can be expressed as the following:

$$I_0^{R*} = I(A_{k0}, W_0, G_0(\cdot), t_{p0}^R, t_{f0}^R) \qquad \forall R \in \{\text{Regime}R\} \tag{3.15}$$

$$I_1^{S*} = I(A_{k1}, W_1, G_1(\cdot), t_{p1}^R, t_{f1}^R, I_0^{R*}) \qquad \forall S \in \{\text{Regime}S\} \tag{3.16}$$

$$q_2^{RS*} = Q(A_{k2}, W_2, G_2(\cdot), I_0^{R*}, I_1^{S*}, p_{k2}^m, t_{p2}^R) \tag{3.17}$$

in which I_0^{R*} is the decision made at $t = 0$ on whether to sell or not, I_1^{S*} is the decision made at $t = 1$ on which type of seller to be, and q_2^{RS*} is the quantity of production, consumption and sales decided at $t = 2$.

In contrast, the simultaneous decision-making gives,

$$I_0^{R*} = I(A_{k0}, W_0, G_0(\cdot), t_{p0}^R, t_{f0}^R) \qquad \forall R \in \{\text{Regime}R\} \tag{3.18}$$

$$q^{RS*} = Q(A_k, W, G(\cdot), p_k^m, t_p^R, t_{f1}^S, I_0^{R*}) \tag{3.19}$$

The differences between sequential and simultaneous decision-making are that some factors like t_f^R do not directly enter into the quantity equation (3.17).

The main theme of this study is to compare empirically (3.16) thorough (3.17) and (3.18) through (3.19) to see which scenario better explains the behavior of cassava producers.

3.1.3 Reasons for favoring the sequential decision-making in this study

The main theme of Bellemare and Barrett (2006) is to test if sequential decision-makings (3.16) and (3.17) better explain livestock traders' behaviors than the simultaneous decision-making (3.19), which this study does not test. Instead, this study tests

Very few sellers in the dataset (7 out of 217 sellers) report sales both at the farm and at the distant market, indicating that most sellers sell cassava either only at the farmgate or only at the distant market. This paper thus argues that many cassava sellers are more likely to decide first where to sell cassava, and decide production and consumption decision.

A sequential decision-making in aforementioned framework is empirically supported over simultaneous decision-making by Bellemare and Barrett (2006) for livestock traders in Kenya and Ethiopia.

The assumption of sequential decision-making in the above framework allows one to empirically estimate supply and demand elasticities with dataset in which cassava producers report price only when they decide to participate in the market.

3.1.4 Why do we analyze the actual production and demand instead of market supply and demand?

(1) We have not only the marketed quantity data, but also the data for home consumption by net sellers and production by net buyers. (2) We can not deny that the home consumption by net sellers has no information available in estimating the market demand behavior by the net buyers. Similarly, we cannot deny that the production behavior by net buyers has no information available in estimating the market supply behavior by the net sellers. In these cases, it is doubtful whether the information contained in actual production and demand behavior should be ignored. The author argues that the information about the actual production and home consumption behavior should be included in the estimation. In order to do that, it is probably better to estimate the elasticity of actual production and demand. The elasticities for these in the market can be automatically obtained from the combination of production and home consumption elasticities.

Assumption of sequential decision making versus simultaneous decision-making The study by Bellemare and Barrett (2006) suggests that producers make market-participation decisions and traded quantities sequentially rather than simultaneously. With an ideal dataset, one interesting research question is whether the same finding holds for cassava producers in Benin. The analysis by Bellemare and Barrett (2006), however, cannot be applied to this study due to the following reasons; 1)In Benin dataset, we observe cassava price only for sellers and buyers, but not for autarkic producers. 2)the dataset used by Bellemare and Barrett (2006) contains more detailed information about the variable costs and fixed cost of market sales or purchases, while Benin dataset contains very limited information about those costs. The finding by Bellemare and Barrett (2006), however, does provide one empirical example in which producers in the agricultural sector make market participation decisions not based on the expected price. This study thus argues that the same assumption for cassava producers in Benin may be appropriate and (3.14) can be consistently estimated without having the cassava price for autarkic producers if the self-selection bias is appropriately corrected.

4 Estimation of the model

This section first describes the estimation procedures used in this study, and then discuss issues associated with the estimation model.

4.1 Estimation procedure

In order to consider the potential simultaneity and dual criteria selection bias discussed in the conceptual framework in the previous section, this study proposes the following estimation procedure;

1. Estimate the market participation equation and seller-type equation by ordered probit (4.1) and probit (4.2)

$$I_i^{\text{sell or not}} = \text{Ordered probit}\left(\alpha^{\text{op}} + x_i^{\text{op}}\gamma_{\text{op}} + u_i^{\text{op}}\right) \quad (4.1)$$

$$I_i^{\text{seller type}} = \text{Probit}\left(x_i^{\text{pr}}\gamma_{\text{pr}} + u_i^{\text{pr}}\right) \quad (4.2)$$

2. For each observation i, obtain the Inverse Mill's Ratio $\hat{\lambda}_i^{\text{op}} = \frac{\phi(x_i^{\text{op}}\hat{\gamma}_{\text{op}})}{\Phi(x_i^{\text{op}}\hat{\gamma}_{\text{op}})}$ and $\hat{\lambda}_i^{\text{pr, off-farm}} = \frac{\phi(x_i^{\text{pr}}\hat{\gamma}_{\text{pr}})}{\Phi(x_i^{\text{pr}}\hat{\gamma}_{\text{pr}})}$, $\hat{\lambda}_i^{\text{pr, on-farm}} = -\frac{\phi(\psi_i^{\text{pr}}\hat{\gamma}_{\text{pr}})}{1-\Phi(\psi_i^{\text{pr}}\hat{\gamma}_{\text{pr}})}$ in which $\phi(\cdot)$ and $\Phi(\cdot)$ are probability density function and distribution function from normal distribution, respectively.

3. Estimate the equations for on-farm sellers (4.3) and off-farm sellers (4.4),

$$q_i^h = \alpha^h + p_i^{\text{farmgate}}\beta_p^h + z_i^h\beta_z^h + \beta_h^{\text{op}}\hat{\lambda}_i^{\text{op}} + \beta_h^{\text{pr}}\hat{\lambda}_i^{\text{pr, on-farm}} + u_i^h \quad (4.3)$$

$$q_i^h = \alpha^h + p_i^{\text{market}}\beta_p^h + z_i^h\beta_z^h + z_i^{\text{PTC}}\beta_{\text{PTC}}^h + \beta_h^{\text{op}}\hat{\lambda}_i^{\text{op}} + \beta_h^{\text{pr}}\hat{\lambda}_i^{\text{pr, off-farm}} + u_i^h \quad (4.4)$$

$$h = (\text{production, sales, consumption})$$

separately by using the two-stage least squares (2SLS) or jointly using the three-stage least squares (3SLS), with $\hat{\lambda}$'s included into the instrumental variables.

Several issues must be discussed regarding the procedures (4.1) through (4.4). In particular, the following subsections 4.1.1 through 4.1.2 discuss why this study apply dual-λ approach, how the estimations account for unobserved proportional transactions cost (PTC) for off-farm sellers. The other issues including technical details are discussed in the appendix C.

4.1.1 Correction of sample selection bias associated with multiple decision-making criteria

Vijverberg (1995) summarizes some of the methods used in the past studies which include conditional logit model and nested logit model.

Using conditional logit model assumes the independence of irrelevant alternatives (IIE). For example, suppose the change in characteristics of distant cassava market affects the probability that a producer becomes off-farm seller and thus the probability that a producer becomes one of the other three types, on-farm seller, stay autarkic, become buyer. IIE assumes that the proportion of each probability that a producer becomes on-farm seller, stay autarkic, or buyer remains unchanged.

This studies argue that, for cassava producers who often face significant fixed cost of participating the market, the decision of where to sell (choice between regime 2 and regime 3) is more likely to be nested in the preceding decision of whether to sell cassava at all. If so, the IIE assumption is unlikely to hold and thus Conditional logit model may be inappropriate.

In order to incorporate the nested structure in multiple decision-making, nested-logit model may be used. Vijverberg (1995), however, points out several shortcomings of nested logit approach. Nested logit model combines multiple selection effects into one selection effect. The combined selection effects sometimes, if not often, appear insignificant even when each selection effect is significant, if those selection effects somehow cancel each other out. Vijverberg (1995) is concerned that the consistency of OLS estimate is in doubt if each selection effect is significant but insignificant if combined together. This study, therefore, apply dual-λ approach used in Vijverberg (1995).

4.1.2 Correction for the unobserved PTC

Our interest here is the relationship between production, sales, consumption and the farmgate price of cassava. While on-farm sellers report the farmgate sales price of cassava, off-farm sellers report the sales price received at the market. The estimation of (4.4) for off-farm sellers thus needs to be corrected for the presence of unobserved PTC.

Two methods are often considered by the literature to account for unobserved PTC in this context. One way is to estimate (4.4) which uses reported market price and variables z_i^{PTC}'s which potentially explain PTC. Another way is to obtain predicted PTC and convert the reported market price into farmgate price using the predicted PTC.

Key, Sadoulet, and de Janvry (2000) uses the former approach while Vakis, Sadoulet, and de Janvry (2003) uses the latter. Key, Sadoulet, and de Janvry (2000), however, suggest that both approaches assume that unobserved PTC is a linear function of observed variables and are interlinked. This study, although follows Key, Sadoulet, and de Janvry (2000), argues that the former approach is better due to its robustness compared to the latter approach.

Methodology by Key, Sadoulet, and de Janvry (2000) and Vakis, Sadoulet, and de Janvry (2003)

Key, Sadoulet, and de Janvry (2000) estimates the model,

$$q^j = \alpha^j + \beta_p^j P + \gamma x_{\text{shifter}}^j + \beta_{PTC}^j z_{PTC} \quad j = (\text{sales}, \text{purchase}) \tag{4.5}$$

in which z_{PTC} includes transport costs per unit, distance to market and other related factors, arguing that PTC can be estimated as a linear combination of several factors such as $\text{PTC}_i^j = \alpha^j + \beta_{PTC}^j z_{PTC,i}$ (j = (sales, purchase); i = household). Vakis, Sadoulet, and de Janvry (2003) proposes the estimation,

$$\begin{aligned} PTC_i^{\text{Transport}} &= \alpha_{PTC} + \beta_{PTC} z_{PTC} + \gamma \lambda_{PTC} + u_{PTC} && (4.6) \\ i &: \text{Household who reports transport costs} \\ \widehat{PTC}_i &= \hat{\alpha}_{PTC} + \hat{\beta}_{PTC} z_{PTC} + \hat{\gamma} \lambda_{PTC} && (4.7) \end{aligned}$$

in which λ_{PTC} is Inverse Mills Ratio associated with the probability that a household i reports transport costs. Vakis, Sadoulet, and de Janvry (2003) then calculate the predicted value $\widehat{PTC}_i$ (including $\hat{\gamma}\hat{\lambda}_{PTC}$) for all households. (4.6) provides insights into which variables should go into z_{PTC}, which are then included into (4.4) to control for the unobserved PTC.

Although the method (4.5) is appropriate since the form of PTC is very difficult to specify, it raises some complications to the estimation of (4.4). Strictly speaking, the elasticities of production, sales and consumption and the coefficients for z_{PTC} should not be estimated independently. More explicitly, the inclusion of z_{PTC} into (4.4) requires the estimation of the following equations;

$$\begin{aligned} q_i^{\text{production}} &= (p_i^m - z_i^{PTC} \beta_q^{PTC}) \beta_q^p + z_i^q \beta_q^z + \beta_q^{\text{op}} \hat{\lambda}_i^{\text{op}} + \beta_q^{\text{pr}} \hat{\lambda}_i^{\text{pr}} + u_i && (4.8) \\ q_i^{\text{consumption}} &= (p_i^m - z_i^{PTC} \beta_c^{PTC}) \beta_c^p + z_i^q \beta_c^z + \beta_q^{\text{op}} \hat{\lambda}_i^{\text{op}} + \beta_q^{\text{pr}} \hat{\lambda}_i^{\text{pr}} + v_i && (4.9) \end{aligned}$$

which is a system of equations with constraint ($\beta_{PTC,h}$ = $\beta_{PTC,c}$). The estimation of (4.8) and (4.9) is difficult since the aforementioned constraint is actually non-linear (we estimate $\beta^h = \beta_{PTC,h}\beta_h$, $\beta^c = \beta_{PTC,c}\beta_c$ with non-linear constraint $\frac{\hat{\beta}^h}{\hat{\beta}^c} = \frac{\hat{\beta}_h}{\hat{\beta}_c})^2$.

The aforementioned constraints are necessary only if we assume that not only the statistical significance but also the magnitude of β^{PTC} is informative (so that $\hat{\beta}^{PTC} \cdot z_{PTC}$ is the actual measure of PTC.). Since the assumption has not been well tested in the past literature, it is beneficial to estimate another different model with alternative approach.

4.1.3 Estimation of PTC

Following Vakis, Sadoulet, and de Janvry (2003) and Henning and Henningsen (2007), I estimate the PTC in the following way;

$$\ln\left(\frac{Tr_i}{q_i^{\text{sale}}}\right) = \alpha_{\text{PTC}} + \beta_{\text{PTC}} z_i^{\text{PTC}} + \hat{\lambda}_i + u_i^{\text{PTC}} \tag{4.10}$$

in which λ_i corrects for the factors that make a seller to report transportation costs. λ_i is calculated using variables so that PTC_i is calculated for all households.

The equation (4.10) itself does not come into the estimation procedure (4.1) through (4.4). The regression (4.10), however, provides insights into what variables should be in z_i^{PTC} equation (4.4).

[2]Although we use natural log of price in our estimation, the same argument holds.

4.2 Theory behind the estimation procedures (4.1) through (4.4)

We have the following assumptions with regard to the error terms in (4.1) through (4.4) for each $h \in \{\text{production, sales, consumption}\}$ (notation h is omitted for simplicity);

$$\begin{pmatrix} u_i^{\text{op}} \\ u_i^{\text{pr}} \\ u_i^{\text{on}} \\ u_i^{\text{off}} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & & \\ \rho_{\text{pr}}^{\text{op}} & 1 & & \\ \sigma_{\text{on}}^{\text{op}} & \sigma_{\text{on}}^{\text{pr}} & \sigma_{\text{on}}^2 & \\ \sigma_{\text{off}}^{\text{op}} & \sigma_{\text{off}}^{\text{pr}} & \sigma_{\text{off}}^{\text{on}} & \sigma_{\text{off}}^2 \end{bmatrix} \right) \tag{4.11}$$

and

$$I_i^{\text{seller}} = \begin{cases} 1 & \text{if} \quad u_i^{\text{op}} > \alpha_2^{\text{op}} - x_i^{\text{op}} \gamma_i^{\text{op}} \\ 0 & \text{otherwise} \end{cases} \tag{4.12}$$

$$I_i^{\text{off-farm}} = \begin{cases} 1 & \text{if} \quad u_i^{\text{pr}} > -x_i^{\text{pr}} \gamma_i^{\text{pr}} \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

If $\rho_{\text{pr}}^{\text{op}} \neq 0$ we run ordered probit and probit models jointly, and if $\sigma_{\text{off}}^{\text{on}} \neq 0$ we run (4.3) and (4.4) jointly as a system of equations. Our preliminary results only show the case in which $\rho_{\text{pr}}^{\text{op}} = 0$ and $\sigma_{\text{off}}^{\text{on}} = 0$, so that each equation is estimated separately.

5 Dataset

This paper uses Benin Small Farmer Survey[3] collected by the IFPRI and LARES (Laboratoire d'Analyse Regionale et d'Expertise Sociale). Table 14 (p.40) provides the summary statistics of relevant parameters.

The dataset contains the information for 899 households. Among those 899 households, this study starts out with 552 cassava producing households which report the quantity of cassava harvested. The survey focus on the economic activities between April 1997 and March 1998. Out of the 552 cassava producing households, this study drops 9 households that are both sellers and buyers of cassava as is done in Key, Sadoulet, and de Janvry (2000) and Renkow, Hallstrom, and Karanja (2004) since the model used in this study fails to explain the behavior of those 9 households. The initial analysis is conducted for the remaining 543 cassava producing households.

Among the 543 cassava producing households are 217 net sellers, 310 autarkic households and 16 net buyers. The table 14 reveals the following picture of cassava producing households in the dataset. The size of cassava harvest varies considerably across households. Approximately 75 % ((2154 - 534) / 2154) of cassava harvested are on average sold by sellers, while a median seller sells only 37.5 % ((400 - 250) 400). Most of the autarky households are small scale compared to net sellers and even to net buyers.

Most households are located in the rural area. The nearest passable road and paved road are 1km and 12 km away for a median household. Many households are also distantly located from their farms (2km for a median household).

Many households rely on crop sales for their major source of income. On average the income from crop sales accounts for 55% of the total income. Cassava sales on average con-

[3]Benin: small farmer survey, 1998. 2004. Washington, D.C.: International Food Policy Research Institute (IFPRI)(datasets). "http://www.ifpri.org/data/benin01.htm"

tributes to 12 % of the total income and 32 % of the total crop sales.

There are several variables that are subject to the potential simultaneity problems. First variable is the price data, although the simultaneity of the price data is not due to the simultaneity in price and quantity as in the aggregate market since individual households are assumed not to affect the market price. The simultaneity of the price data is rather due to how it is obtained. Not all the price data were reported as unit price, but rather as the total value of sales or purchases. 192 out of 217 sellers and all net buyers report total value of transaction and quantity, instead of the unit price. This way of calculation does not always, although likely to cause simultaneity. There is little evidence that only the price data calculated by total values divided by quantity is causing simultaneity. While the correlation coefficient between ln(sales) and ln(price) are -.450 for the total values, it is -.766 for data reported as unit price. Therefore it is still arguable that, if correctly instrumented by instrumental variables, the regression of quantity on the price still produce a valid estimates.

Table 2: Some variables included in estimation and expected signs

	N	Selection		Production		Sales		Consumption	
		Sell	Where	On farm	Off farm	On farm	Off farm	On farm	Off farm
Raw Data									
Cassava price	233			+	+	+	+	-	-
				Input prices					
Hire labor or not	222			-	-	-	-	?	?
Time for leisure	540			-	-	-	-	?	?
Time for leisure^2	540			-	-	-	-	?	?
Opportunity cost of land	540			-	-	-	-	?	?
Farm size	540			-	-	-	-	?	?
Distance to plot	540			-	-	-	-	?	?
				PTC related					
Nearest phone service (km)	539			-	-	-	-	+	+
Nearest passable road (km)	539			-	-	-	-	+	+
Nearest paved road (km)	543			-	-	-	-	+	+
Membership to cooperative	543			+	+	+	+	?	?
Access to credit	543			+	+	+	+	?	?
Have extension agent or not	543			+	+	+	+	-	-
Own car/truck	543			+	+	+	+	?	?
Own motorcycle	543			+	+	+	+	?	?
Own bicycle	543			+	+	+	+	?	?
				Shifter					
Age of hhd head (years)	543	?	?					?	?
Household size	543	?	?					+	+
Education of hhd head	543	+	+	+	+	+	+	?	?
Gender of head	542	-	-						
Total income	542	-	-						
Retired income	543	+	+						
Total asset	543	+	+						
Distance to the source water	542	-	-						
Storage capacity	543	?	?						
Have extension agent or not	543	+	+						
# of traders in village									
Grow									
Maize	543	?						+	+
Cotton	543	+	+	+	+	+	+	?	?
Beans / Cow peas	539	-	-						
Chili	543	-	-						
Okra	542	-	-						
Rice	543	+	+						
Sorghum / millet	543	+	+						
Sweet potato	543	-	-						
Taro	543	?	?						
Tomato	543	+	+						
Yams	543	+	+						

6 Preliminary results

This section discusses the highlight of results using a series of tables. The tables in this section use two symbols, $\hat{\beta}$ for estimated coefficient, and $\hat{S}(\beta)$ for standard errors of the estimated coefficient. The presented results are still preliminary versions and are likely to change in the future. The interpretation of the results at this point focuses on the overall meaning of each regression.

6.1 Two selection stages

The results of regressions are presented in Table 28. For simplicity, I use notations as in Table 3 in the following discussion of results.

Table 3: Definition of parameters

Notation	Definition
$\varepsilon^{\text{pro}}_{\text{on farm}}$	elasticity of production for on-farm sellers
$\varepsilon^{\text{pro}}_{\text{off farm}}$	elasticity of production for on-farm sellers
$\varepsilon^{\text{sales}}_{\text{on farm}}$	elasticity of sales for on-farm sellers
$\varepsilon^{\text{sales}}_{\text{off farm}}$	elasticity of sales for on-farm sellers
$\varepsilon^{\text{con}}_{\text{on farm}}$	elasticity of consumption for on-farm sellers
$\varepsilon^{\text{con}}_{\text{off farm}}$	elasticity of consumption for on-farm sellers

6.2 Analysis of proportional transactions costs (PTC)

The regression (4.10) was run to identify some of the important factors that affect the unobserved PTC. As was discussed earlier, some off-farm sellers report more than one sales transactions including price, distance traveled, transportation costs paid. Following Vakis, Sadoulet, and de Janvry (2003), I here treat each sales transaction as individual observation, and run (4.10) for 54 reported sales transactions. The results are reported in Table 4.

For distance variables, I use the distance to assembly point and the consumption market (km) reported for each village by the village leader, instead of actual distance traveled by off-farm sellers. Although the actual distance traveled has better explanatory power, it may be endogenous to the PTC, and also the variable is not reported for on-farm sellers, which is problematic when we compare on-farm sellers with off-farm sellers in the later sections. I also assume that the actual distance traveled may depend on the sellers characteristics which can be in part explained by other observable characteristics such as the years of education of household head, whether to have a membership to the cooperative, and distance to the phone services. As expected, some of those characteristics significantly affect the ratio between transportation costs and sales quantity.

Table 4: Estimated log(PTC)

ln(Transportation costs / sales quantity)	$\hat{\beta}$	$\hat{S}(\beta)$
ln(distance to assembly point (km))	.474*	.254
ln(distance to consumption (km))	.211	.201
distance to phone (km)	.006	.004
household head education (year)	-.138**	.052
belong to cooperative	-1.217*	.553
λ_{report}	-.128	.108
λ_{opro}	.084***	.019
λ_{pro}	-.032	.479
Constant	.186	.559
p-value (overall significance)	.000	
R^2	.305	
No of clusters (village)	12	
No of obs	54	

Significance levels : $*$: 10% $**$: 5% $***$: 1%

The results in Table 4 suggest that the variables like years of education of household head, membership to cooperatives and distance to phone should be included in the production and consumption equations for off-farm sellers.

Vakis, Sadoulet, and de Janvry (2003) suggests that the predicted value from the regression (4.10) should be added in the production and consumption equations instead of adding each explanatory variable. While their approach is beneficial in terms of preserving the degrees of freedom in the production and consumption equations, this study argues that there are certain drawbacks. First, the assumption that the PTC is a linear functions of certain explanatory variables may not hold in our dataset, which makes the estimation more robust by adding each explanatory variables instead of predicted PTC. Second, the predicted PTC is obtained from the regression and thus obtained with standard errors. Including predicted PTC into our production and consumption equations complicates the estimation of standard errors in these equations since the standard errors then must be corrected for not only for λ's but also for predicted PTC. For these reasons, I add each explanatory variable to off-farm seller's production and consumption equations instead of predicted PTC in the following section.

6.3 Decision on market participation and seller type

Table 28 summarizes the results of ordered probit regression (4.1) and probit regression (4.2).

The results from ordered probit regression suggest that the market participation decision by cassava producers may be highly influenced by which region they live and what ethnic group they belong to. It is plausible that the region of residence influences the market participation decisions since regions often differ in many socio economic characteristics which are unobserved in the dataset. It is, however, less clear how the ethnicity affects the market participation decisions. One possibility is that the dummy variable for ethnicity explains many other observed variables so that ethnicity dummy variables are highly significant and many other variables are insignificant. Dropping the ethnicity dummies, however, significantly lowers the overall explanatory power of the model as is indicated by the Likelihood Ratio test in the second row from the bottom of table 28. I therefore conclude that the ethnicity of a cassava producer contains important information not contained in other observable variables which determine his market participation decision.

Table 5: Results of ordered probit and probit regression (4.1) and (4.2)

Dependent variable	Ordered Probit buyer = 0, autarky = 1 seller = 2		Probit on farm seller = 0 off farm seller = 1	
	$\hat{\gamma}_{opro}$	Std.err	$\hat{\gamma}_{pro}$	Std.err
Household size	.002	.023	.027	.030
Gender of hhd head	-.709*	.408	1.112**	.499
Age of household head	.001	.007	-.017*	.009
Education of head	-.037	.036	-.051*	.031
Dependency			-.165	.143
Agricultural asset (1000 US$)	.069	.075	.291**	.137
Nonag asset (1000 US$)	.339	.208	.130	.132
Farm size (ha)	-.043	.034		
Retired income (1000 US$)	.368	.594	-.548	.486
Storage capacity (1000t)	.004	.779		
Distance to phone (km)	.006	.007	-.047***	.014
Distance to passable road (km)	-.017**	.008		
Distance to plot (km)	.056	.041		
number of traders in village	3.536***	.369	.157**	.072
ag member			-.157	.231
time leisure			3.272***	1.103
maize seller			-.018	.321
cotton seller			1.637**	.656
bambara nuts grower	.233	.562		
bean grower	-.096	.203	.008	.297
chili grower	-.157	.320	.209	.636
cotton grower	.061	.419	-4.957***	1.563
finger millet grower	1.304	1.556		
fruit grower	-1.746**	.714	.113	.974
ground nuts grower	.186	.251	.299	.272
maize grower	.232	.410	.518	.844
okra grower	.238	.343	.637	1.236
onion grower	.554	2.592		
other vegetables grower	.810**	.388	.470	.641
rice grower	-.235	.373	-2.000	1.271
sorghum grower	-.557	.390	-3.287***	1.121
soya grower	.577	1.201		
taro grower	.579	1.268	1.463	1.497
tomato grower	-.178	.276	.240	.314
yams grower	.017	.409	.089	.548
dep2	1.866**	.779	-6.501***	1.471
dep3	1.066**	.523	-.027	1.061
dep4	.775	.798	-8.160***	1.617
dep5	3.735***	.698	-5.941***	1.332
dep6	.810	.723		
ethnic2	-.040	.614	2.818***	.751
ethnic3	.612	.463	1.059***	.404
ethnic4	1.003**	.483	1.301	1.224
ethnic5	1.792**	.898		
ethnic8	-1.889***	.679		
ethnic9	2.243***	.849		
ethnic10	1.321	1.809		
ethnic11	.899***	.345		
a_1	-.300	.847		
a_2	4.073***	.885		
Log-likelihood	-131.07		-78.78	
p-value for overall significance	.000		.000	
% of correct prediction				
0	.07		.70	
1	.93		.92	
2	.95			
p-value for ethnic = 0 (LR test)	.000			
No of obs.	535		207	

Significance levels : $*$: 10% $**$: 5% $***$: 1%

6.4 Production, sales and consumption decisions

Tables 6 through 8 present the results of important variables for each equation. The comprehensive results of regressions are presented in Tables 27 (p.47) through 29 (p.49) in the appendix.

6.4.1 Cassava production equation

Table 6 indicates the following. There is stronger evidence that cassava production by on-farm sellers is elastic to the farmgate price. The elasticity of production by off-farm sellers is inconclusive since $\hat{\varepsilon}^{p}_{\text{off farm}}$ is not significantly different from zero, but also not significantly different from $\hat{\varepsilon}^{p}_{\text{on farm}}$. When two types of sellers are combined, the production elasticity reduces in magnitude.

(Other variables)...

The coefficients of the form of cassava (flour, dried tuber) need to be interpreted carefully. First, it must be noted that the base results without dummies for the form are for the fresh tuber cassava. The dummy for flour is one if the cassava producer report the harvest quantity (kg) in flour form, and similarly for dried tuber. All households used in the regression report only one form of cassava. The coefficients for the form contains two effects, one of which is the difference in quantity due to conversion (ex. quantity reduces when cassava changes from fresh tuber to dried tuber), and the other is the difference in the value per quantity after the difference in quantity is controlled for.

With approximated conversion factor from Indonesia (1kg of fresh tuber $\approx$ 0.4kg of dried tuber and flour) as a reference, it is possible to extract the value of converting the form of cassava from the coefficient estimates. The $\hat{\delta}$ and its standard error $\hat{S}(\delta)$ in the tables 6 through 8 can be interpreted in the following way. For example, .143 in production equation means that, for cassava on-farm sellers .143kg of fresh tuber cassava plus added value of converting fresh tuber to flour is equivalent to 1kg of fresh tuber cassava. The form dummies are, however, included merely to reduce the omitted variable bias in estimation of the elasticities, and the actual estimate of value of converting forms of cassava should be estimated not from the coefficients in this study but in the different specifications of model.

6.4.2 Cassava Consumption equation

When estimated separately, home consumption of cassava by on-farm sellers appear responsive to the farmgate price while off-farm sellers' consumption is again inconclusive. When two types of sellers are combined, the η^{con} is estimated to be significantly negative ($\hat{\eta}^{\text{con}} = -.528$).

(Other variables)...

Table 6: Cassava production with sellers separated

Dependent variable	On Farm Sellers				Off Farm Sellers			
ln(Production(kg))	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$
ln(Price)	.421**	.173			.238	.182		
Household size					.181***	.047		
Education of head	.066***	.023			.029	.053		
Total asset	.608**	.240						
Farm size	.087**	.034			.086	.061		
Bicycle, own					-.338	.315		
Motorcycle, own					-.306	.382		
Car/truck, own					1.874*	.988		
Distance to phone (km)					.021***	.007		
number of traders in village	.158**	.068						
ag member					-.815**	.322		
cotton grower	-1.235**	.550			.039	1.138		
maize grower	-.632*	.363			.622	.773		
fresh tuber	-3.018***	.550	.143	.00	-.190	.397	2.25	.14
dried tuber	-1.835***	.470	.448	.01	.235	.536	3.67	.71
ln(consumption market (km))					.242	.172		
distance to sales point					-.002	.013		
squared distance to sales point					.000	.000		
dep2	2.158***	.215						
dep4	1.275***	.305						
λ_{opro}	.047*	.028			-.170***	.055		
λ_{probit}	-.012	.096			.575*	.339		
Constant	5.608***	.808			3.589**	1.430		
No of obs.	113				60			
No of cluster (village)	30							

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 7: Cassava consumption with sellers separated

Dependent variable	On Farm Sellers				Off Farm Sellers			
ln(Consumption(kg))	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$
ln(price)	-.651***	.224			-.579**	.227		
Household size	.129***	.046			.148***	.055		
Education of head					.137**	.066		
Dependency	-.269**	.111						
Total asset	.722*	.380						
Total income (1000 US$)	.262*	.144			.139*	.080		
Bicycle, own					-.076	.384		
Motorcycle, own					.266	.482		
Car/truck, own					1.501	1.080		
Distance to phone (km)					-.005	.008		
Distance to plot (km)	-.014*	.007						
ag member					.499	.415		
maize grower	-.297	.682			.654	.979		
yams grower	-.979	.708						
flour	.803	.676	7.06	4.57	.948**	.479	7.28	2.16
dried tuber	.658	.830	6.85	7.36	1.427**	.667	13.09	15.20
dep2					.141	.197		
dep3					-.013	.016		
dep4					.000	.000		
ethnic2	.309	.474						
ethnic3	.013	.892						
ethnic8	-1.214	1.254						
ethnic9	1.519	1.523						
ethnic11	-.031	.260						
λ_{opro}	-.195***	.056			-.123*	.069		
λ_{probit}	.028	.146			.620	.436		
Constant	7.909***	1.312			4.814***	1.755		
No of obs.	106				59			
No of cluster (village)	29				28			

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 8: Cassava production and consumption with sellers pooled

Dependent variable	ln(Production(kg))				ln(Consumption(kg))			
	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$	$\hat{\beta}$	$\hat{S}(\beta)$	$\hat{\delta}$	$\hat{S}(\delta)$
ln(Price)	.200*	.109			-.528***	.131		
Household size	.066**	.031			.105***	.034		
Gender	-.310	.337						
Education of head	.082***	.024			.001	.029		
Total asset	.265	.205			.000	.268		
Farm size	.112***	.036			.144***	.044		
Bicycle, own	-.080	.157			-.025	.190		
Motorcycle, own	-.181	.190			-.052	.236		
Car/truck, own	-.685	1.280			-1.057	1.597		
Distance to phone (km)	6.995	.006			.001	.006		
number of traders in village	.086	.073						
ag member	-.226	.155			-.349*	.191		
cotton grower	-.590	.471						
maize grower	-.264	.352			-.114	.477		
yams grower	.537*	.292						
fresh tuber	-1.803***	.294	.433	.00	-.024	.337	2.60	.13
dried tuber	-1.534***	.389	.585	.01	.266	.429	3.60	.41
ln(consumption market (km))	-.095	.085			-.024	.095		
distance to sales point	-.007	.010			-.004	.013		
squared distance to sales point	.000	.000			.000	.000		
dep2	1.755***	.190						
dep6	1.276***	.392						
λ_{opro}	-.029	.041			-.105***	.041		
Constant	6.200***	.600			6.856***	.758		
No of obs.	173				165			
No of cluster (village)	36							

Significance levels : $*$: 10% $**$: 5% $***$: 1%

6.5 Summary of the findings

Table 9 summarizes the estimated elasticities of production, sales and consumption.

Table 9: Elasticities estimates

Method	On farm		Off farm		Pooled	
Production	.421	(.173)	.238	(.182)	.200	(.109)
Consumption	-.651	(.224)	-.579	(.227)	-.528	(.131)

Table 10 summarizes ρ, the estimates of correlation coefficients with p-values in parenthesis. Significant ρ indicates the presence of correlation between u_i^{opro}, u_i^{pro} in (4.1) and (4.2), and u_i^h in (4.2), which can cause the sample-selection biases in the estimation. For some equations (for example, sales equation for off-farm sellers), the results indicate the non-zero correlations between all three error terms. The findings support the methodology in this study which include two selection terms (λ_{opro} and λ_{pro}) to obtain the consistent estimates of elasticities.

The sign of ρ_{pro} has the opposite implication in the case of ρ_{pro}. Applying the argument by Dolton and Makepeace (1987), if there are two types of seller a producer can choose to be, a positive ρ_{pro} in on-farm seller equation means that the expected value of q^h decreases as the producer is more likely to be an on-farm seller. For example, a significantly positive $\rho_{pro}(= .629)$ for off-farm sellers sales equation, the more likely a producer (who has decided to be a seller) chooses to be an off-farm seller, the smaller quantity he will sell as an off-farm seller.

Table 10: Estimated ρ (p-value in parenthesis)

	Method	On farm	On farm	Off farm	Off farm	Pooled	Pooled
Production	ρ_{opro}	.062	(.037)	-.166	(.054)	.020	(.652)
	ρ_{pro}	-.016	(.128)	.562	(.331)		
Consumption	ρ_{opro}	-.185	(.053)	-.152	(.045)	-.115	(.001)
	ρ_{pro}	.026	(.136)	.857	(.264)		

6.6 Comparison of two models

Table 11 summarizes the results of the non-nested J-test suggested by Davidson and MacKinnon (1981). The results from the J-test can be interpreted as the following. The model with two types of sellers separated seems to better explain the behaviors of on-farm sellers, than the model with two types of sellers pooled together. On the contrary, the former model does not seem to have any better explanatory power to explain the behavior of off-farm sellers compared to the latter model.

The findings from J-test are rather mixed partly because many estimation models are still misspecified, as evidenced by generally low p-values.

Table 11: Non-nested *J*-test

	Separate		Pooled
	On-farm sellers	Off-farm sellers	
Production	.895	.000	.001
Consumption	.003	.094	.007

The .895 in the upper-left cell is the *p*-value of the coefficient on predicted value obtained from pooled estimation in the separate estimation. *p*-value of .895 means that the pooled estimation of production does not add any explanatory power to the estimation of production for on-farm sellers.

References

Bellemare, M., and C. Barrett. 2006. “An Ordered Tobit Model of Market Participation: Evidence from Kenya and Ethiopia.” *American Journal of Agricultural Economics* 88:324–337.

Catsiapis, G., and C. Robinson. 1982. “Sample selection bias with multiple selection rules: An application to student aid grants.” *Journal of Econometrics* 18:351–368.

Cragg, J. 1971. “Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods.” *Econometrica* 39:829–844.

Davidson, R., and J.B. MacKinnon. 1981. “Several Tests for Model Specification in the Presence of Alternative Hypotheses.” *Econometrica* 49:781–793.

de Janvry, A., and E. Sadoulet. 2005. “Progress in the modeling or rural households’ behavior under market failures.” In A. de Janvry and R. Kanbur, eds. *Poverty, Inequality and Development, Essays in Honor of Erik Thorbecke*. Kluwer publishing, chap. 8.

Dolton, P.J., and G.H. Makepeace. 1987. “Interpreting Sample Selection Effects.” *Economics Letters* 24:373–379.

Fafchamps, M., E. Gabre-Madhin, and B. Minten. 2005. “Increasing returns and market efficiency in agricultural trade.” *Journal of Development Economics* 78:406–442.

Fafchamps, M., and R. Hill. 2005. “Selling at the farmgate or traveling to market.” *American Journal of Agricultural Economics* 87:717–734.

Goetz, S. 1992. “A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa.” *American Journal of Agricultural Economics* 74:444–452.

Greene, W.H. 2003. *Econometric Analysis*. Pearson Education.

—. 1981. “Sample selection bias as a specification error: a comment.” *Econometrica* 49:795–798.

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

Heltberg, R., and F. Tarp. 2002. “Agricultural Supply Response and Poverty in Mozambique.” *Food Policy* 27:103–124.

Henning, C., and A. Henningsen. 2007. “Modeling Farm Households’ Price Responses in the Presence of Transaction Costs and Heterogeneity in Labor Markets.” *American Journal of Agricultural Economics* 89:665–681.

Holloway, G., C.B. Barrett, and S. Ehui. 2005. “The Double-Hurdle Model in the Presence of Fixed Costs.” *Journal of International Agricultural Trade and Development* 1:17–28.

Key, N., E. Sadoulet, and A. de Janvry. 2000. “Transactions costs and agricultural household supply response.” *American Journal of Agricultural Economics* 82:245–259.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, England: Cambridge University Press.

Qaim, M. 2001. "A prospective evaluation of biotechnology in semi-subsistence agriculture." *Agricultural Economics* 25:165–175.

Renkow, M., D.G. Hallstrom, and D.D. Karanja. 2004. "Rural infrastructure, transactions costs and market participation in Kenya." *Journal of Development Economics* 73:349–367.

Strauss, J. 1984. "Marketed Surplus of Agricultural Households in Sierra Leone." *American Journal of Agricultural Economics* 66:321–31.

Toquero, Z., J.B. Duff, T.L. Anden, and Y. Hayami. 1975. "Marketable Surplus Functions for a Subsistence Crop: Rice in the Philippines." *American Journal of Agricultural Economics* 57:705–709.

Vakis, R., E. Sadoulet, and A. de Janvry. 2003. "Measuring Transactions Costs from Observed Behavior: Market Choices in Peru." CUDARE Working Papers, Department of Agricultural and Resource Economics, University of California Berkeley, USA.

Vijverberg, W.P.M. 1995. "Dual selection criteria with multiple alternatives: migration, work status, and wages." *International Economic Review* 36:159–185.

A Description of dataset

Table 12: Definition of terms in this paper

Terms	Description
Department	Benin's political unit, equivalent to prefecture: there were 6 departments at the time of the survey in 1998. See figure 2(a)(p.41) for the location of each department. (There are 12 departments in Benin in 2008.)
Off farm cassava seller	Households which produce, sell cassava and report sales point other than "farm" and distance to the sales point
On farm cassava seller	Households which produce, sell cassava and report sales point "farm"

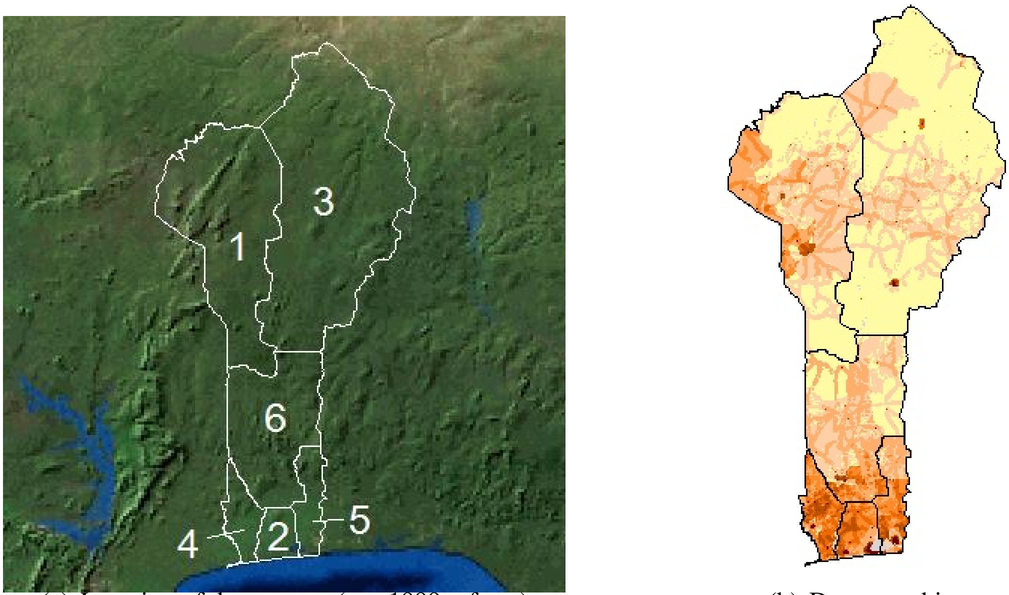
Table 13: Description of some variables

Variable	Description
Asset, agricultural	cart, plow, harrow, tractor, cattle, work cattle, goats/mutton, donkey/horses, pigs, poultry, other animals
Asset, non-agricultural	chairs, tables, beds, other furniture, heater/oven, electric fan, radio/cassettes/CD, TV/VCR, sewing machine, bicycle, motorcycle, car/truck, refrigerator
Dependency	$\frac{\text{\#of kids}(\leq 14 \text{ years old})\text{and elderly}(\geq 60 \text{ years old})}{\text{\#of members between 15 and 59}}$

Table 14: Summary statistics of cassava producing households

	On farm seller		Off farm seller		Autarky		Buyer	
No of observations	136		78		311		15	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Production (*t* /year)	5.6	(1.5)	3.2	(1.8)	.4	(2.0)	.5	(1.8)
Consumption (*t* /year)	.8	(1.4)	.6	(2.2)	.3	(1.2)	.9	(1.1)
Sales (*t* /year)	4.8	(1.6)	2.5	(2.1)				
Household size	8.2	(.6)	9.1	(.6)	9.2	(.6)	11.5	(.4)
Dependency	1.2	(.8)	1.1	(.8)	1.2	(.7)	1.1	(.6)
Age of household head	44.6	(.3)	46.3	(.3)	46.3	(.3)	48.5	(.3)
Education of head (year)	2.7	(1.3)	1.7	(1.6)	1.4	(2.2)	1.1	(2.7)
% of female head	3.7		6.4		4.8		13.3	
% of female member	49.8		49.6		48.9		49.0	
Farm size (ha)	3.3	(1.3)	4.1	(1.0)	5.5	(.9)	6.3	(1.1)
Cassava planted area (ha)	.8	(1.3)	1.0	(1.0)	.9	(.9)	1.0	(1.1)
Total asset (US $)	710	(2.0)	1195	(2.4)	955	(1.9)	2747	(2.0)
Agricultural Asset (US $)	331	(3.3)	521	(3.5)	521	(2.6)	2449	(2.1)
Income (US $)	1541	(1.3)	1743	(1.3)	1252	(1.1)	1404	(1.3)
Storage capacity (*t*)	2.9	(2.0)	15.5	(2.9)	22.3	(7.8)	8.7	(1.1)
Distance to								
phone service (km)	12.2	(1.0)	20.4	(1.4)	24.8	(.9)	25.5	(.9)
passable road (km)	2.4	(1.7)	6.8	(2.2)	6.4	(2.2)	18.4	(1.3)
paved road (km)	8.1	(1.2)	17.0	(1.1)	28.5	(1.1)	39.5	(.6)
own farm (km)	4.7	(4.0)	3.3	(1.6)	2.9	(.8)	2.1	(.7)
% access to rotating credit	76		64		49		33	
% own car/truck	0		4		1		0	
% own motorcycle	35		29		27		20	
% own bicycle	62		72		74		80	
cassava sales / total income (%)	14	(.9)	19	(.9)				
crops sales / total income (%)	51	(.6)	64	(.5)	68	(.5)	48	(.8)
cassava sales price (US cent/kg)	49	(1.1)	150	(.6)				

Figure 1: Benin



(a) Location of department (pre-1999 reform) (b) Demographics

Source: FAO

Table 15: Other characteristics of cassava producing households

	Total	On farm seller	Off farm seller	Autarky	Net buyer
% of cassava producers with the following land holding characteristics					
Own	64	61	73	61	72
Use without pay	14	21	9	11	11
Commune property	10	1	1	19	17
Rent	9	10	14	8	0
Sharecrop	1	1	0	1	0
Other	2	5	3	0	0
% of cassava producers who belong to the following groups					
Village group	35	15	35	44	25
Agricultural cooperative	10	17	6	7	6
Both of above	13	6	6	17	25
No	43	62	53	31	44
% of cassava producers who have access to extension agent					
Yes	64	47	70	68	80
No	36	53	30	32	20

Table 16: Number of cassava sellers with each sales destination by regions

		Department (Prefecture)					
	Total	1	2	3	4	5	6
# of cassava producers	544	41	77	119	71	124	112
# of cassava sellers	217	6	62	20	23	90	16
# of on-farm sellers	138	2	45	11	12	68	
# of off-farm sellers	79	4	17	9	11	22	16
# of reported sales point for off-farm sellers	84	4	17	10	13	24	16
Market	56	4	8	8	10	16	10
Family/Friend	21		9	1	1	5	5
Trader	2					1	1
On farm	5	0	0	1	2	2	0
Distance to the sales point (km)							
mean	14	14	9	7	7	9	36
median	6	20	6	8	8	5	11

Table 17: % of cassava producers who purchase inputs for cassava production

	All cassava producers	On farm seller	Off farm seller	Autarky	Buyer
Fertilizer	5	12	5	3	
Pesticides	1	3		0	
Seeds	6	9	6	4	13
Hire labor	41	64	58	26	20

Table 18: Breakdown (%) of cassava consumption, usage by cassava producing households

	All cassava producers	On farm seller	Off farm seller	Autarky	Buyer
Household Consumption	93	92	82	97	98
Livestock feed	0	0		0	0
Gift, payment	6	7	14	3	1
Other	1		4	0	

Table 19: Frequency of cassava planting, harvesting and sales activity by month (%)

	Planting	Harvest	Harvest(sellers)	Sales
January	3	10	7	6
February	5	12	12	15
March	21	13	17	16
April	40	6	12	22
May	15	0	0	12
June	7	1	1	8
July	7	1	1	5
August	3	3	3	6
September	0	1	2	2
October	0	3	3	2
November	0	2	3	1
December	0	13	5	5
Year round	0	34	35	0

Table 20: % of cassava producers who also produce and sell other crops (by crop)

	Sell					Produce				
	All	On farm seller	Off farm seller	Autar-ky	Buyer	All	On farm seller	Off farm seller	Autar-ky	Buyer
bambara nut			0	0	0	2		4	3	5
beans/cow peas	34	21	37	39	38	47	29	37	57	48
chili	11	7	11	13	10	12	7	11	16	10
cotton	40	9	30	56	43	40	9	30	56	43
finger millet			1	0		0		1	0	
fruit		0	0	0	0	1	1	1	1	5
ground nuts		0	0	0	0	25	19	26	26	38
maize	70	71	71	71	33	95	95	95	96	86
okra	13	3	11	17	14	22	7	12	31	24
onion	0			0		0			0	
other vegetables		0	0	0	0	7	4	10	8	5
rice	4	3	3	5	14	7	4	3	9	14
small grains		0	0	0	0	7	1	5	9	10
sorghum/millet	8	1	7	12	0	26	7	12	37	43
soya	1			2		1			2	
sweet potato		0	0	0	0	5	9	3	4	10
taro	1	1	1	1		1	1	1	1	
tomato	21	21	19	23	5	22	21	19	25	5
yams	22	7	16	31	5	38	13	26	52	52

Table 21: Cassava sales price by month (US cents kg)

	All			On farm			Off farm		
	Mean	Median	CV	Mean	Median	CV	Mean	Median	CV
January	83	80	.8	22	23	.5	123	135	.4
February	57	25	.9	32	24	.9	86	98	.7
March	46	29	1.0	37	26	1.1	67	55	.7
April	27	14	1.2	17	13	.7	70	50	.7
May	34	13	1.4	20	13	1.6	128	124	.1
June	91	108	.6	71	47	.9	102	117	.4
July	88	116	.5	17	16	.4	109	118	.3
August	66	45	.8	31	30	.6	87	107	.7
September	114	151	.6	23	23		137	161	.4
October	36	27	.9	36	27	.9			
November	13	13	.7	17	17	.5	3	3	
December	65	33	.8	39	30	1.0	96	114	.6

Table 22: Cassava production forms by types of producers (%)

	On farm	Off farm	Autarky	buyer
Fresh tuber	88	35	17	10
Dried tuber	8	20	73	80
Flour	4	45	10	10

Table 23: Cassava sales forms by department (region) (%)

	dep1	dep2	dep3	dep4	dep5	dep6
Fresh tuber		70	2	37	100	
Dried tuber	100		98	7		21
Flour		30		57		79

Table 24: Cassava sales price by department (region) and forms of sales (US cents / kg)

	Fresh tuber			Dried tuber			Flour		
	Mean	Median	CV	Mean	Median	CV	Mean	Median	CV
Department 1				42	40	.2			
Department 2	13	13	.1				136	125	.2
Department 3	5								
Department 4	43	19	1.3	20	20		103	117	.5
Department 5	33	25	.9						
Department 6				102	120	.4	114	107	.4
All	27	20	1.1	56	45	.7	121	123	.3

Table 25: Cassava sales price by months and forms (US cents kg)

	Fresh tuber				dried tuber				Flour			
	%	Mean	Median	CV	%	Mean	Median	CV	%	Mean	Median	CV
January	4	31	18	1.4	8	120	120	.0	3	119	117	.3
February	12	25	23	.5	15	85	85	1.1	13	120	115	.3
March	22	35	29	.9	31	43	50	.4	2	113	113	.2
April	29	17	13	.6	15	45	45	.2	7	121	120	.1
May	15	16	13	.3	0				10	130	120	.2
June	2	34	33	.4	23	34	35	.1	24	127	120	.2
July	2	15	12	.4	0				17	118	120	.1
August	4	27	20	.5	0				13	123	130	.3
September	1	27	27	.1	0				6	148	170	.3
October	3	26	24	.2	0				1	100	100	
November	2	23	25	.2	0				0			
December	4	37	30	1.0	8	120	120	.0	5	94	117	.7

Table 26: Cassava price with dummy for regions, months and form

price	Coef.	Std. Err.	*p*-value
dep1	-37.85	22.93	.100
dep2	20.56	17.52	.241
dep4	18.42	17.97	.306
dep5	26.95	17.11	.116
dep6	-2.46	18.25	.893
February	-5.22	7.04	.460
March	-1.52	6.84	.824
April	-13.43	7.12	.060
May	-12.70	7.48	.090
June	-7.20	8.11	.376
July	-16.55	8.61	.055
August	-7.04	8.08	.384
September	19.09	11.22	.090
October	-11.56	9.75	.236
November	-11.01	10.91	.314
December	-4.66	8.44	.581
flour	29.70	10.78	.006
fresh	-78.42	11.48	.000
Constant	86.39	20.40	.000

B Complete Results of the regressions

Table 27: Cassava production with sellers separated

Dependent variable	On Farm Sellers (2SLS)			Off Farm Sellers (OLS)		
ln(Production(kg))	$\hat{\beta}$	Std.err		$\hat{\beta}$	Std.err	
		Heckman	Original		Heckman	Original
ln(Price)	.421**	.173	.348	.238	.182	.326
Household size				.181***	.047	.050
Education of head	.066***	.023	.023	.029	.053	.056
Total asset	.608**	.240	.250			
Farm size	.087**	.034	.048	.086	.061	.063
Bicycle, own				-.338	.315	.335
Motorcycle, own				-.306	.382	.415
Car/truck, own				1.874*	.988	1.015
Distance to phone (km)				.021***	.007	.008
number of traders in village	.158**	.068	.083			
Membership				-.815**	.322	.337
cotton grower	-1.235**	.550	1.118	.039	1.138	1.180
maize grower	-.632*	.363	.331	.622	.773	.804
flour	-3.018***	.550	.773	-.190	.397	.513
dried tuber	-1.835***	.470	.565	.235	.536	.571
ln(consumption market (km))				.242	.172	.195
distance to sales point				-.002	.013	.013
squared distance to sales point				.000	.000	.000
dep2	2.158***	.215	.304			
dep4	1.275***	.305	.215			
λ_{opro}	.047*	.028	.031	-.170***	.055	.056
λ_{probit}	-.012	.096	.141	.575*	.339	.369
Constant	5.608***	.808	1.416	3.589**	1.430	2.004
ρ_{opro}	.062			-.166		
ρ_{pro}	-.016			.562		
p-value for overall significance	.000			.000		
Identification tests (p-value)						
H_0: Underidentified	.006			.042		
H_0: Not Overidentified				.051		
First stage R^2	.242			.348		
No of obs.	113			60		
No of cluster (village)	30					

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 28: Cassava consumption with sellers separated

Dependent variable	On Farm Sellers (2SLS)			Off Farm Sellers (OLS)		
ln(Production(kg))	$\hat{\beta}$	Std.err		$\hat{\beta}$	Std.err	
		Heckman	Original		Heckman	Original
Price	-.651***	.224	.349	-.299	.205	.322
Household size	.129***	.046	.037	.185***	.039	.047
Dependency	-.269**	.111	.134			
Total asset	.722*	.380	.231	.441***	.113	.135
Total income	.262*	.144	.117			
Distance to plot (km)	-.014*	.007	.004			
maize grower	-.297	.682	.489			
yams grower	-.979	.708	.606	-.135	.366	.415
Flour	.803	.676	.768	.437	.381	.557
dried tuber	.658	.830	.329	.575	.524	.655
ethnic2	.309	.474	.351	-.672*	.362	.427
ethnic3	.013	.892	.768	.340	.464	.552
ethnic5				.528	.734	.866
ethnic8	-1.214	1.254	.599			
ethnic9	1.519	1.523	.860	1.254	1.052	1.264
ethnic11	-.031	.260	.228	-.939**	.402	.461
λ_{opro}	-.195***	.056	.045	-.171***	.051	.060
λ_{probit}	.028	.146	.160	.965***	.297	.348
Constant	7.909***	1.312	1.573	5.233***	.897	1.267
ρ_{opro}	-.185			-.152		
ρ_{pro}	.026			.857		
p-value for overall significance	.000			.000		
Identification tests (p-value)						
H_0: Underidentified	.000			.000		
H_0: Not Overidentified				.064		
First stage R^2	.528			.560		
No of obs.	106			59		
No of cluster (village)	29			28		

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 29: Cassava production and consumption with sellers pooled

Dependent variable	ln(Production (kg))			ln(Consumption (kg))		
	$\hat{\beta}$	Std.err		$\hat{\beta}$	Std.err	
		Heckman	Original		Heckman	Original
ln(Price)	.200*	.109	.326	-.528***	.131	.295
Household size	.066**	.031	.024	.105***	.034	.036
Gender	-.310	.337	.478			
Education of head	.082***	.024	.024	.001	.029	.032
Total asset	.265	.205	.151	.281	.268	.288
Farm size	.112***	.036	.028	.144***	.044	.047
Total income (1000 US$)				.135**	.052	.058
Bicycle, own	-.080	.157	.194	-.025	.190	.210
Motorcycle, own	-.181	.190	.182	-.052	.236	.254
Car/truck, own	-.685	1.280	.964	-1.057	1.597	1.726
Distance to phone (km)	.007	.006	.008	.001	.006	.006
number of traders in village	.086	.073	.056			
Membership	-.226	.155	.168	-.349*	.191	.207
cotton grower	-.590	.471	.651			
maize grower	-.264	.352	.371	-.114	.477	.557
yams grower	.537*	.292	.342			
flour	-1.803***	.294	.590	-.024	.337	.584
dried tuber	-1.534***	.389	.477	.266	.429	.518
ln(consumption market (km))	-.095	.085	.133	-.024	.095	.102
distance to sales point	-.007	.010	.011	-.004	.013	.014
squared distance to sales point	.000	.000	.000	.000	.000	.000
dep2	1.755***	.190	.246			
dep6	1.276***	.392	.374			
λ_{opro}	-.029	.041	.043	-.105***	.041	.043
Constant	6.200***	.600	1.303	6.856***	.758	1.192
ρ_{opro}	-.062			-.097		
p-value for overall significance	.000			.000		
Identification tests (*p*-value)						
H_0: Underidentified	.032			.000		
H_0: Not Overidentified	.243			.058		
No of obs.	173			165		
No of cluster (village)	36					

Significance levels : $*$: 10% $**$: 5% $***$: 1%

C Correction for standard errors in the second-stage equation

C.1 The estimate of standard errors robust to the intra-cluster heterogeneity

The standard errors for estimation equations (4.3) and (4.4) must be corrected for two sources of heteroskedasticity. The first source of heteroskedasticity is associated with the fact that the survey was sampled using village as the cluster. The second source of heteroskedasticity is due to the addition of λ's into the equations.

The formula for robust variance-covariance matrix under cluster-sample differs from that of usual least squares estimates in the following way:

$$V_{OLS} = s^2 \left(X'X\right)^{-1} \tag{C.1}$$

$$V_{Cluster} = \left(X'X\right)^{-1} \left[\sum_{j=1}^{N_c} u_j^2\right] \left(X'X\right)^{-1} \quad (N_c : \text{\# of clusters}) \tag{C.2}$$

$$u_j^2 = \sum_{j-cluster} e_{ij} x_{ij} \tag{C.3}$$

in which e_{ij} is the residual for observation i in cluster j. The standard errors from variance-covariance matrix (C.2) are robust since they are less sensitive to the form of heteroskedasticity inside each cluster (village in this case) (?).

C.1.1 Variance of error term in second-stage equation

We have the following trivariate normal distribution.

$$\begin{pmatrix} v_1 \\ v_2 \\ u \end{pmatrix} \sim N\left(\begin{bmatrix} a_2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & \\ \rho & 1 & \\ \sigma_{1u} & \sigma_{2u} & \sigma_u \end{bmatrix} \right) \tag{C.4}$$

By modifying the equation (20) in Vijverberg (1995), we can express the relationship between the sample variance σ^2_{LS} in the Least Square estimates and σ_u as:

$$\begin{aligned} \sigma^2_{\text{LS}} = {} & \sigma^2_{1u} Z_1\gamma_1\lambda_1 + \sigma^2_{2u} Z_2\gamma_2\lambda_2 + (\sigma_{1u}\lambda_1 + \sigma_{2u}\lambda_2)^2 + \sigma_u \\ & + \frac{b(Z_1\gamma_1 - a_2, Z_2\gamma_2, \rho)}{\Phi(Z_1\gamma_1 - a_2)} (-\rho\sigma^2_{1u} + 2\sigma_{1u}\sigma_{2u} - \rho\sigma^2_{2u}) \end{aligned} \tag{C.5}$$

in which $\hat{\sigma}_{1u} = \hat{\delta}_{\lambda_1}$, $\hat{\sigma}_{2u} = \hat{\delta}_{\lambda_2}$ and $b(\cdot)$ is the bivariate normal density function for (v_1, v_2) evaluated at $(v_1 = Z_1\gamma_1, v_2 = Z_2\gamma_2)$.

Variance correction assuming uncorrelated decision makings

If the error terms for decision-makings on 2 criteria are uncorrelated, we have

$$\begin{pmatrix} v_1 \\ v_2 \\ u \end{pmatrix} \sim N\left(\begin{bmatrix} a_2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & \\ 0 & 1 & \\ \sigma_{1u} & \sigma_{2u} & \sigma_u \end{bmatrix} \right) \tag{C.6}$$

Then from Catsiapis and Robinson (1982) and Greene (1981), we can derive the correct variance-covariance matrix for the least square estimates.

$$\sigma_{u,i} = \hat{\sigma}_\varepsilon \left[(1-\rho_{12}^2-\rho_{13}^2) + \sum_{j=2}^{3} \rho_{1j}^2 (1 - Z_{ji}\gamma_{ji}\lambda_{ji} - \lambda_{ji}^2) \right] \quad (C.7)$$

$$\hat{\rho}_{1j}^2 = \frac{\hat{\beta}_{\lambda_j}^2}{\hat{\sigma}_\varepsilon^2} \quad (C.8)$$

Extending the analysis in Greene (2003, p.784), $\hat{\sigma}_\varepsilon$ can be calculated by,

$$\hat{\sigma}_\varepsilon = \hat{\sigma}_{ols} + \bar{\hat{\delta}}_{op}\beta_{\lambda_{op}}^2 + \bar{\hat{\delta}}_{pb}\beta_{\lambda_{pb}}^2 \quad (C.9)$$

in which

$$\bar{\hat{\delta}}_{op} = \frac{1}{N_{LS}} \sum_{i=1}^{N_{LS}} \delta_{op,i} = \frac{1}{N_{LS}} \sum_{i=1}^{N_{LS}} \hat{\lambda}_{op,i} \left[\hat{\lambda}_{op,i} - (x_{op,i}\hat{\gamma}_{op,i} - \hat{a}_2) \right] \quad (C.10)$$

$$\bar{\hat{\delta}}_{pb} = \frac{1}{N_{LS}} \sum_{i=1}^{N_{LS}} \delta_{pb,i} = \frac{1}{N_{LS}} \sum_{i=1}^{N_{LS}} \hat{\lambda}_{pb,i} \left[\hat{\lambda}_{pb,i} - x_{pb,i}\hat{\gamma}_{pb,i} \right] \quad (C.11)$$

Note $\bar{\hat{\delta}}_{op}$ and $\bar{\hat{\delta}}_{op}$ are the means only across sellers.

Then the variance-covariance matrix for the 3rd equation is,

$$\begin{pmatrix} \hat{\beta}_x - \beta_x \\ \hat{\beta}_{\lambda_j} - \beta_{\lambda_j} \end{pmatrix} \sim N(0, B'\Psi B) \quad (C.12)$$

where N_{LS} is the number of observations in the LS estimation,

$$B = \left[\begin{bmatrix} X_{LS}' \\ \lambda_j' \end{bmatrix} [X_{LS}, \lambda_j] \right]^{-1} = \begin{bmatrix} X_{LS}'X_{LS} & X_{LS}'\lambda_j \\ \lambda_j'X_{LS} & \lambda_j'\lambda_j \end{bmatrix}^{-1} \quad (C.13)$$

$$\Psi = \Psi_1 + \Psi_2 + \Psi_3 \quad (C.14)$$

$$\Psi_1 = \frac{1}{N_{LS}} \sum_{i=1}^{N_{LS}} \sigma_{u,i}^2 \begin{bmatrix} X_{LS}'X_{LS} & X_{LS}'\lambda_j \\ \lambda_j'X_{LS} & \lambda_j'\lambda_j \end{bmatrix} \quad (C.15)$$

$$\Psi_j(j=2,3) = \frac{1}{N_j N_{LS}} \sum_{i=1}^{N_{LS}} \sum_{\ell=1}^{N_{LS}} \sigma_{i\ell}(\lambda_j) \begin{bmatrix} X_{LS}'X_{LS} & X_{LS}'\lambda_j \\ \lambda_j'X_{LS} & \lambda_j'\lambda_j \end{bmatrix} \quad (C.16)$$

Applying Greene (1981), Ψ_1 can be expressed as,

Greene (1981) suggests the modified version of $\Psi_j(j=2,3)$ as,

$$\Psi_2 = \left(\sum_{i=1}^{N_{LS}} w_{op,i}' x_{op,i} \right) \text{Est.Asy.Var}[\hat{\gamma}_{op}] \left(\sum_{\ell=1}^{N_{LS}} x_{op,\ell}' w_{op,\ell} \right) \quad (C.17)$$

$$= \left(W_{op}' X_{op} \right) \text{Est.Asy.Var}[\hat{\gamma}_{op}] \left(X_{op}' W_{op} \right) \quad (C.18)$$

$$\Psi_3 = \left(\sum_{i=1}^{N_{LS}} w_{pb,i}' x_{pb,i} \right) \text{Est.Asy.Var}[\hat{\gamma}_{pb}] \left(\sum_{\ell=1}^{N_{LS}} x_{pb,\ell}' w_{pb,\ell} \right) \quad (C.19)$$

$$= \left(W_{pb}' X_{pb} \right) \text{Est.Asy.Var}[\hat{\gamma}_{pb}] \left(X_{pb}' W_{pb} \right) \quad (C.20)$$

in which

$$W'_{op} = \beta_{\lambda_{op}} \sqrt{N_{LS}/N_{op}} \begin{bmatrix} X'_{LS} \\ \lambda_j \end{bmatrix} \begin{bmatrix} \delta_{op,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \delta_{op,N_{LS}} \end{bmatrix} \rightarrow (k_{LS}+2) \times N_{LS} \tag{C.21}$$

X_{op} and X_{pb} both only include observations for sellers, and thus have $(N_{LS} \times k_{op})$ and $(N_{LS} \times k_{pb})$ dimensions respectively. Therefore Ψ_2 and Ψ_3 are both $(k_{LS}+2) \times (k_{LS}+2)$ with 2 being the number of selection criteria.