

# How Costly Is It for Poor Farmers to Lift Themselves Out of Subsistence? \*

Olivier Cadot<sup>†</sup>  
Laure Dutoit<sup>‡</sup>  
Marcelo Olarreaga<sup>§</sup>

## Abstract

The main objective of this paper is to provide estimates of the cost of moving out of subsistence for Madagascar's farmers. The analysis is based on a simple asset-return model of occupational choice. Estimates suggest that the entry (sunk) cost associated with moving out of subsistence can be quite large—somewhere between 124 and 153 percent of a subsistence farmer's annual production. Our results make it possible to identify farm characteristics likely to generate large gains if moved out of subsistence, yielding useful information for the targeting of trade-adjustment assistance programs.

JEL classification numbers: F10, O12, O19.

Keywords: threshold regression, switching regression, unknown sample separation, Madagascar, subsistence, entry costs.

---

\*This paper is part of a World Bank research project on Doha agricultural negotiations and rural poverty financed by the Bank-Netherlands Partnership Program. The results and interpretations are the authors' and do not necessarily reflect the World Bank's views. We wish to thank Monah Andriambalo, Caroline Freund, Marc Gurgand, Sylvie Lambert, Alessandro Nicita, Koroutoumo Ouattara, Guido Porto, seminar participants at DELTA, the University of Lausanne, The World Bank and the ETSG conference in Dublin for helpful comments and discussions.

<sup>†</sup>HEC Lausanne, CERDI and CEPR. Email: olivier.cadot@unil.ch.

<sup>‡</sup>HEC Lausanne. Email: laure.dutoit@unil.ch.

<sup>§</sup>Development Research Group, The World Bank, and CEPR. Email: molarreaga@worldbank.org.

# 1 Introduction

Putting to an end two decades of anti-market policies, in the mid-1990s a wave of reforms swept across least developed countries (LDCs), in particular Sub-Saharan Africa. Those reforms were intended, *inter alia*, to improve the efficiency of agricultural markets by eliminating failed marketing boards and to reduce the distortions affecting producer prices and incentives. Ten years on, however, it is fair to say that those reforms (some of which remain controversial; see e.g. Rodrik 1992, McMillan et al. 2002) have been slow to deliver the expected boost to growth and poverty alleviation. The feeble response of poor farmers to price signals is an old observation in development economics and has led to a debate on the rationality of farmers (see de Janvry et al. 1991 for a summary). That debate is now largely settled, if only by de Janvry et al.'s demonstration that a weak supply response to changes in market prices is to be expected in the presence of quantity constraints or "missing markets" in food or labor. Along similar lines, it has been recently argued that reforms affecting market prices can have only indirect effects on poor farmers who live on subsistence farming, out of reach of markets and market reforms. Clearly, the retreat from markets of large numbers of farmers was the old policies' result; however, once those policies had sufficiently damaged markets, the potential reach of reforms was reduced, making poverty persistent.

Interestingly, Brambilla and Porto (2005) have also shown that reforms can have non-monotone effects with an early phase of market disruption (their case study concerned Zambia's cotton reform, which initially led to widespread moral hazard) followed by a reorganization along a different market structure. Transitory shocks of that sort can send farmers momentarily into subsistence. With a combination of re-entry costs and other market failures such as the unavailability of credit, however, the retreat of farmers from the market may become permanent, implying potentially large welfare costs. It is thus crucial to understand how large these entry costs are and what individual characteristics they depend on.

A lot has already been said about entry costs in the theoretical IO literature. In his classic work, Bain (1956) defined them as any cost advantage that incumbents hold over entrants, even if recurring. Those could come from economies of scale, product differentiation, or

capital requirements. Baumol et al. (1981) defined entry costs as “*anything that requires an expenditure by a new entrant into an industry, but that imposes no equivalent cost upon an incumbent*”, stressing the importance of sunk costs. The presence of sunk costs was later on shown by Baldwin (1989), Baldwin et al. (1989) and Dixit (1989) to lead to hysteresis in market structure and trade flows,<sup>1</sup> something that is close to our earlier observation that entry barriers into commercial farming lead to hysteresis in poverty.

If the conceptual issues are clear, the search for empirical evidence on the presence of entry costs is still largely limited to high-tech sectors in industrial countries, where, according to Geroski (1995) “*econometric estimates of the height of entry barriers suggest that they are high*”. In the trade and development literature, the focus has been on the existence of barriers to export. Using Colombian manufacturing census data, Roberts et al. (1997) and Sanghamitra et al. (2001) found that sunk costs related to lack of exporting experience affected the aggregate export response, leading to hysteresis in trade flows. Sivadasan (2003) tested how reducing entry barriers through measures like trade and FDI liberalization would affect manufacturing productivity in India. Porto (2005) attacked the problem from a slightly different angle, looking at how removing government-related barriers to export (transport costs, custom practices, bureaucracy, regulation and corruption) would affect poverty in developing countries. He found those barriers to be a large 24.5% of the value of goods shipped, their removal having a non-negligible effect on poverty.

The type of informal barriers considered by Porto is likely to be relevant for entry into agricultural markets. Remoteness, whether resulting from geography or infrastructure decay, is likely to be a substantial barrier to entry into commercial farming, as is the lack of credit or intermediaries (sometimes from ethnic groups that were marginalized by nationalistic policies). Individual characteristics such as education and farm size are also likely to matter. The interaction between those factors and the presence of entry costs has however been largely untouched. One of the few papers approaching our concerns, Barrett (1997), looked into the effect of food marketing liberalization on trader entry in Madagascar. He showed

---

<sup>1</sup>See also Martin (2002) and Owen et al. (2002) for trade-related issues.

that liberalization of agricultural products marketing induced entry in subgroups of the food marketing channel, but not in groups where entry barriers were high. Vakis et al. (2003) measured the transaction costs occurring when a farmer sells his crops on the market. They showed that the level of transaction costs will determine which market (farmgate, local or distant market) farmers enter.

The literature's relative disregard of entry costs in and out of commercial agriculture is understandable given that those costs can be expected to be small in magnitude, as poor farmers are unlikely to switch altogether to sophisticated input-intensive farming. However what matters is not the absolute magnitude of entry costs but their magnitude relative to the income of subsistence farmers, in particular when credit markets are deficient so that sunk costs must be self-financed. Moreover, populations on the fringe of markets are likely to be vulnerable ones, and even well-wishing reforms can have the unwanted effect of destroying fragile markets, as changes in relative prices can be too small to cover switching costs or make intermediation profitable. When middlemen's trucks no longer come, rural roads are left to decay and entire regions can shrink into subsistence. A report by Oxfam (Oxfam-IDS 1999) argued that this is precisely what happened in Zambia after the maize reform (cited in Winters 2000).

This paper is an attempt to bridge the literature's gap. We propose a method to evaluate entry costs based on a simple asset-return model of farm specialization, the gist of which is as follows. Suppose that being in subsistence or in the market depends on a given household characteristic  $z$ . High- $z$  farmers are in the market whereas low- $z$  are in subsistence, with cutoff  $z^+$ . Each sector of activity can be thought of as an asset whose dividend is the net value of production (evaluated at market prices for subsistence farmers, leaving aside for the moment the issues involved in that valuation). The asset's price is the present discounted value of future dividends and expected capital gains. If the market asset is worth more than the subsistence one for some  $z < z^+$  (i.e. for a subsistence farmer), the difference provides a lower bound on the market entry cost for that farmer. For the marginal ( $z^+$ ) farmer, it is just equal to the entry cost.

Empirically, our strategy consists of estimating a profit equation for each activity (“profit” meaning the net value of production approximated using the data available in the household survey) using a switching-regression technique. This gives us two expected profits for each farmer  $i$ , conditional on  $i$ ’s individual characteristics: one under the “subsistence” regime and one under the “market” regime. For a subsistence farmer, the difference in predicted profits between the market regime (where he is not) and the subsistence regime (where he is) is the opportunity cost of not being in the market, which of course depends on his individual characteristics. The present discounted value of this estimated stream of profit differentials for the marginal farmer is then our estimated entry cost.

As the decision to be in the market or not is endogenous to the profit differential, endogenous switching-regression methods must be used (Maddala and Nelson 1975). An additional difficulty comes from the fact that there is no natural definition of “subsistence farming”, as the proportion of farm output sold on the market is, in the data, a continuous variable.<sup>2</sup> Rather than imposing an arbitrary cutoff in terms of proportion of output self-consumed, we let the data determine the best way of splitting the sample using a two-step procedure developed for switching-regression models with unknown switch point. Namely, in the first step we obtain consistent estimates of all parameters for a given value of the unknown cutoff by maximum likelihood, taking endogeneity into account; in the second step, we search for the cutoff value that yields the *maximum maximum* of the log-likelihood function.

The threshold obtained by this method turns out to be at zero. That is, the sample is split between farmers who are completely cut from markets (about 10% of them) and farmers who sell something. This accords with the intuitive idea that indivisibilities are involved (purchase of capital equipment, availability of intermediaries, transport cost, etc.). Once those hurdles are overcome, adjustment in terms of quantities sold is smoother. Estimated entry costs are large relative to income —between 124 and 153 percent of a subsistence farm’s annual output evaluated at market prices. Interestingly, however, the economic opportunity cost associated

---

<sup>2</sup>As defined by the Encyclopedia Britannica, “subsistence farming is a form of farming in which nearly all of the crops or livestock raised are used to maintain the farmer and his family, leaving little, if any, surplus for sale or trade.” However, most farmers sell some and keep some for home consumption.

with this market failure (i.e. with the fact that subsistence farms with potentially higher returns in the market do not move because of entry costs) adds up to a mere 0.46 percent of GDP, the reason being very low productivity in both sectors. These results, however preliminary, go in the general direction of the literature on trade and poverty, making it difficult to believe that the trade reforms of the 1980s and 1990s were the culprit for the slide into subsistence and poverty of large numbers of Sub-Saharan farmers.

## 2 Model

Suppose that farmers can decide whether to participate in the market or not; if not, their farms are called “subsistence” ones. The numbers of farms in each sector are  $S$  and  $M$  respectively, both functions of time (argument omitted). Time is continuous and the horizon is infinite. Let  $\Pi_j$  ( $j = S, M$ ) be the instantaneous flow of return to farming in sector  $j$ ; in the subsistence sector, it is measured by valuing self-consumed crops at market prices. Let also  $z$  be a time-invariant individual characteristic affecting profits (for now taken as a scalar, although it will be multidimensional in the empirical part), so  $\Pi_j = \Pi_j(z)$ . Note that  $\Pi_M$  and  $\Pi_S$  are distinct functions, although mapping the same argument ( $z$ ) into profits. We will suppose that

$$\frac{\partial \Pi_M}{\partial z} > \frac{\partial \Pi_S}{\partial z} \quad \forall z. \quad (1)$$

This single-crossing condition will, in equilibrium, lead to partial sorting of farms between subsistence and market as a function of  $z$ .

Farmers in  $M$  are periodically thrown into  $S$  by idiosyncratic events (say, the household head falls sick) happening at a constant and exogenous hazard rate  $s$ .<sup>3</sup> Once thrown back into subsistence, farmers can either jump back to the market by incurring a sunk cost  $c$  (common to all) or stay in subsistence. We assume for simplicity that there is no lag between the event

---

<sup>3</sup>Alternatively, one could think of the shock as affecting the terms of trade in a way that makes commercial agriculture less attractive. Such shocks would however be aggregate (correlated across farmers) and would complicate the dynamics without much additional insight.

of being thrown into subsistence and the opportunity to get back to the market.

Given (1), the individual-characteristic variable  $z$  will sort farms in  $S$  between those that decide to move back into the market (high- $z$  ones relative to the rest of the  $S$  sub-population) and those that don't (low- $z$  ones). We will call  $z^-$  the critical value of  $z$  that separates these two sub-populations and  $z^+$  the highest value of  $z$  in  $S$ . Thus, we have in steady-state three sub-populations: market farmers, subsistence farmers who are constrained to stay there by the sunk cost needed to get out, and subsistence farmers who do better by staying in subsistence. For the latter, sorting is completely voluntary and reflects comparative advantage, as in Roy's hunters vs. fishermen story (Roy 1951).

Let  $p$  be the proportion of farmers in  $S$  who would switch were it not for the switching cost;  $p$  and the steady-state values of  $M$  and  $S$  are endogenous, reflecting differential rates of return between market participation and subsistence farming (like in migration models where migration is driven by earnings differentials) whereas  $s$  is exogenous.

Let  $N = M + S$  be the total number of farms, kept fixed throughout. Using hats to denote proportional rates of changes ( $\widehat{S} = dS/S$ ), the evolution of stock variables  $S$  and  $M$  is determined by entry and exit rates and adding up:

$$\widehat{S} = sM - pS. \quad (2)$$

In steady state,  $\widehat{S} = 0$ , so equation (2) becomes

$$M^* = \frac{pS^*}{s} = \frac{p}{s}(N - M^*) \quad (3)$$

or

$$M^* = \left( \frac{p}{s+p} \right) N. \quad (4)$$

The stock of farms in subsistence is

$$S^* = \left( \frac{s}{s+p} \right) N$$

and the frequencies of each of the three subpopulations are respectively  $p/(s+p)$  for market farms ( $z > z^+$ ),  $sp/(s+p)$  for constrained subsistence farms ( $z^- < z < z^+$ ) and  $s(1-p)/(s+p)$  for non-constrained ones ( $z < z^-$ ). Note that, in steady state, all farmers with  $z^- < z < z^+$  have already been thrown out of the market, so only high  $z$  remain there.

Let  $r$  be the interest rate.<sup>4</sup> Let also  $V_M$  and  $V_S$  be value functions corresponding to the present discounted value of future profit streams and capital gains in each sector, given an infinite horizon. All value functions are evaluated at the system's steady state. Farming in sector  $j$  can be thought of as holding an asset whose price is determined by a standard no-arbitrage condition. For a market farm,

$$rV_M = \Pi_M(z) - sc. \quad (5)$$

In (5), the first term on the RHS is a “dividend” term—the instantaneous flow of profits—whereas the second one is an expected capital-loss term equal to the product of the re-entry cost  $c$  by the exit's hazard rate  $s$ . The simplicity of the capital-loss term is due to the assumption of immediate re-entry. One could easily add a waiting time to the story, but the end formula for the entry cost would be unchanged. For a subsistence farm, the expression is even simpler:

$$rV_S = \Pi_S(z). \quad (6)$$

Inequalities between these value functions determine three segments of the farm population. For the lowest- $z$  segment ( $z < z^-$ ),

$$V_M < V_S;$$

that is, sorting is voluntary. For the intermediate- $z$  segment ( $z^- < z < z^+$ ),

$$V_M - c < V_S < V_M;$$

---

<sup>4</sup>With perfect credit markets, the interest rate used in the calculation of NPVs should be the economywide interest rate. However, under credit constraints, it should be instead the shadow interest rate defined by the opportunity cost of funds for farmers.

that is, sorting is involuntary: farmers in that segment would rather be in the market ( $V_S < V_M$ ) but can't because of the entry cost ( $V_M - c < V_S$ ). Finally, high- $z$  farmers voluntarily sort themselves in the market, because

$$V_S < V_M - c.$$

Observationally, farmers in the first and second segments are indistinguishable and are characterized by the same expected return conditional on their individual characteristics, as the function mapping characteristics into farming return depends only on the actual sector of activity (market or subsistence). Therefore, in terms of the relationship between characteristics and returns, there is only one switchpoint at  $z^+$ . For the marginal farm ( $z = z^+$ ),

$$V_M(z) - c = V_S(z).$$

Using (5) and (6), this gives immediately

$$c = \frac{\Delta\Pi}{r + s} \tag{7}$$

where  $\Delta\Pi = \Pi_M - \Pi_S$ . That is, the entry cost is equal to the net present value (discounted at a rate reflecting both the cost of funds and the probability of having to re-incur it) of the profit differential between the two sectors. This simple expression will guide our empirical exploration. Note that, were the assumption of risk-neutrality to be relaxed, the returns differential might as well include a risk premium. This might be important empirically if there was some reason to believe that market farming is inherently riskier than subsistence farming. (This would be the case, for instance, if producer prices were volatile and negatively correlated with consumer prices, implying volatility in real income.) To our knowledge, there is no evidence to back up such conjecture. In the absence of a way of assessing the farmers' degree of risk aversion and hence of deriving a measurable proxy for the risk premium, there is no real point in modeling it explicitly. The reader should however keep

this point in mind as a caveat.<sup>5</sup>

### 3 Estimation

We now turn to an attempt at estimating  $c$  following the logic of expression (7). There are two difficulties. First, as the model’s logic rests on a comparison of financial returns, we must find a way of valuing the return to subsistence activity. We do so in the simplest possible way, by using market prices to value output. However we do not have farm accounting data to assess profits as precisely as if we used corporate balance sheets. In principle, one would want to measure profits as the difference between output value (actual sales plus self-consumed output valued at producer prices) and intermediate consumption. However farmers typically report their income with considerable error, and cost data is non-existent. We go round these difficulties by taking consumption expenditures (reported much more accurately than income in household surveys) net of other (non-agricultural) revenue.<sup>6</sup>

Second, farms must be classified as “subsistence” or “market” ones on the basis of a reasonable criterion. The most natural one is the proportion of farm output that is sold vs. self-consumed. However this settles only half the issue, because the cutoff must then be set somewhere.<sup>7</sup> We choose, as explained in section 3.2 below, to let the data generate the cutoff value endogenously.

#### 3.1 Data

The data is from Madagascar’s four household surveys (1993, 1997, 1999 and 2001), called in French “*Enquête Permanente des Ménages*” (EPM). Each is made of a household survey per se covering about 5,200 households on average and a “community questionnaire” collecting

---

<sup>5</sup>We are grateful to an anonymous referee of the World Bank’s Policy Research Working Papers for attracting our attention to this issue.

<sup>6</sup>We actually also estimated our model using the output value only, assuming zero cost (since the data on costs is not available). The results we found are similar to the results we get with our measure of profits (consumption expenditures-non-agricultural revenue).

<sup>7</sup>See Hotchkiss (1991) for a similar discussion of part-time vs. full-time work.

data on villages. The EPM uses multi-stage sampling, splitting observations into strata and clusters. Strata are identified by province (“Faritany”) and by urban vs. rural area; clusters are identified by administrative units (“Fokontany”) randomly drawn in each stratum. The sample in this paper includes 5,951 agricultural households. Although repeated, the EPM is not a panel as only a small number of households were sampled repeatedly in 1997 and 1999.

Commune data (collected in 2001 and covering 1’385 communities at the level of administrative units above Fokontany) is from the commune census of Cornell University’s ILO program. It includes, inter alia, data on the commune’s level of development, infrastructure and aggregate agricultural production.

Producer prices are taken from the Food and Agriculture Organization (FAO), for all years between 1991 and 2001. Prices are in current Malagasy francs (MGF). Climatic data was taken on the one hand from the Ministry of Agriculture, Livestock and Fisheries of Madagascar and on the other hand from a paper on climatic perturbations (Rambeloalijaona & Randrianarivelo, 2003).

The first column of Table 1<sup>8</sup> shows descriptive statistics computed allowing for the sample design (weights, strata and clusters<sup>9</sup>). The first striking feature of the data is agricultural household’s extreme poverty.<sup>10</sup> With a yearly income of US \$566 for a family of five (implying about 30 cents per day per person) the average Malagasy household lives well under the international dollar-a-day poverty line.

The second is the degree of farm specialization. On average, a farm’s first crop accounts for 80 percent of agricultural income, and about 70 percent of total income comes from crop sales. With 7.5 percent of full income<sup>11</sup> from livestock sales, the diversification of sources of income is somewhat limited.

At about 1.6 hectares, the average Malagasy farm is very small, which means that all but the simplest capital equipment is prohibitively costly to buy at the farm level, and hence

---

<sup>8</sup>The second and third columns will be commented on later.

<sup>9</sup>For more on that, see Deaton (1997).

<sup>10</sup>A thorough analysis of poverty trends in Madagascar can be found in Paternostro et al. (2001).

<sup>11</sup>By “full” income we mean that our measure of income includes the value of the total harvest (evaluated at producer prices) and not only the value of what is sold in the market.

that labor productivity can only be very low.

Land ownership is widespread at about 80 percent of all land.<sup>12</sup> Of that part of land that is not owned, sharecropping accounts for 4% and tenancy for 3.6%. The interest in the distinction between sharecropping and tenancy is that under the former, the farmer owes a fixed percentage of the harvest to the landowner, which provides partial insurance through risk-sharing between the farmer and landowner. Tenants, by contrast, must pay a fixed rent, implying that they bear the entire burden of crop-variability risk. One may therefore expect more risk-taking under sharecropping than under tenancy.<sup>13</sup>

Finally, if all crops, sold or self-consumed, are evaluated at producer prices, potatoes and rice contribute respectively 55% and 26% of “full” agricultural revenue. In fact, only 26 percent of the potato harvest is marketed vs. about 60 percent for rice. In terms of cash revenue, rice contributes about 36 percent of the total. Overall, about two thirds of farm output evaluated at market prices is self-consumed.<sup>14</sup>

Figure 1 shows the evolution of the prices of Madagascar’s main crops over 1991-2001, after normalization at 100 in 1991 to improve comparability. The figure shows high price volatility. Cloves, vanilla and sugar cane, in particular, underwent large price swings (from 100 to over 650 for cloves, to slightly less than 800 for vanilla, with high variability in-between, and to about 900 for cane sugar). Other prices are somewhat less volatile but oscillate between index values 50 and 200. Notice that except for sugar cane, crop that experienced high price volatility are long-term crops, that is, crops that need more than three years to reach maturity.<sup>15</sup> This high price variability over the sample period allows us to simulate the long-run response of farmers to price changes on the basis of observed past behavior.

---

<sup>12</sup>The 15 percent of lands that are neither owned, nor in sharecropping, nor in tenancy are either lent to the household without rent or given by the village to develop it.

<sup>13</sup>Sadoulet et al. (1993) discuss the efficiencies or inefficiencies of sharecropping contracts and show that under extreme poverty or gift exchanges, contract are in fact efficient.

<sup>14</sup>This sticks to Barrett’s (1997) number, where he says that in the province of Antananarivo, only 25-30% of agricultural production is marketed.

<sup>15</sup>See Table 3 for the classification of crops.

## 3.2 Method

Following the logic of section 2, we assume that the function mapping individual characteristics into the return from farming is sector-specific, i.e. is not the same for market vs. subsistence farms. For instance, transportation costs, access to credit and inputs, or any factors affecting the output's quality are likely to matter more for commercial than for subsistence agriculture.<sup>16</sup> This calls for the use of a switching-regression framework. As per the previous section's model, however, we assume that rational farmers base their decision to "migrate" from subsistence to market on a comparison of returns, implying that the switchpoint is endogenous.

Let  $\Pi_{ji}$ ,  $j = M, S$ , be farmer  $i$ 's "profits" (precise variable definitions are discussed below) in sector  $j$ , and let  $\pi_{ji} = \ln \Pi_{ji}$ . Let  $X_i$  be a vector of individual characteristics affecting profits in both sectors (in general, these RHS variables need not be the same in both regimes, but here they are). The model is

$$\pi_{Mi} = X_i' \beta_M + u_{Mi}, \quad (8)$$

$$\pi_{Si} = X_i' \beta_S + u_{Si}, \quad (9)$$

$$I_i^* = (\pi_{Mi} - \pi_{Si}) \delta + Z_i \gamma - v_i \quad (10)$$

where  $u_{Mi}$ ,  $u_{Si}$  and  $v_i$  are error terms. In (8)-(10),  $\pi_{Mi}$ ,  $\pi_{Si}$  and  $I_i^*$  are latent (unobserved) variables; the *observed* level of (log) profits is

$$\pi_i = \begin{cases} \pi_{Mi} & \text{if } I_i^* > 0 \\ \pi_{Si} & \text{if } I_i^* \leq 0. \end{cases}$$

The presence of  $\pi_{Mi}$  and  $\pi_{Si}$  on the RHS of (10) is what makes the switching endogenous, whereas the assumption that the entry cost is the same for all households ensures that (10)

---

<sup>16</sup>For evidence on productivity differentials between subsistence farmers and market participants, see Brambilla and Porto (2005).

holds, the entry cost being subsumed in the equation's constant term.<sup>17</sup> Substituting from (8) and (9), we can write

$$\begin{aligned} I_i^* &= X_i'(\beta_M - \beta_S)\delta + Z_i'\gamma - \varepsilon_i \\ &= W_i'\alpha - \varepsilon_i \end{aligned} \tag{11}$$

where  $W_i' = [X_i', Z_i']$ , and

$$\varepsilon_i = (u_{Si} - u_{Mi})\delta + v_i.$$

Let  $\Sigma$  be the variance-covariance matrix of the error terms  $u_{Si}$ ,  $u_{Mi}$ , and  $\varepsilon_i$ , with the following (standard) assumptions:

**A1**  $\sigma_{MS} = 0$ ;

**A2**  $\sigma_{j\varepsilon} \neq 0$ ,  $j = M, S$ ;

**A3**  $\sigma_\varepsilon^2 = 1$ .

A1 and A3 are identification assumptions discussed in Lee (1979), whereas A2 reflects the switching decision's endogeneity. Given these, we can write

$$\Sigma = \begin{pmatrix} \sigma_M^2 & & \\ 0 & \sigma_S^2 & \\ \sigma_{M\varepsilon} & \sigma_{S\varepsilon} & 1 \end{pmatrix}. \tag{12}$$

Let  $f(\pi_i | \pi_{Mi}, \pi_{Si}, I_i^*)$  be the density function of  $\pi_i$  conditional on  $\pi_{Mi}$ ,  $\pi_{Si}$ , and  $I_i^*$ ,  $f^j(\pi_{ji})$  the density of  $\pi_{ji}$  conditional on regime  $j$ , and

$$q_i = \begin{cases} 1 & \text{if } I_i^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

---

<sup>17</sup>A variable entry cost cannot be identified out of cross-section data, although we provide in section 3.3.4 below a *very tentative* approach to understand how the entry cost might actually vary across households.

Let also  $q_i^* = \text{prob}(q_i = 1) = \text{prob}(I_i^* > 0)$  be the probability that observation  $i$  belongs to regime  $M$ ; this probability is unknown because the cutoff that determines it is itself unknown (a cutoff assigning 10% of the population to the  $M$  regime clearly would not give the same probability as one assigning it 90% of the population). Following Hartley (1978), we can write

$$f(\pi_i | \pi_{Mi}, \pi_{Si}, I_i^*) = \begin{cases} f^M(\pi_{Mi}) & \text{if } I_i^* > 0, \\ f^S(\pi_{Si}) & \text{otherwise} \end{cases} \quad (13)$$

and

$$f(\pi_i | \pi_{Mi}, \pi_{Si}) = q_i^* f^M(\pi_{Mi}) + (1 - q_i^*) f^S(\pi_{Si}). \quad (14)$$

Letting  $h(\pi_i)$  stand for the mixture density on the RHS of (14), the likelihood function corresponding to problem (8)-(10) is

$$L = \prod_{i=1}^n h(\pi_i)$$

and consistent estimates of all parameters can, in principle, be obtained by

$$\max \mathcal{L} \equiv \ln(L) = \sum_{i=1}^n \ln[h(\pi_i)]. \quad (15)$$

The difficulties involved in the application of MLE to (15) are discussed in Quandt and Ramsey (1978) and in the comments published in the same issue of the JASA.<sup>18</sup> Since then, a number of multi-step procedures have been devised to find consistent estimates for the model's parameters, including the unknown cutoff (see e.g. Hansen 2000). We follow here a procedure set up by Hotchkiss (1991) to deal with situations involving both selectivity and an unknown cutoff.

In order to describe the procedure in some detail, we introduce some more notation. Let  $\phi$  and  $\Phi$  denote respectively the density and CDF of the standard normal distribution. We

---

<sup>18</sup>Most of the issues are discussed in Maddala (1983).

write the inverse Mills ratio as

$$\varphi_j(\cdot) = \begin{cases} \phi(\cdot)/\Phi(\cdot) & j = M \\ -\phi(\cdot)/[1 - \Phi(\cdot)] & j = S, \end{cases}$$

with the minus sign in the bottom expression to facilitate the use of the regime index  $j$  later on. For future use, let also  $\rho_j$  be the coefficient of correlation between  $u_{ji}$  and  $\epsilon_i$ , and

$$\eta_{ji} = \left[ \frac{W_i' \alpha - \rho_j (\pi_{ji} - X_i' \beta_j) / \sigma_j}{\sqrt{1 - \rho_j^2}} \right], \quad j = M, S. \quad (16)$$

Lee (1979) showed that

$$E(u_{Mi}^2 | I_i^* > 0) = \sigma_M^2 - \sigma_{M\epsilon}^2 W_i' \alpha \varphi_M(W_i' \alpha)$$

and

$$E(u_{Si}^2 | I_i^* \leq 0) = \sigma_M^2 - \sigma_{S\epsilon}^2 W_i' \alpha \varphi_S(W_i' \alpha).$$

Thus, letting  $\xi_{ji}$  and  $\mu_{ji}$  be error terms and using  $\varphi_{ji}$  as shorthand for  $\varphi_j(W_i' \alpha)$ , we can write

$$\pi_{ji} = X_i' \beta_j - \varphi_{ji} \sigma_{j\epsilon} + \xi_{ji}, \quad j = M, S \quad (17)$$

and

$$u_{ji}^2 = \sigma_j^2 - W_i' \alpha \varphi_{ji} \sigma_{j\epsilon}^2 + \mu_{ji}, \quad j = M, S. \quad (18)$$

Using this, steps 1 and 2 of our procedure, which follow Lee (1979), provide consistent estimates which we use as initial values for the ML estimation of step 3. From then on we follow Hotchkiss's procedure (Hotchkiss 1991), which consists of estimating the model's parameters by maximum likelihood for a given (endogenous) cutoff and then searching for the cutoff yielding the *maximum maximorum* of the log-likelihood function, re-optimizing all parameters at each step:<sup>19</sup>

---

<sup>19</sup>Hansen (2000) proposes a somewhat similar method for situations without selectivity; however his first

**Step 1** An initial (arbitrary) cutoff  $\lambda^*$  is set in terms of the share of a farm's output that is sold on the market (an observable variable). Farms are sorted in either regime using this cutoff and a probit is run on (11), yielding an estimate of  $\alpha$ ,  $\hat{\alpha}$ . Let  $\hat{\varphi}_{ji} = \varphi_j(W_i'\hat{\alpha})$ .

**Step 2** Estimates  $\hat{\varphi}_{ji}$  from the first-stage probit are used in OLS regressions of (17), yielding estimates  $\hat{\beta}_j$  and  $\hat{\sigma}_{j\epsilon}$ . Plugging the former into (8)-(9) yields a vector of residuals  $\hat{u}_{ji}$ . Using  $\hat{u}_{ji}^2$ ,  $\hat{\sigma}_{j\epsilon}^2$ ,  $\hat{\alpha}$  and  $\hat{\varphi}_{ji}$  in lieu of  $u_{ji}^2$ ,  $\sigma_{j\epsilon}^2$ ,  $\alpha$  and  $\varphi_{ji}$  respectively in an OLS regression of (18) finally gives an estimate of  $\sigma_j^2$ .

**Step 3** All parameters being consistently estimated, they are used as initial values for the maximization of

$$\begin{aligned} \ln \tilde{\mathcal{L}} = & \sum_{i=1}^n \left\{ q_i \left[ \ln \phi \left( \frac{\pi_{Mi} - X_i'\beta_M}{\sigma_M} \right) - \ln \sigma_M + \ln \Phi(\eta_{Mi}) \right] \right. \\ & \left. + (1 - q_i) \left[ \ln \phi \left( \frac{\pi_{Si} - X_i'\beta_S}{\sigma_S} \right) - \ln \sigma_S + \ln [1 - \Phi(\eta_{Si})] \right] \right\} \end{aligned} \quad (19)$$

where  $\eta_{ji}$  is defined in (16). With this method, every parameter of (8), (9) and (10) is estimated consistently and asymptotically efficiently. The value of the maximum likelihood is recorded.<sup>20</sup>

**Step 4** The procedure starts again from Step 1 with a different value of  $\lambda^*$  and is repeated until the *maximum-maximorum* of the log-likelihood function is found.

---

step is a set of OLS regressions, each for a given threshold value, the second being the minimization of the sum of squared residuals by choice of the threshold. Note that Hotchkiss's procedure differs from an early one proposed by Quandt (1958) and Silber (1974), in which the first step maximizes the likelihood function and the second step searches for the best cutoff value *holding all other parameters constant*.

<sup>20</sup>The maximization was performed in Stata (Lokshin and Sajaia 2004). Note that Lokshin and Sajaia's procedure allows for weights and clusters, but not for strata. However, as omitting strata generally biases upward the estimated standard errors, this is working against us.

**Step 5** Finally, the entry cost faced by subsistence farmers willing to switch to the market is estimated by

$$\begin{aligned}\hat{c} &= \frac{1}{r+s} \Delta E \left( \Pi_{ji} \left| \widetilde{W}'_i \right. \right) \\ &= \frac{1}{r+s} \left\{ \exp \left[ \widetilde{X}' \widehat{\beta}_M - \widehat{\sigma}_{M\epsilon} \varphi_M(\widetilde{W} \widehat{\alpha}) \right] - \exp \left[ \widetilde{X}' \widehat{\beta}_S + \widehat{\sigma}_{S\epsilon} \varphi_S(\widetilde{W} \widehat{\alpha}) \right] \right\}\end{aligned}\quad (20)$$

where  $\widetilde{X}$  and  $\widetilde{W}$  are evaluated at the marginal farm (the one with the highest predicted probability of being in the market while actually being in subsistence), all parameter estimates are based on the “best” sample split, and suitable proxies (discussed below) are used for the interest rate  $r$  and exit rate  $s$ .

### 3.3 Results

#### 3.3.1 Endogenous classification of farms

The log likelihood was maximized at  $\lambda^* = 0$  (no market participation at all). This splits our sample into 449 households in subsistence and 5502 on the market (equivalent to 0.3 and 3.8 million households respectively). For robustness, a bootstrap procedure was applied to  $\lambda^*$ : 60 samples were drawn from the sample with replacement; for each of these samples, the model was re-estimated and the value of  $\lambda^*$  maximizing the log likelihood was retrieved. The mean value of  $\lambda^*$  over the bootstrap was 0.00256, with a standard error of 2.26E-04. The corner solution ( $\lambda^* = 0$ ) therefore seems to be a good estimate of the true value of  $\lambda^*$ .

As a further check on how reasonable it is to set  $\lambda^*$  at zero, the last two columns of Table 1 report descriptive statistics conditional on sectoral sub-samples constructed using this value. As for gender, the proportion of female household heads is 18% in subsistence vs. 14% in the market, consistent with the common observation that female-headed households are also often poorer. Average age is slightly higher in subsistence, but education shows no difference.

Cropland size is 64% smaller in subsistence than in the market, and a slightly lower proportion (76% vs 80%) is owned. Access to credit is also lower (2.66% of farms in subsistence

vs. 4.26% in the market) although both proportions are extremely low, suggesting largely inexistent rural credit markets. Counterintuitively, market farmers are more remote than subsistence ones, but export crops are known to be grown in relatively isolated areas in Madagascar (see Stifel et al., 2003).

In terms of “full income”, a market farm’s main crop accounts for about 56 percent of the total. In terms of cash income, the proportion is 80%, the difference coming from the large proportion that is self-consumed even by market farms (see supra). If market farms seem more specialized in terms of cash crops than food ones, interestingly subsistence farms are also heavily specialized with the first crop accounting for 70% of the total evaluated at market prices (the top two crops account for a whopping 90% of the total). In other words, subsistence farmers mostly rely on two crops only.

The bottom of the table shows that whereas farmers on the market produce mainly potatoes and rice, subsistence farmers mainly produce rice and cassava, which accords with evidence that rice and cassava are Madagascar’s most important food crops.<sup>21</sup>

### 3.3.2 Profit equations, selection equation and entry costs

The first column in Table 4 shows results for the profit equation of market farms<sup>22</sup>, the second for that of subsistence farms (recall that we are talking here of full income, including self-consumption valued at producer prices), and the third for the selection equation, that is, for equation (10). The dependent variable of the selection regression, “occupation”, is equal to one for a market farm and zero otherwise.

Notice that for the model to be identified, the selection equation must include additional explanatory variables that explain the regime decision only. This is why the variables *changes in past prices*, *number of rich*, *number of poor*, *remoteness index* and *livestock* are part of equation (10) and not of profit equations. Changes in past prices, whether through price index or price volatility, indeed affect the decision to participate in subsistence or in the

---

<sup>21</sup>The main food crops of Madagascar are rice, maize, cassava, sweet potatoes and groundnuts and the most important cash crops are coffee, vanilla, cloves, sugarcane, cotton and cocoa.

<sup>22</sup>The definition of the explanatory variables is in Table 2.

market: if a farmer's past prices follow a negative slope, he will change his cropping scheme and this will have an impact on his participation decision. However, there is no evidence that past prices should influence present prices. These two variables are therefore not part of the two profit equations. Finally, *remoteness*, *number of rich*, *number of poor* and *livestock* were only included in the selection equation even if they could also influence profits. We exclude them from the profit equations because they were never significant.

Interestingly, household characteristics have a significant effect on profits in both cases (market and subsistence) but less on occupational choice. Unsurprisingly, the household head's schooling has a positive impact on profit in both cases. Age affects subsistence profits along an inverted U-shape but does not affect market profits.

The presence or size of factors of production (cropland size, loan, inputs, fertilizer use and steer) in general have a positive effect on market profits and on subsistence profits; as for occupational choice, larger farms in terms of acreage have a higher probability of being in the market. Ownership structure seems to affect profits only in the market sector, with a positive effect of tenancy. Quite surprisingly, climatic variables only show clear effect on the occupational choice.

Remoteness (national road, transport cost and remoteness index) has the overall expected effect, as the presence of a national road in the village has a positive effect on market profits. Moreover the remoteness index increases the probability of being on the market, as already noted in our discussion of descriptive statistics.

Finally and most interestingly, the evolution of prices as measured by our composite price indices affects occupational choice in ways that accord with intuition.<sup>23</sup> The more a farm's producer prices increased in the past, the greater its probability of being on the market. Conversely, the greater the volatility of those prices, the smaller this probability.

In sum, household characteristics, production factors, community characteristics and, most importantly, prices, generally affect profits and occupational choices in expected ways.

---

<sup>23</sup>The price index and volatility were not included in the profit equations, since these variables reflect the movement of past prices. It is difficult to imagine that past prices can affect present profits. For explanations on the construction of the price index and price volatility variables, see appendix A.3.

A last comment concerns the statistics indicated at the bottom of the table.  $Sigma_1$ , and  $Sigma_2$  are the square-roots of the variance of the residuals of equations (8) and (9) and  $rho_1$  and  $rho_2$  are the correlation coefficients.<sup>24</sup> Only  $rho_1$  (the correlation coefficient between occupational choice and the market profit equation) is significant. Since it is positive, a farmer on the market has higher profits than those of a “random” one.

Table 6 provides the mean conditional profits under the two regimes:  $E[\pi_{Mi}|I_i^* \leq 0]$ ,  $E[\pi_{Si}|I_i^* \leq 0]$ ,  $E[\pi_{Mi}|I_i^* > 0]$  and  $E[\pi_{Si}|I_i^* > 0]$ , for different values of  $\lambda^*$ . Subsistence farmers’ average agricultural profits is 30% lower than the profit of market farmers. This can be explained by higher productivity, access to better inputs, etc. Controlling for their characteristics, the switch from subsistence to market participation could increase subsistence farmers income by 36 percent. Similarly, those farmers out of subsistence could see their income fall by half if they were to move into subsistence.

In order to compute the entry costs going along with a move from subsistence towards the market (equation (7)), we must approximate  $r$ , the interest rate, and  $s$ , the probability of exit from the market sector. To be as close as possible to reality, we proxy  $r$  with the interest rate on microcredit. Yearly microcredit rates in Madagascar typically hover between 27% and 43%. An interest rate of 30% therefore seems to be a good approximation. As for  $s$ , it was calculated using the survey’s two-year panel<sup>25</sup> by calculating the proportion of 1997 market farms that switched to subsistence between 1997 and 1999. This proportion was then adjusted to find a yearly percentage of farmers switching into subsistence. We find a value of  $s$  of 4.08 percent. From equation (7), the entry cost was computed by dividing the opportunity cost by the sum of  $r$  and  $s$ , that is, by 0.3408.

Using these, our estimate of the entry cost is 536’385 Malagasy francs i.e. 139% of subsistence farmers agricultural profits. The lower bound for the entry cost estimated with the 95% confidence interval is 481’496 Malagasy francs, or 124% of subsistence farmers’ profits. The upper-bound estimate amounts to 591’275 Malagasy francs, i.e. 153% of profits.

---

<sup>24</sup>Since  $rho_1$  and  $rho_2$  are not equal to zero, it means that  $\sigma_{1\epsilon}$  and  $\sigma_{2\epsilon}$  cannot take the value zero. Recalling the comment of page 5 on these two variables, this means that the switch is indeed endogenous.

<sup>25</sup>The data contain a panel over about 1000 households, surveyed in 1997 and 1999.

The annual economic opportunity cost of subsistence farming is obtained by taking the mean of  $E[\Delta\pi_{ji}|I_i = 0]$  over subsistence households that are not moving to the market in each year.<sup>26</sup> The estimate of the opportunity cost is 167'800 Malagasy francs, or 43 percent of profits. Dividing the annual opportunity cost by GDP and taking the weighted sum, we have that moving all subsistence farmers into the market would add an annual 0.46 percent to Madagascar's GDP. Entry costs and the consequent subsistence of farmers can be very large when measured at the level of the individual facing the cost. However, at the aggregate level, the costs do not seem to be very large, echoing the literature on trade reform and adjustment costs (see Matusz and Tarr, 1999).

### 3.3.3 Robustness

For robustness, Tables 5a-5c provide first-stage ML profit estimates for values of  $\lambda^*$  set at 10, 20 and 30 percent respectively, allowing for some market sales by subsistence farmers. Unsurprisingly, the larger is  $\lambda^*$ , the more different are the regression results. Few changes are observed on the two profit equations, but results seem less robust for the occupational choice equation. Note from Table 6 that, as we raise the value of  $\lambda^*$ , the entry cost associated with subsistence farming declines.

Table 5d provides profit estimates for  $\lambda^* = 0$  but with a decomposition of price variables into short-term crops and long-term crops, depending on the length of time it takes for them to grow.<sup>27</sup> While long term price index and volatility variables are significant, only the short term volatility variable is significant. All other coefficients appear to be in line with those reported in Table 4.

A last robustness check is provided in Table 5e. In order to test for production aggregation problems, the switching regression model is estimated on a subsample composed of households producing mainly rice,<sup>28</sup> Madagascar's most common crop. With 84% of Malagasy

---

<sup>26</sup>That way, we estimate accurately the opportunity cost, since we only take into account farmers that are really facing the cost.

<sup>27</sup>For the classification of the crops, see table 3.

<sup>28</sup>Since in most of the case, households have two important crops (e.g, she mainly produces vanilla for exports, but also a large amount of rice for self consumption), we considered both of them as main crops.

households producing rice as one of their first two crops, we have a large enough subsample (4996 observations). Regression results on rice are shown in Table 5e. As there is little change between whole-sample and rice-sample estimates, crop aggregation does not seem to have a critical impact on our conclusions.

Table 6a provides the rice subsample’s conditional profits for different values of  $\lambda^*$ . It shows that both subsistence and market profits are slightly higher in the subsample than in the full sample. Rice thus seems to be more profitable than the “average” crop. The entry cost’s upper- and lower-bound estimates are close to the full-sample ones.

### 3.3.4 Policy implications

As a market failure, the presence of barriers to entry into commercial farming obviously has potential policy implications. However what corrective measures are appropriate depends on the nature of the barriers. If they are private (say, the need to purchase capital equipment), then one would presume that improved access to credit would be key to reducing them. If, by contrast, they are collective (say, the quality of the local road infrastructure) then government action is needed.

In order to shed —very preliminary— light on this question, we show in Table 7 the results of a simple regression of  $\Delta\hat{\pi} \equiv \hat{\pi}_M - \hat{\pi}_S$ , the (predicted) opportunity cost of subsistence farming, on household characteristics. The idea of this regression is to identify individual characteristics correlated with high opportunity costs and hence with large gains from switching to commercial farming. This exercise should be interpreted with caution, however, if only because for all farmers other than the “marginal” one, the negative earning differential measured by  $\Delta\hat{\pi}$  (negative because the first term is hypothetical while the second term is actual) is only a lower bound on the actual entry cost.

Large households seem to have higher opportunity costs, possibly reflecting the fact that they have lower per-capita income and hence less surplus to purchase capital equipment to switch to the market. Households with a more educated and younger head and a larger farm have the lowest opportunity cost, suggesting, in accordance with intuition, that they are

more reactive to economic incentives.

The presence of a producer association widens the opportunity cost, suggesting that it raises the return to commercial farming (possibly by improving the farmers' bargaining power vis-à-vis middlemen) without making it easier for subsistence farmers to switch. Access to credit also reduces negative earnings differentials, suggesting that credit does help to overcome barriers to entry in the market. However as discussed elsewhere in this paper one shouldn't make too much of this given the very low proportion of farms with access to credit. Finally, the presence of a national road in the village raises the opportunity cost of being in subsistence, suggesting that it raises the return to commercial farming (as one would expect) without helping subsistence farmers to do the switch. Taken together, the combination of the effects of credit and transportation leans more in the direction of private than public entry barriers.

## 4 Concluding remarks

As discussed in the introduction, so far empirical work on entry barriers has largely left out agricultural markets. Yet, these markets are important in poor countries where trade-adjustment assistance mechanisms can be expected to be weak; moreover, the extreme poverty of rural households in least-developed countries like Madagascar makes them particularly vulnerable to even small deteriorations in their terms of trade. Recent work on trade and poverty has also highlighted the fact that agricultural markets themselves are fragile and can be destroyed when institutions (like intermediation) or infrastructure (like rural roads) are left to decay, which can happen when changes in relative prices are perceived to last. Restoring those markets can be long and costly.

This suggests that rural households in the grey zone between market participation and subsistence should receive particular attention in the design of reforms. Of course, the incentives faced by those households can be assessed only indirectly using the household surveys' noisy information, and there is anyway no natural definition of subsistence farming.

Notwithstanding these difficulties —for which we have no perfect fix— our analysis yields a number of results.

We define the cost of moving out of subsistence as the present discounted value of the expected opportunity cost of not switching for the “marginal farmer” —the one whose fundamentals make him just indifferent between “going commercial” and not. We then approximate this magnitude empirically by comparing the monetary equivalent of the return to subsistence vs. commercial farming.

The most striking insight coming out of our empirical analysis is that the order of magnitude of this entry cost seems to be very large: more than one year of the typical subsistence farmer’s output valued at market prices. Such a large entry barrier implies the persistence of relatively large returns differentials between subsistence and market farming. It also makes subsistence farming “sticky”, a sort of poverty trap.

Could our entry-cost calculation be biased upward? Because it is essentially an NPV calculation, there are two possible sources of upward bias: either the profit differential appearing on the numerator (the opportunity cost from not being on the market) is too high, or the interest and exit rates appearing in the denominator are too low. We use very high microcredit rates as interest rates, so the interest rate is unlikely to be the problem. There is more uncertainty about the accuracy of our exit rate, which was estimated on a short sample period (1997-99) and may not be representative of long-run exit rates. We have, however, no outside information to benchmark it and so no prior about whether it might be too high or too low. As for the denominator, by contrast, there may be an argument. Because we value subsistence output at producer prices, it is indeed possible that we under-estimate the return to subsistence farming. One might argue for using consumer prices to value output that is for consumption; as consumer prices are above producer prices, this would raise the “shadow” return to subsistence farming and hence reduce the profit differential (the opportunity cost from not being in the market) and the inferred entry cost. This should be kept in mind as a caveat, but the difference between producer and consumer prices is unlikely to change drastically the flavor of our result. Even cut in half, the entry cost would still be

several months' worth of production.

Because Malagasy agriculture has a very low productivity, it turns out that the opportunity cost of subsistence farming (the income loss attributable to the failure of farmers to switch to the market) is small relative to the overall economy (although the whole economy is characterized by low productivity): moving all farmers out of subsistence would add a meagre 0.46 percent to Madagascar's GDP.

One important question for which we can provide only a very preliminary approach is whether entry costs are private (say, the upfront purchase of capital) or collective (say, road improvements). Both conjectures are consistent with the evidence we find of unused arbitrage opportunities; yet, they would call for very different policy remedies. We tried to get a first shot at this issue by exploring the correlation between negative earnings differentials and individual characteristics, including access to credit and transport infrastructure. We take the result that access to credit seems to reduce unused opportunities while road infrastructure enlarges them as —very preliminary— evidence that entry barriers into commercial farming seem to have to do with private rather than public goods, although further research is clearly called for.

## References

- [1] Bain Joe S. (1956), “Barriers to New Competition”; *Cambridge, Harvard University Press*.
- [2] Baldwin Robert E.(1989), “Sunk-Cost Hysteresis”, *NBER Working Paper 2911*.
- [3] Baldwin Robert E., Paul Krugman (1989), “Persistent Trade Effects of Large Exchange Rate Shocks”, *Quarterly Journal of Economics*, 104, 635-654.
- [4] Barrett, Christopher (1997), “Food Marketing Liberalization and Trade Entry: Evidence from Madagascar”; *World Development*, 25, 763-777.
- [5] Baumol William J., Robert D. Willig (1981), “Fixed Costs, Sunk Costs, Entry Barriers, and Sustainability of Monopoly”, *Quarterly Journal of Economics*, 96(3), 405-431.
- [6] Brambilla, Irene and Guido Porto (2005), “Farm productivity and crop choice: evidence from cotton marketing reforms in Zambia”, mimeo, Yale University.
- [7] Deaton, Angus (1997), “The Analysis of Household Surveys”, *The World Bank, The John Hopkins University Press, Baltimore*.
- [8] De Janvry Alain, Marcel Fafchamps, Elisabeth Sadoulet (1991), “Peasant Household Behavior with Missing Markets: Some Paradoxes Explained”; *Economic Journal*, 101, 1400-1417.
- [9] Dixit, Avinash (1989), “Entry and Exit Decisions under Uncertainty”, *Journal of Political Economy*, 97, 620-638.
- [10] Geroski P.A. (1995), “What do we know about entry?”; *International Journal of Industrial Organization*, 13, 421-440.
- [11] Hansen, Bruce E. (2000), “Sample Splitting and Threshold Estimation”; *Econometrica*, 68(3), 576-603.

- [12] Hartley, Michael J. (1978), “Estimating Mixtures of Normal Distributions and Switching Regressions: comment”; *Journal of the American Statistical Association* , 73, 730-738.
- [13] Hotchkiss, Julie L. (1991), The Definition of Part-Time Employment: A Switching Regression Model with Unknown Sample Selection,” *International Economic Review*, 32, 899-917.
- [14] Lee, Lung-Fei (1979), “Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables,” *Econometrica*, 47, 977-996.
- [15] Lokshin, M., Z. Sajaia (2004), “Maximum-likelihood estimation of endogenous switching regression models,” *Stata Journal, StataCorp LP*, 4(3), 282-289.
- [16] Maddala, G.S., and F. Nelson (1975), “Switching Regression Models with Exogenous and Endogenous Switching”; *Proceedings of the American Statistical Association*, 423-6.
- [17] Maddala, G.S. (1983), “Limited-Dependent and Qualitative Variables in Econometrics”, *Econometric Society Monographs No.3, Cambridge University Press, New York*.
- [18] Martin, Stephen (2002), “Sunk Cost and Entry”; *Presidential Address Industrial Organization Society*, 20, 291-304.
- [19] Matusz, Steven and David Tarr (1999), “Adjusting to trade policy reform”, Policy Research Working Paper 2142, The World Bank.
- [20] McMillan, Margaret, Dani Rodrik and Karen Welch (2002), “When economic reform goes wrong: cashewnuts in Mozambique”, CEPR discussion paper.
- [21] Owen Robert F., David Ulph (2002), “Sunk Costs, Market Access, Economic Integration, and Welfare”, *Review of International Economics*, 10, 539-555.
- [22] Oxfam-IDS (1999), *Liberalization and Poverty*; Final report to DFID.
- [23] Paternostro, S., J. Razafindravonona, D. Stifel (2001), “Changes in Poverty in Madagascar : 1993-1999, ” Cornell Food and Nutrition Policy Program Working Paper 120.

- [24] Porto Guido G. (2005), “Informal Export Barriers and Poverty ”, *Journal of International Economics* , 66, 447-70.
- [25] Quandt, Richard (1958), “The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes”; *Journal of the American Statistical Association*, 53, 873-880.
- [26] — and J. Ramsey (1978), “Estimating Mixtures of Normal Distributions and Switching Regressions”; *Journal of the American Statistical Association* , 73, 730-738.
- [27] Rambeloalijaona, J.S. and M.L. Randrianarivelo (2003), “Early Warning in Africa : Madagascar Case Study”, *Second International Conference on Early Warning*.
- [28] Roberts Mark J., James R. Tybout (1997), “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs”, *The American Economic Review*, 87, 545-564.
- [29] Rodrik Dani (1992), “The Limits of Trade Policy Reform in Developing Countries”, *The Journal of Economic Perspectives*, 6, 87-105.
- [30] Roy, A. D. (1951), “Some Thoughts on the Distribution of Earnings”; *Oxford Economic Papers* , 3, 135-146.
- [31] Sadoulet E., S. Fukui, A. de Janvrz (1993), “Efficient share tenancz contracts under risk: The case of two rice-growing villages in Thailand”, *Journal of Development Economics*, 45, 225-243.
- [32] Sanghamitra Das, Mark J. Roberts, James R. Tybout (2001), “Market Entry Costs, Producer Heterogeneity, and Export Dynamics”; *NBER Working Paper* 8629.
- [33] Silber, W. L. (1974), “The Market for Federal Agency Securities: Is There an Optimum Size of Issue?”; *Review of Economics and Statistics* , 56, 14-22.

- [34] Stifel, David, B. Minten and P. Dorosh (2003), “Transaction Costs and Agricultural Productivity : Implications of Isolation for Rural Poverty in Madagascar”; *IFPRI MSSD discussion paper*, 56.
- [35] Sivadasan J. (2003), “Barriers to Entry and Productivity: Micro- Evidence from Indian Manufacturing Sector Reforms”; *Graduate School of Business, University of Chicago*.
- [36] Vakis Renos, Elisabeth Sadoulet, Alain de Janvry (2003), “Measuring Transactions Costs from Observed Behavior: Market Choices in Peru”, *CUDARE Working Papers, University of California, Berkeley*, 962.
- [37] Winters, L. Alan (2000), “Trade Policies for Poverty Alleviation”; in A. Matoo and P. English, eds., *Trade and the WTO: A Handbook*; The World Bank, pp. 28-38.

## A Construction of variables

The construction of the EPM variables was guided by Appendix 1 and 2 of a paper by Paternostro et al (2001). In these appendix, the authors give a comprehensive description of the EPM surveys of 1993, 1997 and 1999. Most of all, they point out the differences appearing between them. Taking their comments into consideration, we built variables that are as consistent as possible across surveys.

The description of how variables were created will be done only when the construction was not straightforward and where questions might be raised.

### A.1 Expenditures

The expenditures variable was constructed from the expenditure sections of the EPM. Following Paternostro et al (2001), we omitted items that did not appear in all surveys. In the end, this variable is composed of every-day consumption (food, clothes, hobbies....), health expenditures and schooling expenditures.

### A.2 Autoconsumption

As it clearly appears in Paternostro et al, building a measure of autoconsumption raises many problems. In this paper, autoconsumption is a key variable, since it is one of the components of the household's profit, which is the dependent variable of the model. For that reason, we paid particular attention to the construction of this variable.

As for the previous variable, only goods that showed in all surveys were included in the autoconsumption bundle. In addition, we limited our measure to autoconsumption of crops, ignoring autoconsumption of livestock and other produce resulting from cattle breeding, of product of fishing or hunting and of processed products.

Autoconsumption is evaluated at both retail price and producer price. It is the latter that is used in the measure of profits, since autoconsumption is compared with sales, which are logically valued at producer prices.

Selling prices are relatively easy to get, since in each survey, farmers are asked how much they sold of each crop and how much money they brought from it. It is then straightforward to calculate unit producer prices for each crop in each farm. However, because of differences in quality of the farmers' products<sup>29</sup>, producer prices are regressed on the household's characteristics. The fitted price obtained from this regression is free of quality differences. Autoconsumption is then evaluated at this fitted farmgate price.

We apply the same method as Paternostro et al<sup>30</sup> to compute the retail prices. Since all surveys contain a detailed expenditure section, we could calculate purchase prices for most of the crops. However, as Paternostro et al. explain in their paper, the prices we

---

<sup>29</sup>Difference in the prices of the same good sold by various farmers will partly be due to dissimilarities in the quality of the goods.

<sup>30</sup>For details, see Paternostro et al (2001), p.79.

could derive that way are rather suspect. To circumvent this problem, we make use of retail prices found in the 1999 community questionnaire. These prices were directly noted down from the displays by the survey investigator, in each community. Comparing these prices with the 1999 selling prices, a markup is calculated. This markup is then applied to the fitted producer prices in the four surveys. The hence calculated retail prices are then used to evaluate autoconsumption.

### A.3 Evolution of prices

We would expect that one of the most important decision variable is the changes in prices of the crops the farmer grows. We therefore compute price indexes reflecting the general change in prices of the farmer's crops.

The evident pitfall of such an exercise is the determination of the period over which the price indexes should be calculated. In order to be closer to reality, crops are split into two groups depending on the time needed between the planting of the crop until the first harvest. This can take from as little time as 3 months (rice for example) to as much as 8 years (cloves). Crops that need 1.5 year or less to reach maturity are classified as "short term crops". The rest is described as "long term crops".<sup>31</sup> The distinction between the two groups is important with respect to the decision process of farmers. Undeniably, the impact of a change in the price of a crop will be different whether the farmer faces a short term crop or a long term crop. In the former case, a sudden fall in the price will prompt the farmer to drop that crop for another, more profitable crop. In contrast, in the latter case, a sudden diminution of the price will not have much impact on the farmer's cropping decision. Indeed, not only is the profitability of a long term crop determined over the several past years, but also, the investment that growing such a crop represents will hinder the farmer from taking any rushed decision.

With this in mind, a Tornqvist price index was computed, where long term crop price changes were observed on 10 years and short term crop ones were taken on 3 years. In other words, we took the average change in price over 10 years for the long term crops, when possible<sup>32</sup> and on 3 years for the short term crops. Tornqvist price index has therefore the following form :

$$Tornqvist_i = \prod_{l=1}^L (\Delta p_{il})^{share_{il}} \prod_{s=1}^S (\Delta p_{is})^{share_{is}} \quad l, s \in \{1, 4\}, \quad L + S = 4, \quad {}^{33} \quad (21)$$

---

<sup>31</sup>See Table 3 for the classification of crops.

<sup>32</sup>As mentioned before, the FAO database contains producer prices from 1991 to 2001. The 1993 long term price index was thus calculated over the period 1991-1993, whereas the 2001 index was computed over the period 1991-2001. This divergence is certainly not desirable, but inescapable because of the limitation of the data.

<sup>33</sup>The total of the long term and short term crops must equal four, since farmers can grow four crops at most.

where  $i$  stands for the household,  $l$  for long term crops,  $s$  for short term crops and

$$\Delta p_{il} = \sum_{n=1}^9 \frac{1}{9} \left( \frac{p_{l,n}}{p_{l,n-1}} \right), \quad (22)$$

$$\Delta p_{is} = \sum_{n=1}^2 \frac{1}{2} \left( \frac{p_{l,n}}{p_{l,n-1}} \right), \quad (23)$$

where  $n$  stands for years, and

$$share_{ic} = \frac{\text{valued harvest}_{ic}}{\text{total valued harvest}_i} \quad c \in [l, s]^{34} \quad (24)$$

In order to control for the volatility of prices over the years, another Tornqvist index was computed by taking the standard error of the price changes instead of the averages :

$$S_E \text{Tornqvist}_i = \prod_{l=1}^L (S_E \Delta p_{il})^{share_{il}} \prod_{s=1}^S (S_E \Delta p_{is})^{share_{is}} \quad (25)$$

where  $S_E \Delta p_{il}$  is the standard error of the price changes of long term over the 10 years and  $S_E \Delta p_{is}$  is the standard error of the price changes of short term over 3 years.

Finally, note that the Tornqvist price index is referred to as ‘‘Price index’’ in the regressions and the volatility price index as ‘‘SE Price index’’.

---

<sup>34</sup>The harvest of household  $i$  at time  $t$  (recall that each household is observed once only) is valued at the prices of the different crops at time  $t$ .

**Table 1: Descriptive statistics**

Statistics	Total	Market	Subsistence
Sample size	5951	5502	449
Family size			
Mean	5.09	5.08	5.14
Std. Error	0.05	0.05	0.14
Household's head, percentage by sexe			
Male	85.60	85.88	81.86
Female	14.40	14.12	18.14
Household's head age distribution (% population)			
15-30 years	23.29	23.53	19.99
30-45 years	38.74	38.93	36.18
45-60 years	24.85	24.52	29.23
60 years and more	12.99	12.88	14.59
Household's head highest level of education attained (% population)			
Primary	84.73	84.69	85.21
Std. Error	0.6	0.6	1.8
Secondary	14.97	15.00	14.61
Std. Error	0.6	0.6	1.8
University	0.29	0.3	0.17
Std. Error	0.1	0.1	0.2
Cropland surface (ha)			
Mean	1.61	1.65	1.06
Std. Error	0.05	0.06	0.07
Percentage of owned land			
Mean	79.94	80.08	76.48
Std. Error	0.8	0.8	2.4
Percentage of land in sharecropping			
Mean	4.06	3.96	5.47
Std. Error	0.3	0.3	1.12
Percentage of rented land			
Mean	3.64	3.56	4.64
Std. Error	0.3	0.27	0.9
Percentage of households with outstanding agricultural loan			
Mean	4.15	4.26	2.66
Std. Error	0.39	0.4	0.1
Remoteness (% population being)			
Least remote	34.83	34.57	38.29
Std. Error	2.01	1.7	3.0
Fairly remote	43.19	43.02	45.41
Std. Error	2.33	1.9	3.2
Most remote	21.96	22.39	16.27
Std. Error	2.02	2.1	2.9
Annual expenditures (mean)			
MGF	389230.82	397496.50	278969.91
Std. Error	9794.90	10224.74	12846.21
\$, PPP	566.52	578.55	406.03
Std. Error	14.88	15.71	18.70
Annual profits (mean)			
MGF	626374.10	639355.54	453208.10
Std. Error	13344.29	13728.22	27877.13
\$, PPP	911.67	930.57	659.63
Std. Error	19.42	19.98	40.57
Share of income of livestock sales in full income (mean)			
Mean	7.46	6.55	19.70
Std. Error	0.27	0.23	1.12
Share of crop sales in total income			
Mean	71.0	76.30	0
Std. Error	0.8	0.63	0
Share of crop sales in crop harvest (in value)			
Mean	38.05	40.90	0
Std. Error	0.67	0.6	0
Share of crop sale in total crop sales (mean)			
Main crop	80.50	80.50	0
Std. Error	0.55	0.55	0
Second main crop	15.25	15.25	0
Std. Error	0.37	0.37	0
Third crop	3.46	3.46	0
Std. Error	0.18	0.18	0
Fourth crop	0.78	0.78	0
Std. Error	0.06	0.06	0
Share of crop harvest in harvest of all crops (mean)			
Main crop	56.54	55.58	69.28
Std. Error	0.77	0.8	1.6
Second main crop	24.09	24.37	20.47
Std. Error	0.45	0.5	1.0
Third crop	13.56	13.99	7.76
Std. Error	0.39	0.4	0.8
Fourth crop	5.81	6.06	2.49
Std. Error	0.27	0.3	0.3
Most cultivated crops (share of crop's total harvest in population total harvest)	Potato 55.22 Rice 25.84 Cassava 6.47 Coffee 2.65 Maize 1.78	Potato 57.18 Rice 24.87 Cassava 6.09 Coffee 2.65 Maize 1.66	Rice 68.76 Yam 16.39 Sweet potato 9.55 Maize 5.16 Coffee 4.04
Most cultivated crops in value (% crop's population total sales in population total sales)	Rice 36.39 Coffee 11.49 Vanilla 10.56 Cassava 7.98 Tobacco 4.26	Rice 36.39 Coffee 11.49 Vanilla 10.56 Cassava 7.98 Tobacco 4.26	- - - - -

**Table 2: Description of variables**

Variable	Signification
Profits	Household's annual agricultural profits. The profits are composed of the agricultural sales and of the value of crop autoconsumption.
HH size	Number of people in the family outside the household's head.
HH head age 2	Household's heads' age (squared age), in logs.
HH head schooling	Household's head's last achieved school year. 1-6 : primary school: 7-13, secondary school, 14-18: university.
Members age	Household's members' average age.
Cropland size	Number of squared meters of land the household crops or has in fallow, in logs <sup>a</sup>
Owner	Proportion of owned cropped land.
Tenant	Proportion of rented cropped land.
Sharecropper	Proportion of cropped land in sharecropping.
Outstanding loan	Dummy indicating whether the household has a loan meant for agriculture. 1=yes, 0=no.
Inputs	Dummy indicating whether the household used inputs in its production. 1=yes, 0=no.
Price index lt	Price index for long term crops, in logs.
Price index st	Price index for short term crops, in logs.
SE. Price index lt	Standard error of the price index for long term crops, in logs.
SE. Price index st	Standard error of the price index for short term crops, in logs.
Temp. amplitude	Temperature amplitude : range between the highest recorded temperature and the lowest one.
Disturbances	Index indicating the risk of atmospheric disturbances. Goes from 0=low risk to 5=high risk.
Transport cost	Cost of transporting 50 kg of rice to the nearest town.
Farmers association	Dummy indicating whether the village where the household lives has an association of farmers. 1=yes, 0=no.
Share agri. Pop.	Share of the population of the community in the agricultural sector.
Fertilizer use	Share of the farmers of the community who use fertilizers. 0=0%, 1=5%, 2=5-25%,., 5=75%.
National road	Dummy indicating if a national road runs throughout the community. 1=yes, 0=no.
Steer	Number of steers in the community.
Important crop	Dummy indicating if the household's first or second crop corresponds to the most important crop (in surface or in value) in the community. 1=yes, 0=no.
Number of rich	Percentage of rich people in the community.
Number of poor	Percentage of poor people in the community. <sup>b</sup>
Remoteness index	Index indicating how remote the community is. This index takes into account the infrastructure of the community and transport facilities. For more details, see the Ilo census of the Cornell University. Goes from 1=least remote to 5 most remote.
Livestock	Number of cows and pigs in the community.
Regional dummies	Dummies for the 6 Faritanys (provinces). Each dummy equals to 1 if the village is in this province and 0 otherwise.
Year dummies	Dummies indicating to which EPM the household belong. 1=yes, 0=no.

<sup>a</sup>For variables that have zero values, we added 1 to the variable so that the log is always feasible.

<sup>b</sup>Number of rich and number of poor do not add to 1, since the data gives in addition the number of middle-rich and of middle-poor.

**Table 3: *Short term and long term crops***

<b>Short term crops</b> <b>(Less than 18 months</b> <b>from planting to</b> <b>harvest)</b>	<b>Time from</b> <b>planting to</b> <b>harvest</b>  <b>Months</b>	<b>Long term crops</b>  <b>More than 18 months</b> <b>from planting to</b> <b>harvest</b>	<b>Time from</b> <b>planting to</b> <b>harvest</b>  <b>Years</b>
Ananas	12	Apple	2-5
Banana	9-12	Avocado	4-10
Beans	2-3	Cloves	6-8
Beetroot	1	Cocoa beans	4-5
Brede (leaves)	2-6	Coconut	3-8
Cabbage	7-12	Coffee	3-4
Carrots	2-3	Litchi	3-5
Cassava	12	Mandarin	3-5
Cotton	18	Mango	3-5
Cucumber	2-3	Oranges	3-5
Garlic	8-12	Peaches	3-4
Leeks	8-12	Pepper	3-8
Lenses	6	Vanilla	4-7
Lettuce	3-4		
Maize	3		
Melon	6-12		
Onions	1-3		
Peanuts	4-5		
Peas	2-3		
Potato	3		
Pumpkin	4-6		
Rice	3-6		
Soja	3		
Sugar cane	18		
Sweet potato	3-7		
Taro	6-12		
Tobacco	5-8		
Tomatoes	2-3		
Watermelon	6-12		
Wheat	6		
Yam	9-12		

Table 4: *Switching regression results,  $\lambda = 0$*

	Inprofit <sub>1</sub>	Inprofit <sub>0</sub>	occupation
HH size	0.07*** [14.15]	0.04** [2.07]	-0.01 [0.85]
HH head age	0.672 [1.04]	4.76*** [2.69]	1.66 [1.09]
HH head age 2	-0.08 [0.94]	-0.63*** [2.63]	-0.25 [1.19]
HH head schooling	0.04*** [10.33]	0.04*** [3.79]	0.02 [1.45]
Members age	0.14*** [6.54]	-0.02 [0.93]	0.05 [0.72]
Cropland size	0.19*** [11.29]	0.21*** [2.77]	0.13*** [3.24]
Outstanding loan	0.18*** [2.99]	0.29*** [2.57]	0.10 [0.46]
Inputs	0.17*** [3.90]	0.07 [0.80]	0.8 [0.64]
Fertilizer use	-1.54E-03 [0.12]	-0.02 [0.88]	0.04 [1.03]
Steer	4.89E-07 [0.24]	-3.42E-06 [1.13]	-6.04E-06* [1.80]
Owner	0.02 [0.50]	-0.13 [0.80]	0.17 [1.44]
Sharecropper	0.01 [0.07]	-0.26 [1.07]	-0.13 [0.70]
Tenant	0.15* [1.93]	0.29 [1.38]	-0.05 [0.22]
Temp. amplitude	0.01 [1.19]	0.01 [0.64]	-0.01 [0.46]
Disturbances	0.02 [0.61]	0.07 [0.89]	-0.15*** [2.69]
Transport cost	3.40E-07 [0.19]	-1.54E-06 [0.46]	-4.31E-06 [1.51]
National road	0.08* [1.85]	-0.01 [0.14]	0.04 [0.43]
Farmers association	0.07 [1.49]	-0.13 [1.62]	0.13 [1.21]
Share agri. pop.	-1.72E-03* [1.65]	-2.08E-03 [0.81]	-3.89E-03 [1.50]
Important crop	-4.09E-03 [0.12]	-0.05 [0.61]	-0.15 [1.61]
Year1993	-0.25*** [3.32]	-0.74** [2.48]	0.56*** [2.86]
Year1997	0.33*** [5.61]	0.19 [1.33]	-0.17 [1.19]
Year1999	0.33*** [5.49]	0.21 [1.44]	-0.03 [0.16]
Faritany2	-0.20*** [3.28]	0.06 [0.49]	0.10 [0.74]
Faritany3	-0.19** [2.28]	-0.48* [1.73]	0.55*** [2.93]
Faritany4	-0.15 [1.26]	-0.31 [1.07]	0.65** [2.23]
Faritany5	-0.23*** [2.62]	-0.38*** [1.69]	0.30 [1.46]
Faritany7	0.08 [1.04]	-0.15 [0.66]	0.38** [2.03]
Number of poor			-0.01*** [2.59]
Number of rich			-0.01** [2.14]
Remoteness index			0.09* [1.81]
Livestock			-1.65E-06 [0.11]
Price index			2.06*** [6.74]
SE Price index			-0.04** [2.19]
Constant	9.15*** [8.14]	2.06 [0.62]	-2.47 [0.81]
Sigma <sub>1</sub>	0.62*** [26.49]		
Sigma <sub>2</sub>	0.62*** [12.53]		
rho <sub>1</sub>	0.53*** [3.62]		
rho <sub>2</sub>	0.08 [0.13]		
$\hat{\sigma}_\lambda$	2.26E-04		
Observations	5951		
Wald chi2(28) 1171.50 Prob > chi2=0.000			
Wald test of indep. eqns. : chi2(1) = -9.7E+06 Prob > chi2 = 1.0000			

Robust z statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 5a: *Switching regression results,  $\lambda = 0.1$*

	<b>lnprofit<sub>1</sub></b>	<b>lnprofit<sub>0</sub></b>	<b>occupation</b>
HH size	0.06*** [13.82]	0.05*** [3.59]	-0.01 [1.31]
HH head age	0.52 [0.84]	3.61** [2.36]	0.90 [0.69]
HH head age 2	-0.06 [0.73]	-0.48** [2.33]	-0.15 [0.82]
HH head schooling	0.04*** [10.14]	0.05*** [4.46]	0.02* [1.87]
Members age	0.14*** [6.47]	0.07 [0.97]	0.06 [1.16]
Cropland size	0.19*** [11.31]	0.21*** [4.80]	0.10*** [2.93]
Outstanding loan	0.17** [2.53]	0.19* [1.84]	0.06 [0.35]
Inputs	0.20*** [4.36]	-0.02 [0.29]	0.07 [0.73]
Fertilizer use	-5.46E-04 [0.04]	-0.02 [0.79]	0.07** [2.24]
Steer	4.61E-07 [0.23]	-2.62E-06 [0.91]	-2.28E-06 [0.68]
Owner	0.02 [0.54]	-0.01 [0.12]	0.19** [2.03]
Sharecropper	0.02 [0.29]	-0.30 [1.38]	-0.14 [0.78]
Tenant	0.16** [2.00]	0.12 [0.72]	-0.04 [0.22]
Temp. amplitude	4.12E-03 [0.40]	0.03 [1.47]	-0.04 [1.53]
Disturbances	0.02 [0.55]	0.02 [0.35]	-0.11** [2.02]
Transport cost	9.36E-08 [0.06]	-4.48E-07 [0.15]	-6.31E-06** [2.30]
National road	0.07* [1.73]	-0.02 [0.24]	-2.44E-03 [0.03]
Farmers association	0.06 [1.40]	-0.02 [0.28]	0.03 [0.34]
Share agri. pop.	-1.37E-03 [1.36]	-2.89E-03 [1.29]	-3.60E-03 [1.53]
Important crop	-0.01 [0.29]	-0.01 [0.09]	-0.05 [0.68]
Year1993	-0.18** [2.40]	-0.64*** [4.26]	0.10 [0.60]
Year1997	0.35*** [5.70]	0.13 [1.17]	-0.20 [1.45]
Year1999	0.36*** [5.67]	0.12 [1.00]	0.01 [0.05]
Faritany2	-0.19*** [3.09]	-0.01 [0.08]	0.14 [1.11]
Faritany3	-0.15* [1.87]	-0.27 [1.37]	0.48*** [2.82]
Faritany4	-0.07 [0.54]	-0.34 [1.44]	0.82*** [3.12]
Faritany5	-0.19** [2.13]	-0.36* [1.92]	0.39** [2.14]
Faritany7	0.08 [1.09]	0.11 [0.53]	0.44*** [2.70]
Number of poor			-4.16E-03** [2.21]
Number of rich			-3.23E-03 [1.32]
Remoteness index			0.07 [1.56]
Livestock			-1.18E-05 [0.96]
Price index			1.99*** [8.08]
SE Price index			-0.03* [1.74]
Constant	9.27*** [8.01]	4.20 [1.46]	-1.15 [0.47]
Sigma <sub>1</sub>	0.61*** [22.79]		
Sigma <sub>2</sub>	0.69*** [11.26]		
rho <sub>1</sub>	0.51*** [3.54]		
rho <sub>2</sub>	0.23 [0.82]		
Observations	5823		

Wald chi2(28) 1196.55 Prob > chi2=0.000

Wald test of indep. eqns. : chi2(1) = -1.0e+07 Prob > chi2 = 1.0000

Robust z statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 5b: *Switching regression results,  $\lambda = 0.2$*

	<b>lnprofit<sub>1</sub></b>	<b>lnprofit<sub>0</sub></b>	<b>occupation</b>
HH size	0.06*** [13.03]	0.05*** [4.82]	-0.02* [1.84]
HH head age	-0.11 [0.17]	3.57*** [2.92]	0.51 [0.42]
HH head age 2	0.02 [0.26]	-0.47*** [2.85]	-0.1 [0.61]
HH head schooling	0.03*** [9.42]	0.04*** [5.14]	0.02** [2.37]
Members age	0.16*** [6.96]	0.03 [0.62]	0.09* [1.69]
Cropland size	0.19*** [10.58]	0.23*** [6.51]	0.07** [2.33]
Outstanding loan	0.24*** [3.52]	0.17** [2.03]	0.15 [1.02]
Inputs	0.19*** [3.98]	0.04 [0.53]	0.02 [0.20]
Fertilizer use	4.82E-03 [0.33]	-0.02 [1.01]	0.05* [1.79]
Steer	3.14E-07 [0.16]	-1.14E-06 [0.46]	-1.42E-06 [0.46]
Owner	-0.01 [0.34]	0.05 [0.57]	0.17* [1.79]
Sharecropper	0.02 [0.27]	-0.17 [1.03]	-0.10 [0.53]
Tenant	0.15* [1.95]	0.07 [0.48]	0.10 [0.59]
Temp. amplitude	3.85E-03 [0.39]	0.03* [1.85]	-0.04 [1.58]
Disturbances	0.03 [1.06]	0.02 [0.54]	-0.12** [2.24]
Transport cost	7.65E-07 [0.57]	-1.47E-06 [0.62]	-5.94E-06** [2.31]
National road	0.07* [1.70]	0.01 [0.14]	-0.01 [0.15]
Farmers association	0.06 [1.39]	0.02 [0.36]	0.03 [0.34]
Share agri. pop.	-1.23E-03 [1.21]	-2.43E-03 [1.45]	-4.09E-03** [1.97]
Important crop	0.01 [0.34]	-0.06 [1.19]	-1.87E-03 [0.03]
Year1993	-0.08 [1.07]	-0.73*** [5.36]	0.19 [1.19]
Year1997	0.35*** [5.49]	0.21** [2.48]	-0.18 [1.39]
Year1999	0.34*** [5.17]	0.24*** [2.81]	-0.08 [0.60]
Faritany2	-0.22*** [3.39]	-0.03 [0.37]	0.15 [1.13]
Faritany3	-0.19** [2.33]	-0.12 [0.88]	0.33** [2.02]
Faritany4	-0.09 [0.71]	-0.27* [1.79]	0.59** [2.50]
Faritany5	-0.20** [2.21]	-0.32** [2.31]	0.45** [2.52]
Faritany7	0.08 [1.04]	0.03 [0.24]	0.53*** [3.24]
Number of poor			-4.03E-03** [2.16]
Number of rich			-0.01*** [2.59]
Remoteness index			0.07* [1.77]
Livestock			-1.69E-05 [1.54]
Price index			1.90*** [7.48]
SE Price index			-0.02 [1.20]
Constant	10.39*** [8.79]	3.93* [1.73]	-0.41 [0.18]
Sigma <sub>1</sub>	0.59*** [19.22]		
Sigma <sub>2</sub>	0.67*** [19.51]		
rho <sub>1</sub>	0.41** [2.28]		
rho <sub>2</sub>	0.16 [0.92]		
Observations	5599		

Wald chi2(28) 1102.04 Prob > chi2=0.000  
Wald test of indep. eqns. : chi2(1) = -1.1e+07 Prob > chi2 = 1.0000  
Robust z statistics in brackets  
\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 5c: *Switching regression results,  $\lambda = 0.3$*

	<b>lnprofit<sub>1</sub></b>	<b>lnprofit<sub>0</sub></b>	<b>occupation</b>
HH size	0.07*** [13.02]	0.05*** [5.98]	-0.02** [2.39]
HH head age	0.20 [0.31]	2.04* [1.94]	0.47 [0.41]
HH head age 2	-0.02 [0.21]	-0.26* [1.84]	-0.10 [0.66]
HH head schooling	0.04*** [9.10]	0.04*** [5.76]	0.02*** [2.71]
Members age	0.17*** [6.78]	0.05 [1.11]	0.13*** [2.71]
Cropland size	0.18*** [9.47]	0.23*** [8.13]	0.07** [2.34]
Outstanding loan	0.26*** [4.67]	0.10 [1.28]	0.22* [1.66]
Inputs	0.19*** [4.06]	0.07 [1.24]	0.02 [0.24]
Fertilizer use	8.83E-04 [0.06]	-0.01 [0.28]	0.05* [1.72]
Steer	-3.30E-07 [0.16]	3.42E-07 [0.14]	-9.16E-07 [0.31]
Owner	-0.05 [1.22]	0.06 [0.71]	0.04 [0.42]
Sharecropper	-0.01 [0.13]	-0.12 [0.89]	-0.09 [0.51]
Tenant	0.11 [1.41]	0.10 [0.80]	0.08 [0.47]
Temp. amplitude	0.01 [0.57]	0.01 [1.06]	-0.03 [1.46]
Disturbances	0.02 [0.89]	0.03 [0.83]	-0.17*** [3.32]
Transport cost	3.54E-07 [0.25]	-1.35E-07 [0.07]	-6.90E-06*** [2.68]
National road	0.06 [1.61]	0.02 [0.28]	-0.04 [0.44]
Farmers association	0.08* [1.95]	0.02 [0.31]	0.07 [0.80]
Share agri. pop.	-1.62E-03 [1.60]	-1.47E-03 [0.85]	-0.01*** [3.24]
Important crop	-0.01 [0.29]	-0.03 [0.59]	-0.06 [0.87]
Year1993	-0.01 [0.09]	-0.68*** [5.38]	0.28* [1.80]
Year1997	0.36*** [5.44]	0.25*** [3.27]	-0.18 [1.47]
Year1999	0.33*** [4.79]	0.25*** [3.29]	-0.15 [1.20]
Faritany2	-0.24*** [3.51]	-0.06 [0.86]	0.12 [0.98]
Faritany3	-0.19** [2.40]	-0.12 [1.00]	0.42*** [2.70]
Faritany4	-0.07 [0.54]	-0.22 [1.57]	0.50** [2.21]
Faritany5	-0.17* [1.88]	-0.32** [2.46]	0.55*** [3.22]
Faritany7	0.07 [0.87]	0.09 [0.71]	0.67*** [4.41]
Number of poor			-2.67E-03 [1.54]
Number of rich			-0.01** [2.40]
Remoteness index			0.05 [1.28]
Livestock			-2.31E-05** [2.04]
Price index			1.98*** [9.05]
SE Price index			2.78E-04 [0.02]
Constant	9.89*** [8.14]	6.51*** [3.34]	-0.23 [0.11]
Sigma <sub>1</sub>	0.57*** [18.26]		
Sigma <sub>2</sub>	0.65*** [22.41]		
rho <sub>1</sub>	0.31 [1.55]		
rho <sub>2</sub>	0.17 [1.15]		
Observations	5372		
Wald chi2(28) 1057.36 Prob > chi2=0.000			
Wald test of indep. eqns. : chi2(1) = -1.1e+07 Prob > chi2 = 1.0000			
Robust z statistics in brackets			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Table 5d: *Switching regression results with short term and long term prices,  $\lambda = 0$*

	<b>lnprofit<sub>1</sub></b>	<b>lnprofit<sub>0</sub></b>	<b>occupation</b>
HH size	0.07*** [14.24]	0.04* [1.92]	-0.01 [0.62]
HH head age	0.68 [1.14]	3.71** [2.03]	2.65* [1.66]
HH head age 2	-0.08 [1.03]	-0.48* [1.94]	-0.38* [1.77]
HH head schooling	0.04*** [10.43]	0.04*** [3.70]	0.02 [1.57]
Members age	0.14*** [6.58]	-0.03 [0.28]	0.04 [0.51]
Cropland size	0.18*** [11.20]	0.20** [2.20]	0.13*** [2.97]
Outstanding loan	0.18*** [2.99]	0.38** [2.53]	0.11 [0.55]
Inputs	0.17*** [3.97]	0.03 [0.27]	0.10 [0.79]
Fertilizer use	-1.78E-03 [0.13]	-0.02 [0.72]	0.05 [1.49]
Steer	4.66E-07 [0.23]	-1.40E-06 [0.44]	-5.44E-06* [1.67]
Owner	0.02 [0.40]	-0.18 [1.04]	0.19 [1.62]
Sharecropper	0.01 [0.15]	-0.31 [1.26]	-0.06 [0.32]
Tenant	0.15* [1.92]	0.31 [1.44]	-0.06 [0.27]
Temp. amplitude	0.01 [1.19]	0.02 [1.03]	-0.02 [0.52]
Disturbances	0.02 [0.69]	0.11 [1.27]	-0.17*** [2.95]
Transport cost	2.91E-07 [0.19]	-2.33E-06 [0.69]	-2.66E-06 [0.75]
National road	0.08** [1.97]	-0.04 [0.52]	0.01 [0.08]
Farmers association	0.06 [1.56]	-0.15* [1.87]	0.16 [1.40]
Share agri. pop.	-1.66E-03 [1.62]	-1.74E-03 [0.64]	-3.88E-03 [1.49]
Important crop	-3.57E-03 [0.11]	-0.02 [0.21]	-0.20** [2.07]
Year1993	-0.26*** [3.51]	-0.89** [2.41]	0.72*** [3.44]
Year1997	0.33*** [5.72]	0.14 [0.91]	-0.15 [0.99]
Year1999	0.34*** [5.53]	0.11 [0.74]	0.03 [0.21]
Faritany2	-0.20*** [3.36]	0.06 [0.41]	0.04 [0.30]
Faritany3	-0.19** [2.38]	-0.61** [2.15]	0.48** [2.49]
Faritany4	-0.15 [1.31]	-0.51 [1.53]	0.75** [2.49]
Faritany5	-0.23*** [2.59]	-0.56** [2.42]	0.27 [1.35]
Faritany7	0.07 [0.97]	-0.20 [0.86]	0.23 [1.05]
Number of poor			-0.01*** [2.63]
Number of rich			-0.01* [1.94]
Remoteness index			0.08 [1.42]
Livestock			4.84E-06 [0.30]
Price index long term			8.38*** [4.11]
SE Price index long term			-3.20*** [3.54]
Price index short term			0.52 [1.47]
SE Price index short term			-0.06*** [2.84]
Constant	9.09*** [8.09]	3.77 [1.11]	-4.10 [1.39]
Sigma <sub>1</sub>	0.61** [26.17]		
Sigma <sub>2</sub>	0.62*** [9.84]		
rho <sub>1</sub>	0.43** [2.32]		
rho <sub>2</sub>	-0.09 [0.15]		
Observations	5955		

Wald chi2(28) 1193.52 Prob > chi2=0.000

Wald test of indep. eqns. : chi2(1) = -9.6e+06 Prob > chi2 = 1.0000

Robust z statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 5e: *Switching regression results on rice subsample,  $\lambda = 0$*

	lnprofit <sub>1</sub>	lnprofit <sub>0</sub>	occupation
HH size	0.06*** [12.62]	0.04* [1.81]	-0.01 [0.81]
HH head age	0.46 [0.70]	4.53* [2.37]	1.33 [0.81]
HH head age 2	-0.05 [0.59]	-0.60** [2.32]	-0.22 [1.00]
HH head schooling	0.03*** [9.73]	0.04*** [3.66]	1.27E-03 [0.12]
Members age	0.14*** [5.94]	2.05E-03 [0.02]	0.06 [0.82]
Cropland size	0.21*** [11.11]	0.21** [2.57]	0.16*** [3.78]
Outstanding loan	0.17** [2.51]	0.35** [2.35]	0.13 [0.62]
Inputs	0.14*** [3.30]	0.08 [0.79]	0.05 [0.39]
Fertilizer use	0.01 [0.63]	-0.03 [1.05]	0.02 [0.60]
Steer	7.53E-07 [0.37]	-2.77E-06 [0.91]	-2.27E-06 [0.63]
Owner	0.02 [0.41]	-0.12 [0.66]	0.25** [2.03]
Sharecropper	-0.03 [0.36]	-0.25 [1.02]	-0.06 [0.32]
Tenant	0.11 [1.42]	0.31 [1.42]	0.05 [0.20]
Temp. amplitude	0.02 [1.44]	0.02 [0.62]	0.02 [0.64]
Disturbances	0.01 [0.24]	0.03 [0.34]	-0.15*** [2.66]
Transport cost	-4.15E-07 [0.25]	-3.14E-06 [0.85]	-5.76E-06*** [2.66]
National road	0.08* [1.92]	-0.04 [0.42]	0.05 [0.41]
Farmers association	0.03 [0.81]	-0.12 [1.40]	0.06 [0.52]
Share agri. pop.	-2.57E-05** [2.37]	-1.65E-03 [0.59]	-3.87E-03 [1.40]
Important crop	0.02 [0.52]	-0.05 [0.54]	-0.12 [1.11]
Year1993	-0.33*** [4.47]	-0.77** [2.30]	0.45** [2.20]
Year1997	0.35*** [5.94]	0.12 [1.19]	-0.21 [1.46]
Year1999	0.33*** [5.43]	0.20 [1.26]	-0.09 [0.57]
Faritany2	-0.20*** [3.38]	0.08 [0.58]	0.06 [0.45]
Faritany3	-0.16* [1.96]	-0.34 [1.11]	0.52*** [2.66]
Faritany4	-0.19* [1.71]	-0.21 [0.73]	0.37 [1.27]
Faritany5	-0.13 [1.45]	-0.24 [0.92]	0.31 [1.50]
Faritany7	0.09 [1.26]	-0.09 [0.34]	0.37* [1.89]
Number of poor			-0.01** [2.54]
Number of rich			-0.01** [2.45]
Remoteness index			0.08 [1.56]
Livestock			-1.23E-06 [0.08]
Price index			2.17*** [5.88]
SE Price index			-0.05* [1.86]
Constant	9.34*** [7.62]	2.49 [0.70]	-2.05 [0.66]
Sigma <sub>1</sub>	0.61*** [24.93]		
Sigma <sub>2</sub>	0.63*** 9.61		
rho <sub>1</sub>	0.52*** [3.35]		
rho <sub>2</sub>	0.13 [0.24]		
Observations	4996		
Wald chi2(29) 1153.65 Prob > chi2=0.000			
Wald test of indep. eqns. : chi2(1)= -8.0e+06 Prob > chi2=1.000			
Robust z statistics in brackets			
* significant at 10%; ** significant at 5%; *** significant at 1%			

**Table 6: Conditional profits**

		Subsistence, $I_i = 0$	Market, $I_i = 1$	Opportunity cost	Entry cost		
					Lower bound	Marginal HH	Upper bound
$\lambda = 0$	$\pi_{M_i}$	210399.5	551470.3				
	Se	(3877.9)	(7688.2)				
	$\pi_{S_i}$	387379.1	526704.5	167800.1	481496.9	536385.9	591274.9
	Se	(13999.8)	(12818.9)	(9543.9)		(28004.6)	
$\lambda = 0.1$	$\pi_{M_i}$	326586.0	563272.6				
	Se	(4947.8)	(7784.7)				
	$\pi_{S_i}$	427452.2	531248.9	133216.1	393341.8	429835.5	466329.2
	Se	(10988.6)	(11332.1)	(6345.4)		(18619.2)	
$\lambda = 0.2$	$\pi_{M_i}$	394503.4	571292.7				
	Se	(6331.5)	(7955.8)				
	$\pi_{S_i}$	489019.1	542381.9	103329.2	404515.2	433759.0	463002.8
	Se	(10196.8)	(9393.6)	(5084.8)		(14920.3)	
$\lambda = 0.3$	$\pi_{M_i}$	463059.6	580895.0				
	Se	(7336.1)	(8705.2)				
	$\pi_{S_i}$	513787.9	561932.1	96145.7	400168.3	426384.5	452600.7
	Se	(8424.9)	(8894.3)	(4558.4)		(13375.6)	

**Table 6a: Conditional profits on rice subsample**

		Subsistence, $I_i = 0$	Market, $I_i = 1$	Opportunity cost	Entry cost		
					Lower bound	Marginal HH	Upper bound
Rice, $\lambda = 0$	$\pi_{M_i}$	231262.7	561033.0				
	Se	(4437.8)	(8604.2)				
	$\pi_{S_i}$	391486.0	548657.1	181749.2	520575.5	580631.8	640688.1
	Se	(14083.4)	(14345.4)	(10442.5)		(30641.0)	
Rice, $\lambda = 0.1$	$\pi_{M_i}$	427582.5	573416.8				
	Se	(6743.8)	(8628.9)				
	$\pi_{S_i}$	439159.1	575805.2	157048.6	477889.3	518329.4	558769.5
	Se	(10983.6)	(12862.3)	(7031.6)		(20632.7)	
Rice, $\lambda = 0.2$	$\pi_{M_i}$	427658.8	581833.3				
	Se	(7169.3)	(8781.8)				
	$\pi_{S_i}$	503923.1	562510.2	114272.0	357045.9	390150.0	423254.1
	Se	(10576.6)	(10554.4)	(5756.1)		(16889.8)	
Rice, $\lambda = 0.3$	$\pi_{M_i}$	502038.1	591717.1				
	Se	(8374.7)	(9437.7)				
	$\pi_{S_i}$	529234.2	581513.4	110794.7	391321.7	419105.6	446889.5
	Se	(9009.2)	(10137.5)	(4831.0)		(14175.5)	

**Table 7: Opportunity cost regressions**

	Subsistence farmers
Expenditure per capita	-0.04 [1.13]
HH size	0.09*** [9.40]
HH head age	-10.51*** [9.32]
HH head age 2	1.38*** [9.01]
HH head schooling	-0.01* [1.94]
Members age	0.51*** [10.54]
Cropland size	-0.04** [1.98]
Owner	0.72*** [7.52]
Sharecropper	1.02*** [7.58]
Tenant	-0.71*** [3.37]
Outstanding loan	-1.21*** [3.01]
Farmers association	0.81*** [14.14]
Share agri. pop.	0 [0.34]
Transport cost	0.00*** [3.72]
National road	0.37*** [5.91]
Year1993	1.49*** [10.00]
Year1997	0.47*** [4.25]
Year1999	0.41*** [4.42]
Faritany1	-0.31*** [3.50]
Faritany2	-1.42*** [9.62]
Faritany3	0.21** [2.10]
Faritany4	-0.07 [0.42]
Faritany5	0.08 [0.83]
Constant	16.79*** [7.97]
Observations	371
R-squared	0.71

Absolute value of t statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Figure 1: Evolution of prices

