UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Challenging small-scale farming, a non-parametric analysis of the (inverse) relationship between farm productivity and farm size in Burundi

Marijn Verschelde
Marijke D'Haese
Glenn Rayp
Ellen Vandamme

October 2011

2011/745

D/2011/7012/50

Challenging small-scale farming, a non-parametric analysis of the (inverse) relationship between farm productivity and farm size in Burundi

Marijn Verschelde* Marijke D'Haese† Glenn Rayp‡
Ellen Vandamme§

Abstract

We use a nonparametric approach to investigate the relationship between farm productivity and farming scale. A Kernel regression is used on data of mixed cropping systems to study the determinants of production including different factors that have been identified in literature as missing variables in the testing of the inverse relationship such as soil quality, location and household heterogeneity. Household data on farm activities and crop production was gathered among 640 households in 2007 in two Northern provinces of Burundi. Five production models were specified each with different control variables. Returns to scale are found to depend on the farm scale. Our results qualify to a large extent the finding of an inverse relationship between farm size and productivity, though without fully explaining it. Other factors that affect significantly positive production include the soil quality and production orientation towards banana or cash crop production. Production seems to be negatively affected by field fragmentation.

Keywords: inverse relationship, farm size, nonparametric, Burundi
JEL classification: D24, O13, Q12, Q18

*Corresponding author, SHERPPA, Department of General Economics, Ghent University, Tweekerkenstraat 2, 9000 Gent, Belgium. E-mail: marijn.verschelde@ugent.be. Tel:+32(0)9/264.35.03, Fax:+32(0)9/264.35.99.

†Department of Agricultural Economics, Ghent University, Coupure links 653, 9000 Gent, Belgium. E-mail: marijke.dhaese@ugent.be.

‡SHERPPA, Department of General Economics, Ghent University, Tweekerkenstraat 2, 9000 Gent, Belgium. E-mail: glenn.rayp@ugent.be.

§Department of Agricultural Economics, Ghent University, Coupure links 653, 9000 Gent, Belgium. E-mail: ellen.vandamme@ugent.be.

1 Introduction

The (possibly inverse) relationship between farm size and land productivity has been heavily debated in literature for decades (see e.g. Wiggins et al. (2010)). Given that it contradicts economic theory, which implies that factor productivity should be equal across farms or between the plots of a single household, all along attempts are made to explain the occurrence of this inverse relationship. Several obvious and less obvious reasons and explanations for the inverse relationship between farm size and farm productivity (IR) have been put forward and tested, but none have yet been able to provide an explanation for the IR.

A first obvious reason is the presence of imperfect factor markets (Feder, 1985). This includes failures in different types of production factor markets: land market (Platteau, 1996; Heltberg, 1998), credit market (Assunção and Ghatak, 2003), insurance market (Dercon and Krishnan, 1996) and labour market (Feder, 1985; Barrett, 1996; Assunção and Braido, 2007). Malfunctioning or a complete absence of these markets will lead to suboptimal resource allocation on farm level implying inefficiencies. An important cause of the presence of imperfect labour markets in developing countries is claimed to be labour supervision cost (Feder, 1985; Lipton, 2010). As hired labour is assumed to be less motivated and effective, it takes more productive family labour to supervise hired labour which decreases overall labour productivity at farm level. This would explain why labour and farm productivity are lower on large farms, which require more hired labour. Assunção and Braido (2007) and Barrett et al. (2010) argue that the imperfect market hypotheses imply the presence of unobservable variation between households that leads to differences in the input intensity levels which are correlated with farm area. Therefore, they add a set of household specific characteristics such as household size, dependency ratio, and gender of the household head in testing the inverse relation between farm size and productivity. However, none of the studies cited up to now has proven household characteristics to solely explain the IR.

A second important explanation questions whether the IR between farm size and productivity emerges (or not) due to omitted variables. Soil quality is mentioned as an important but often neglected explanatory variable. Differences in soil quality lead to differences in soil productivity which clearly affect output (Sen, 1975), with small farmers being more productive because of having plots of better quality. In addition, farming practices and production methods might vary according to farm size, leading to differences in yields and productivity (Byiringiro and Reardon, 1996; Schultz, 1964; Assunção and Braido, 2007; Lipton, 2010). All revised studies on this issue show a decrease in the severity of the IR when controlling for soil quality (Benjamin, 1995; Lamb, 2003; Assunção and Braido, 2007; Barrett et al., 2010), but none has found that the IR disappears when controlling for soil quality. Differentiation in farm management skills as an explanatory variable of farm productivity was tested using panel data in which was allowed for household-specific fixed effects (Lipton, 2010). Though Lipton (2010) argues that differentiation in management was not yet thoroughly tested in empirical research, the existing evidence doesn't point to managerial skills as the determinant which explains the IR.

A third explanation of the IR is related to methodological issues. The debate in literature points to the struggles with methodological problems in proving the IR. As one of the unsettled issues, Lipton (2010) mentions that big farms cannot be considered as linear blowups of small ones. Incentives to use inputs vary with the production scale, i.e. bigger farms use a different technology than small farms. Most empirical studies on the IR are based on cross sectional data used for econometric models that fail to capture for non-linearities and that impose a common specification (parameters) for the whole sample they analyze. Moreover, the scale ranges that are allowed in the models may be too small to measure scale effects (Collier and Dercon, 2009).

This paper addresses in particular the latter issue. We analyze factors influencing farm production including scale using a non-parametric estimation of the production function estimated for a unique dataset in the North of Burundi. By using a nonparametric approach we are able to track heterogeneity in productivity effects of increased access to production factors. Our rich dataset allows controlling for several of the missing variables mentioned above. We account for mixed output by calculating the market values of all crops produced while allowing for mixed cropping systems. The relationship between inputs and farm output – here measured as market value of crop and coffee production – is not linear, which parametric models fail to capture. We find that returns to scale are increasing with the size of the farms and qualify the occurrence of the IR to a large extent. However, for the sample we analyze, we fail to reject the inverse relationship. In the next section we describe the data and methods used in the analysis. The third section presents our estimation results. Conclusions are drawn in the fourth section.

2 Methodology

2.1 Data

Household data on farm activities was gathered in 2007 in two densely populated provinces in the North of Burundi, Ngozi and Muyinga. The provinces were chosen because they are among the most populated of the country. Both provinces cover an area of 2300 km^2 and 1.4 million inhabitants; this is 13% of the total surface of Burundi and 19% of the population. Both provinces are densely populated with 475 inhabitant per km^2 in Ngozi and 322 inhabitants km^2 in Muyinga. Economic activity outside agriculture is very limited in both provinces, except for the city of Ngozi which is the third largest city of Burundi. Burundi has the sad record of being one of the poorest countries in the world. With a GNI of 390$ (PPP) per capita it is ranked at the bottom of the group of low-income countries (WorldBank, 2011). In the Human Development Index ranking of 169 countries, it is at the 166th place (UNDP, 2010). The country seems to have much against it when trying to succeed in promoting economic growth; its size is rather small, it is landlocked, with limited natural resources and it is prone to ethnic conflict. The economy depends largely on agriculture; more than one third of the total GDP is derived from agricultural production and more than 90% of employment is allocated to the agricultural sector. Agriculture also plays a vital role in the trade balance as more than 90% of foreign exchange earnings is derived from the export of coffee although the contribution of this export to the country's GDP is rather small (CIA, 2010).

In total 640 farm households were questioned; 360 in the Nogzi Province and 280 in Muyinga Province. All 16 municipalities of the two provinces were covered (nine in Ngozi Province and seven in Muyinga), per province ten villages where selected based on geographical distribution and in every village four households were randomly selected. The interviews were held in Kirundi in collaboration with a team of the University of Burundi. Because of missing data, 20 farms had to be excluded from the data analysis.

For each household, two questionnaires were used; a first questionnaire collected information on household and farm characteristics. A second questionnaire was used to gather information on each plot the farmer owned. The result is a very rich dataset with detailed and reliable information on farm scale (production level, size, labour input, farm inputs), the farming system (crop choices and cash crops) as well as on the farmer's evaluation of the soil quality, and steepness of the different fields. The latter is particularly important given the area is particularly hilly. In order to avoid measurement error in farm size, positions were measured using GPS (GPSmap 60CSx), which allowed for a measurement of the plot size with a precision of 5 meters.

2.2 Variables included in the model

The output is measured by the sum of the market value of all crops produced irrespective of whether these are sold or consumed by the household. Farm production for each food crop is multiplied by the average market price of the respective crops. The level of marketing by the farmers is so low that no individual farm-gate prices could be captured. Furthermore, the diversity of the mixed cropping produce made it not possible to use other quantity measures. E.g., the alternative of caloric content couldn't be used because it would exclude the possibility to account for the value of coffee production.

Factors influencing production are production factors (land, labour, inputs), while controlling for location, farm management, soil quality and household characteristics. As land input, the farm area that is actually used for cultivating food and cash crops is included. Two different sources of labour are distinguished, namely family labour (expressed in person units) and hired labour (expressed in paid wages). One other type of non-labour inputs is included: the sum of the expenditure on seed, chemicals and agricultural equipment.

Four different types of control variables are included: location, farm management, soil quality and household heterogeneity. Location is considered by adding a dummy for the province. As the capital of the Ngozi province is the third largest city in Burundi, access to assets and markets in this province might be significantly higher than in Muyinga. Indicators for farm management are the cropping pattern, fragmentation index and production technology used. A mixed cropping pattern is quantified by the share of the total cropping surface used for either: staple crops, cash crops, banana or other crops. Land fragmentation is assessed by the Simpson index. This index varies from zero to one and is calculated by dividing the total sum of the different field surfaces squared by the square of total cropping area ($S = \sum s_i^2 / (\sum s_i)^2$). Farms with higher land fragmentation will demonstrate a higher Simpson index. Two dummies are included to account for the use of chemicals and animal manure as soil improving farming techniques. Farmers were asked to assess the steepness of the plot and soil quality of each of their plots on a scale from one to four. This resulted in the calculation of two variables, one variable that indicates the share of the total cropping surface that has a steep slope and a second variable representing the share of the total cropping surface with good quality soil.

Finally, we control for household heterogeneity by including the following variables: age of the household head, the share of household income derived from off-farm activities and a dummy for extension (whether or not the household has been visited by an extension officer). A descriptive analysis for all variables included in the model is given in Table 1.

Variables	Ngozi province		Muyinga province		Entire sample		Test
							t-test
Agricultural output (1,000BIF)	1029.67	(1062.04)	787.60	(948.41)	921.13	(1019.01)	2.99**
Farm size (ha)	0.87	(1.44)	1.29	(1.89)	1.13	(1.66)	-2.26**
Farm size per person (ha/pers)	0.18	(0.24)	0.25	(0.35)	0.21	(0.29)	-2.68**
Size cultivated land (ha)	0.76	(1.1)	0.99	(1.45)	0.86	(0.52)	-2.12**
Size cultivated land per person (ha/pers)	0.14	(0.20)	0.19	(0.29)	0.16	(0.25)	-2.59**
Family labour (nb)	2.74	(1.34)	2.51	(1.10)	2.64	(1.24)	2.30**
Labour cost (paid wage, 1,000BIF)	39.34	(13.66)	23.91	(100.77)	32.42	(118.35)	1.66**
Cost for seeds (1,000BIF)	20.46	(34.00)	17.62	(20.70)	19.18	(28.82)	1.28
Costs for chemicals (1,000BIF)	8.45	(20.56)	1.10	(5.98)	5.16	(16.19)	6.29**
Costs for agricultural material (1,000BIF)	4.47	(9.65)	3.76	(6.87)	4.15	(8.52)	1.02
Total cost production inputs (1,000BIF)	33.38	(48.38)	22.49	(25.00)	28.49	(39.98)	3.61**
Share staple crops (%)	52.51	(19.57)	61.88	(18.81)	56.71	(19.78)	-6.04**
Share coffee (%)	13.77	(13.62)	9.22	(10.71)	11.73	(12.60)	4.65**
Share banana (%)	20.78	(14.60)	18.05	(12.29)	19.55	(13.67)	2.53**
Share non-productive land use (%)	12.93	(17.27)	10.84	(17.02)	11.99	(17.18)	1.52
Share in the marsh (%)	9.33	(12.28)	2.87	(6.29)	6.40	(10.54)	8.46**
Share under steep slope (%)	20.52	(29.85)	17.57	(29.59)	19.20	(29.75)	1.23
Share good quality soil (%)	49.51	(37.53)	46.49	(41.43)	48.15	(39.32)	0.94
Fragmentation index (0-1)	0.23	(0.14)	0.24	(0.14)	0.24	(0.14)	-0.51
Age of hhhead (years)	41.36	(12.41)	40.01	(12.89)	40.75	(12.64)	1.32
Share income off-farm (%)	37.45	(3.59)	39.16	(32.04)	38.22	(32.33)	-0.65
							χ^2-test
Use of chemicals (% yes)	83		65		75		26.27**
Use of animal manure (% yes)	61		49		56		9.78**
Extension visit (% yes)	21		57		37		82.62**
Observations	342		278		620		

Significance levels : $*$: 5% $**$: 1% $***$: 0.1%
Note: t-values test for differences between the province means

Table 1: Descriptive analysis dependent, independent and control variables included in model

2.3 Nonparametric regression approach

The empirical model is defined by a $n \times 1$ dependent scalar Y (farm output), a $n \times q$ multivariate regressor X (inputs) and an additive error ϵ.

$$Y_i = g(X_i) + \epsilon_i \text{ , with } i = 1, ..., n \tag{1}$$

This production function can be estimated by imposing a parametric form. The vast majority of papers impose a Cobb-Douglas (CD) specification. Log output is defined as a linear function of the log of the q regressors, with additive error.

$$\ln Y_i = \alpha + \sum_{k=1}^{q} \beta_k \ln X_{ik} + \epsilon_i \tag{2}$$

It is from the assumed Cobb-Douglas specification of the production function that Assunção and Braido (2007) derive the yield approach that they prefer. The Cobb-Douglas specification implies that the factor elasticities are independent from scale and hence equal for all farms. However, if there are non-linearities or interactions in the true model, the empirical model is misspecified and the coefficients are inconsistent under a log-linear specification (Henderson and Kumbhakar, 2006). A flexible parametric alternative is the translog specification; quadratic effects and interaction effects are introduced in the empirical model.

$$\ln Y_i = \alpha + \sum_{k=1}^{q} \beta_k \ln X_{ik} + 0.5 \sum_{k=1}^{q} \sum_{l=1}^{q} \beta_{kl} \ln X_{ik} \ln X_{il} + \epsilon_i \tag{3}$$

In some cases, the translog specification can give economically unreasonable estimates, caused by (1) failure to capture all nonlinearities in the true model (Henderson and Kumbhakar, 2006), and (2) the high multicollinearity or low degrees of freedom as result of the inclusion of quadratic effects and interactions.

In order to allow for farm-specific input returns and yet avoiding to impose *'a priori'* a functional relationship between the output scalar and regressors, nonparametric approaches can be used[1]. In a nonparametric (generalized) kernel regression, $E[Y|X = x]$ is estimated by means of a localized regression (one could note it as $\hat{g}(x) = E[Y|\text{X close to x}]$).

Kernel weight functions are used to give more weight to observations near the observation point. Window widths impose the window of localization. If the window is large, the

[1] See Li and Racine (2007) for an extensive overview of the used kernel regression approach

curve will be a smooth straight line. On the other hand, if the window width is small, non-linearities are allowed for and the curve becomes less smooth. It is intuitively clear and shown in literature that the choice of the weighting function is of far less importance than the choice of the window of localization - which we will discuss below.

We use kernel weights (l^c, l^u, l^o) with window widths $(\lambda^c, \lambda^u, \lambda^o)$ to specify the weight function for $x = [x^c, x^u, x^o]$, where x^c is a vector of continuous values, x^u is a vector of unordered discrete values, x^o is a vector of ordered discrete values. In specific, we specify a standard normal kernel function l^c to weight the continuous variable x_k^c (see (4)). An Aitchison and Aitken (1976) kernel l^u is specified to weight discrete unordered variable x_l^u with c_l categories and $\lambda_l^u \in [0, (c_l - 1)/c_l]$ (see (5)). To weight the ordered discrete value x_m^o, we use a Wang and van Ryzin (1981) kernel function with $\lambda_m^o \in [0, 1]$ (see (6)).

$$l^c\left(\frac{X_{ik}^c - x_k^c}{\lambda_k^c}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_{ik}^c - x_k^c}{\lambda_k^c}\right)^2} \tag{4}$$

$$l^u(X_{il}^u, x_l^u, \lambda_l^u) = \begin{cases} 1 - \lambda_l^u \text{ if } X_{il}^u = x_l^u, \\ \lambda_l^u/(c_l - 1) \text{ otherwise} \end{cases} \tag{5}$$

$$l^o(X_{im}^o, x_m^o, \lambda_m^o) = \begin{cases} 1 \text{ if } X_{im}^o = x_m^o, \\ (\lambda_m^o)^{|X_{im}^o - x_m^o|} \text{ otherwise} \end{cases} \tag{6}$$

To allow for a multivariate regression, we use - as is common practice - product kernels. The product kernel of x^c is $W_{\lambda^c}(X_i^c, x^c) = \prod_{k=1}^q (\lambda_k^c)^{-1} l^c((X_{ik}^c - x_k^c)/\lambda_k^c)$. For x^u, the product kernel is defined as $L_{\lambda^u}(X_i^u, x^u) = \prod_{l=1}^r l^u(X_{il}^u, x_l^u, \lambda_l^u)$. The product kernel of x^o is $L_{\lambda^o}(X_i^o, x^o) = \prod_{m=1}^s l^o(X_{im}^o, x_m^o, \lambda_m^o)$. All together, we can specify a Racine and Li (2004) generalized kernel function as $\mathcal{K}_\gamma(X_i, x) = W_{\lambda^c}(X_i^c, x^c) L_{\lambda^u}(X_i^u, x^u) L_{\lambda^o}(X_i^o, x^o)$, with $\gamma = (\lambda^c, \lambda^u, \lambda^o)$.

Two approaches were considered to estimate $E(Y|X = x)$. First, the Nadaraya-Watson estimator, which takes the kernel weighted average of the observed Y_i values and normalizes it by the sum of the kernel weighted averages (see (7)). This is the so called local-constant approach as it specifies a locally averaged constant value y for each observation point. It can be obtained as the solution of a in (8). Second, the local-linear estimator, which estimates a local linear relation for each observation point by obtaining a and b in (9). If bandwidths are very large in a local-linear regression and there is thus no local weighting, we have the parametric least squares estimator. The least squares estimator can thus be seen as a special case of the local-linear estimator (Li and Racine, 2007, p. 83). We opt for the local-linear regression as it has better boundary properties than the local-constant regression (Hall et al., 2007) and nests least squares as a special case.

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i \mathcal{K}_\gamma(X_i, x)}{\sum_{i=1}^{n} \mathcal{K}_\gamma(X_i, x)} \tag{7}$$

$$\min_{a} \sum_{i=1}^{n} (Y_i - a)^2 \mathcal{K}_\gamma(X_i, x) \tag{8}$$

$$\min_{\{a,b\}} \sum_{i=1}^{n} (Y_i - a - (X_i - x)'b)^2 \mathcal{K}_\gamma(X_i, x) \tag{9}$$

As discussed, the choice of multivariate bandwidth γ is of crucial importance. We opt for the often used data-driven approach that minimizes the asymptotic integrated mean squared error (AIMSE): the least-squares cross-validation approach as defined in (10).

$$CV(\gamma) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{g}_{-i}(X_i))^2 t(X_i) \tag{10}$$

where $\hat{g}_{-i}$ is the leave-one-out local-linear kernel estimator of $E(Y_i|X_i)$, and $0 \leq t(\cdot) \leq 1$ is a weight function that serves to avoid difficulties caused by dividing by 0 or by the slower convergence rate arising when X_i lies near the boundary of the support of X. Simulation results of Li and Racine (2004) show that cross-validated local-linear regressions indeed choose much larger bandwidths if the true relationship is linear.[2]

3 Results

3.1 Description of the farming system related to farm size

The farming system in Burundi consists of small peasant landholdings (of generally less than 1 ha per family as illustrated in Figure 1), very small plots with double cropping,

[2]We opt for this approach over the AIC CV approach as the least-squares CV approach is more used in the literature and is faster to compute.

manual self-subsistence farming with little marketed surplus (Cochet, 2004). Crop production is done on both the hill side and in the drained marshes. Two distinct cropping systems were distinguished on each landholding. A first system consisted of separate plots cultivated with mixed crops (grains, pulses, tubers and coffee), and, a second system was based on banana production (see also Cochet (2004)). Several authors emphasize the importance of banana production in the current farming system (Rishirumuhirwa and Roose, 1998; Cochet, 2004). It seems as if the banana has over the years replaced cattle production which requires more land and other natural resources. The most important food crops produced and consumed in the study area were sweet potatoes, beans, cassava, banana and flour of maize (FAOSTAT, country profile, 2005). Except for banana and coffee, most farmers did not market produce and even when they did sell, it was mainly surplus sales of very small quantities.

The average farm size in our sample was 1.12ha however about 45% of the farms in the sample were smaller than 0.5ha. Farms were larger in Muyinga compared to the more densely populated Ngozi Province (see Table 1). The distribution of land over the sample was rather unequal. Moreover, compared to a previous survey and with Rishirumuhirwa and Roose (1998), we find an increase of inequality in access to land, which resulted in an increased number of very small scale farms (smaller than 0.5ha). Furthermore farms were highly fragmented with on average more than eight plots on the hillside (collines), and one to two plots in the swamps (marsch). The relatively large farms in our sample are deliberately not excluded from the analysis as they may contain valuable information which can be studied separately with a nonparametric model.

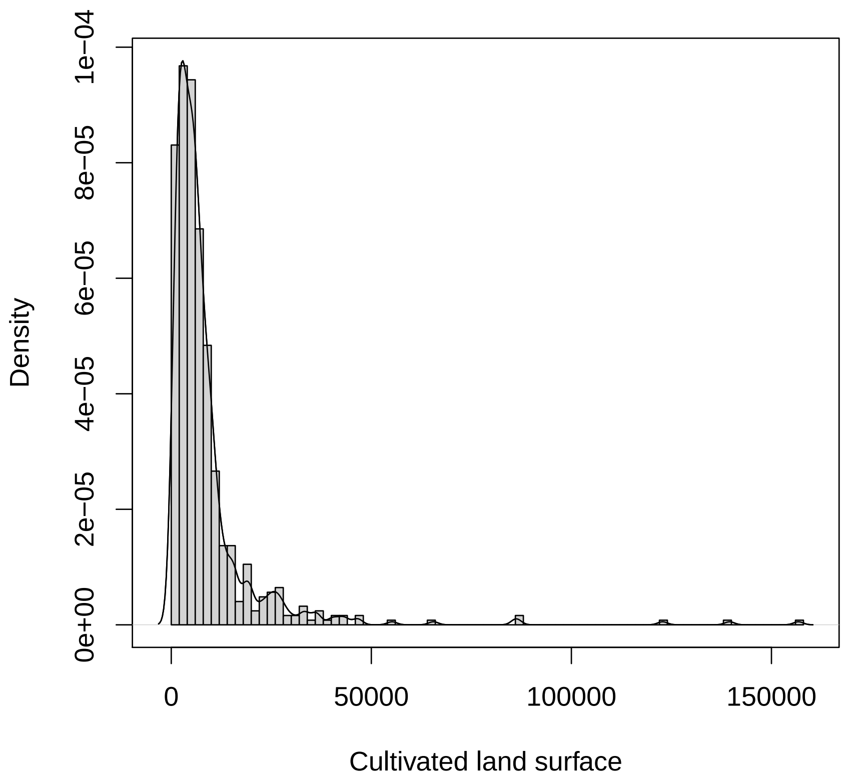


Figure 1: Density plot of farm sizes in the sample (m^2)

Symptomatic for the very poor livelihoods of the farm households in the study area, was their high level of food insecurity. In our survey, we registered the HFIAS score (Household Food Insecurity Access Score, USAID, Coates et al., 2007), which shows that only 7% of the households could be considered food secure (results not shown in the table). Two thirds of all households interviewed were even labelled severely food insecure. These figures coincide with FAO data indicating that 68% of the total population is undernourished (FAO, 2009).

Results presented in Table 2 suggest that farm size, production strategy, crop productivity and farm production may be related, although not all effects tend to go into the same direction. Large farms showed slightly different land use patterns compared to small farms. Larger farms tended to attribute a larger share of their total farm surface to other non-production activities such as forestry and fallow land whereas small farms used most of their land for staple food production rather intensively. However, the share of production area dedicated to cash crops, i.e. coffee production, did not significantly differ according to farm size quartiles. Small farms were using a larger proportion of the total production surface for banana production while larger farms used relatively more land for bean production (not detailed in the table). Farm proportions dedicated to other important crops in the area such as tubers and cereals did not differ between the land size quartiles and are therefore not reported. Crop diversification seems to be larger on larger farms, which is supposed to make them less prone to risks of crop failure compared to small less diversified farms.

Variables	First quartile		Second quartile		Third quartile		Fourth quartile		Test
									F-stat
Agricultural output (1,000BIF)	429.55	(344.18)	601.84	(370.03)	902.33	(616.38)	1750.80	(1580.97)	68.06**
Farm size (ha)	0.22	(0.11)	0.56	(0.24)	0.91	(0.33)	2.82	(2.62)	119.46**
Farm size per person (ha/pers)	0.06	(0.05)	0.12	(0.08)	0.18	(0.12)	0.48	(0.48)	84.08**
Size cultivated land (ha)	0.16	(0.07)	0.41	(0.08)	0.71	(0.11)	2.2	(2.1)	111.95**
Size cultivated land per person(ha/pers)	0.05	(0.05)	0.13	(0.09)	0.18	(0.12)	0.48	(0.48)	84.08**
Family labour (nb)	2.15	(0.68)	2.56	(1.17)	2.86	(1.39)	2.97	(1.43)	14.40**
Labour cost (paid wage, 1,000BIF)	7.34	(29.60)	10.22	(24.63)	22.04	(44.64)	90.08	(219.64)	18.19**
Seed cost (1,000BIF)	13.06	(21.05)	15.60	(17.99)	19.95	(20.23)	28.12	(44.11)	8.44**
Costs for chemicals (1,000BIF)	1.60	(0.05)	3.72	(11.21)	4.86	(14.90)	10.49	(25.32)	8.91**
Costs for material (1,000BIF)	2.59	(3.58)	3.57	(6.82)	4.32	(6.94)	6.13	(13.32)	4.86**
Total cost inputs (1,000BIF)	17.21	(23.44)	22.89	(29.08)	29.13	(32.54)	44.75	(59.54)	14.58**
Labour cost per ha (1,000BIF/ha)	28.19	(109.9)	17.80	(37.94)	23.45	(47.16)	34.78	(100.38)	1.25
Seed cost per ha (1,000BIF/ha)	72.40	(115.96)	28.71	(30.72)	23.46	(25.53)	14.61	(28.85)	25.87**
Costs chemicals per ha (1,000BIF/ha)	7.11	(19.23)	6.52	(20.07)	5.09	(16.11)	4.60	(12.17)	0.73
Costs material per ha (1,000BIF/ha)	14.86	(23.65)	6.36	(8.98)	4.99	(8.35)	2.76	(5.83)	21.32**
Total cost inputs per ha (1,000BIF/ha)	94.37	(131.49)	41.59	(46.03)	33.55	(34.80)	21.97	(36.38)	28.99**
Share staple crops (%)	54.88	(21.27)	54.72	(19.41)	61.13	(18.63)	56.12	(19.20)	3.63**
Share coffee (%)	12.04	(14.45)	12.54	(11.61)	10.48	(11.28)	11.87	(12.84)	0.76
Share of banana (%)	22.53	(14.83)	19.67	(14.78)	18.04	(11.26)	17.99	(13.12)	3.82**
Share of non-productive land use (%)	10.55	(17.15)	13.07	(18.56)	10.35	(15.70)	14.01	(17.05)	1.76
Share in the marsh (%)	8.32	(13.83)	5.94	(8.87)	5.73	(9.48)	5.78	(9.05)	2.22.**
Share under steep slope (%)	18.65	(30.79)	20.30	(30.53)	16.37	(27.03)	21.48	(30.52)	0.86
Share good quality soil (%)	44.56	(38.74)	43.38	(38.34)	47.11	(38.92)	57.56	(40.04)	4.26**
Fragmentation index	0.30	(0.16)	0.23	(0.12)	0.23	(0.13)	0.19	(0.13)	17.27**
Age of hhhead (years)	37.00	(11.37)	40.15	(12.39)	42.24	(13.46)	43.63	(12.37)	8.36**
Share income off-farm (%)	44.07	(33.71)	40.92	(33.65)	37.90	(31.12)	30.05	(29.26)	5.46**
									χ^2-test
Use of chemicals (% yes)	65.2		75.5		72.9		85.2		16.74**
Use of manure (% yes)	40.6		54.8		58.1		68.4		24.71**
Extension visit (% yes)	25.8		34.2		45.2		43.9		16.36**
Observations	155		155		155		155		

Significance levels : $*$: 5% $**$: 1% $***$: 0.1%

Table 2: Descriptive analysis for different quartiles of farm size (N=620)

The allocation of labour seems to be closely related to farm size with larger farms allocating more family labour and spending more funds on hired labour. However, the level of labour per land unit was significantly higher for smaller farms as family labour per land unit was larger for small farms and wages paid for hired labour per land unit were not significantly different from larger farms. Investments in agricultural production were measured by the expenditure on seed, agricultural material and chemicals. These investments increased significantly with increasing production area. Smaller farms spent significantly more money per land unit on seed and agricultural material. Investments in chemicals such as fertilizer and pesticides were not different across the land size quartiles; these chemicals were used with the same, generally very low, intensity on both small and large farms. However the likelihood of using chemicals was larger on larger farms. On top of this, the likelihood of using specific soil improving techniques (manure, compost, mulching) was higher for the quartile with the largest farms. These findings suggest differences in the production strategies related to differences in cropping area. These differences in crop production strategies might lead to different production outcomes and even more so to differences in farm productivity.

3.2 Results on Farm Size and Productivity from the Kernel Estimations

The nonparametric approach makes no *'a priori'* assumptions on the functional relationship between the dependent variable and regressors. Using cross-validation, the trade-off between bias (for a given model, larger for a smooth, linear curve) and variance (larger for a wiggly, non-linear curve) is settled. As there is too few variation in family labour, it is inappropriate to consider this as a continuous variable. Therefore, we define family labour as an ordered discrete variable.[3] In contrast to the parametric models, an ordered discrete variable can be included as one variable in a nonparametric model.

We illustrate the nonparametric results by showing directly the estimated level of output as a function of the value of a respective independent variable, holding the other regressors equal to respectively the median for continuous variables or modus for (ordered) discrete variables. In addition, we show 95% bootstrap confidence intervals. A significantly increasing (resp. decreasing) curve illustrates a significant positive (resp. negative) effect of the regressor on agricultural production.[4]

The base model includes as independent variables, size land used for agricultural production, family labour, cost of hired labour, cost of inputs used, and a dummy for the province

[3] Results do no alter when we consider family labour as a continuous variable.

[4] The nonparametric model allows for interactions between all regressors. 3-D plots of estimated interactions between regressors are available on request.

(see Figure 2).[5] The model shows significant effects of cultivated land and cost of hired labour. The model confirms that production was higher in Ngozi compared to Muyinga. An increase in family labour did not significantly contribute to production, indicating a very low (zero) marginal productivity of family labour. There is a clear non-linear relationship between hired labour and agricultural output.

Because of the high correlation (0.44) between land surface and hired labour, the effects of the two variables are difficult to disentangle. The farm scale is therefore considered as a combination of both.[6] In Figure 7(a) and in Table 4 in Appendix, we define the scale of the farm by the respective quantiles of hired labour and land surface used for production. A scale of 0 (resp. 1) means that the farm uses the minimum (resp. maximum) level of hired labour (larger than 0) and the minimum (resp. maximum) surface for production found in the data. Figure 7(a) illustrates that returns to scale of hired labour and land surface are a function of the scale of the farm. Relatively small farms are found to have returns to scale close to 0. Relatively large farms have returns to scale not far below 1. The assumption that returns to scale are not scale dependent - as imposed in the CD model and shown by the horizontal black line - is thus rejected at the 95% confidence interval.

A second model checks for the influence of spatial effects within the provinces (see Figure 3). The results show a clear difference between the first eight villages that are in the Muyinga Province and the eight villages in the Ngozi Province. This confirms the effects of the Provincial dummy found in the base model.

In a third model, we control for land use (see Figure 4). The effects of cultivated land, costs for hired labour and intermediary inputs, and location are similar as for the base model. Farms with a larger share of banana cultivation are found to have a higher agricultural output. As the only cash crop, the share of coffee planted contributes positively

[5]Preliminary analysis showed that inclusion of dummies for 'no labour cost' and 'no costs intermediary inputs' does not improve the localized model as the dummies are 'smoothed out' (bandwidth equal to 1). Table 3 in Appendix summarizes the used LSCV bandwidths. We make use of the 'np' package in R of Hayfield and Racine (2008).

[6]We do not consider the scale effect of intermediary inputs in this analysis because 1) the use of intermediary inputs is not highly correlated to land surface (correlation of 0.12) and 2) the effect of intermediary inputs is insignificant. It should be mentioned that both the physical and economic access to intermediary inputs are rather problematic in the area studied.

to production. Again, Figure 7(c) shows that returns to scale are scale dependent.
Model 4 checks for the effects of field characteristics such as the steepness of the plots, perceived soil quality, share of land in marches, application of manure and chemical fertilizers, plot fragmentation (see Figure 5). Steepness of the plots is particularly relevant for this hilly environment. The share of the farm located in the marches is of importance for the production of vegetables. The marches are drained and mostly used for vegetable production. Fragmentation is an important problem. The average number of plots on the farms in the sample is 6.6, with the largest quartile having on average eight plots. We find a non-significant negative effect of steepness of the plots. Fragmentation has a significant non-linear effect at the 90% confidence interval. Perceived soil quality is found to be highly significant. Field characteristics are thus important determinants of agricultural production. The results of the base model concerning the inputs hold. We find a non-linear effect of hired labour on agricultural production and returns to scale that are dependent of the sale of the farm (see Figure 7(d)).
Finally a fifth model checks the effect of off-farm income in total household income, the access to extension services and the age of the head of the farm household (see Figure 6). We do not find significant effects of the three variables. The effect of farm size cultivated is not significant in this model. In contrast to the previous models, we find a significant non-linear positive effect for intermediary inputs in this fourth model. However, as the three added variables are not significant, the model should be interpreted with caution. If we drop the three variables, we return to the base model with a significant effect of land surface and a strong non-linear effect of hired labour. Again, model 4 finds that returns to scale are dependent on the scale of the farm (see Figure 7(e)).

In sum, based on this sample of small scale farms, we find clear indications of the size dependency of returns to scale. In Figure 7 this can also be seen by comparing the results of the generalized kernel estimation of the five models with the estimated scale parameter assuming a Cobb-Douglas parametric specification, shown by the horizontal straight line. For the largest firms, defined in terms of land size and the number of hired labourers combined, the returns to scale are significantly higher than would be estimated assuming parameter constancy. In addition, for these firms the returns to scale are almost constant, whereas for the smallest firms, the returns to scale are about 0.2, again significantly lower than the estimation assuming returns to scale independent from the farm size. Hence, our estimations results qualify the occurence of the Inverse Relationship to a substantial extent, yet without fully rejecting it. In addition, we find strong effects of crop choice and field characteristics. The agricultural returns from small-scale fragmented production on low soil quality plots are expected to be very low.

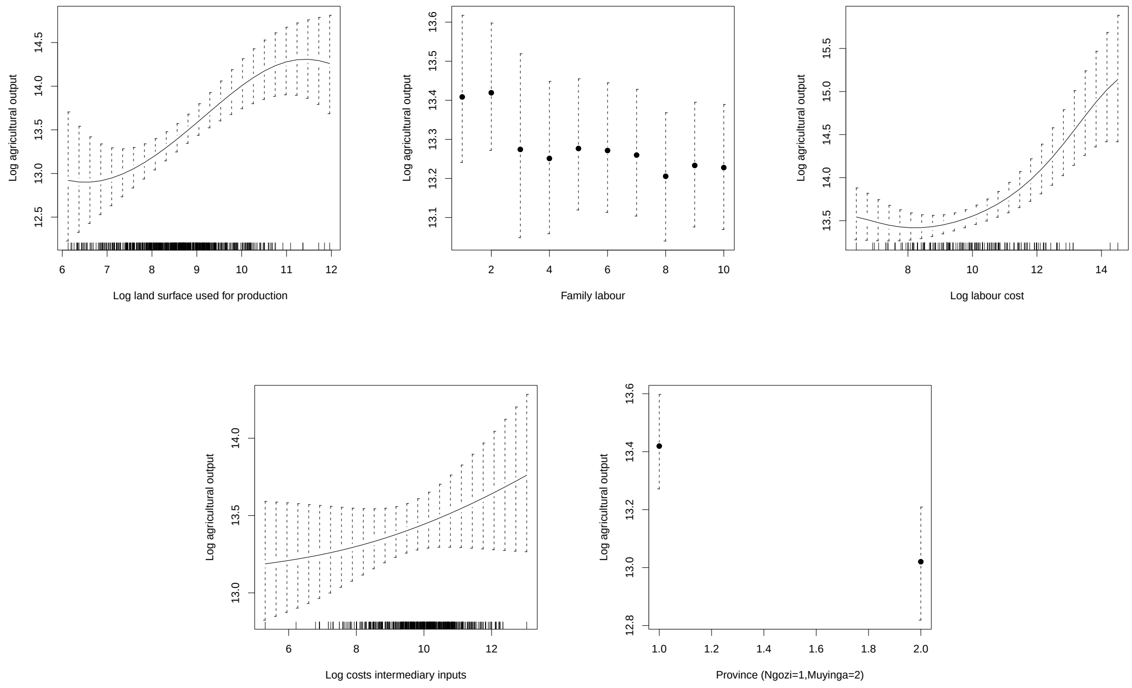


Figure 2: Base Model

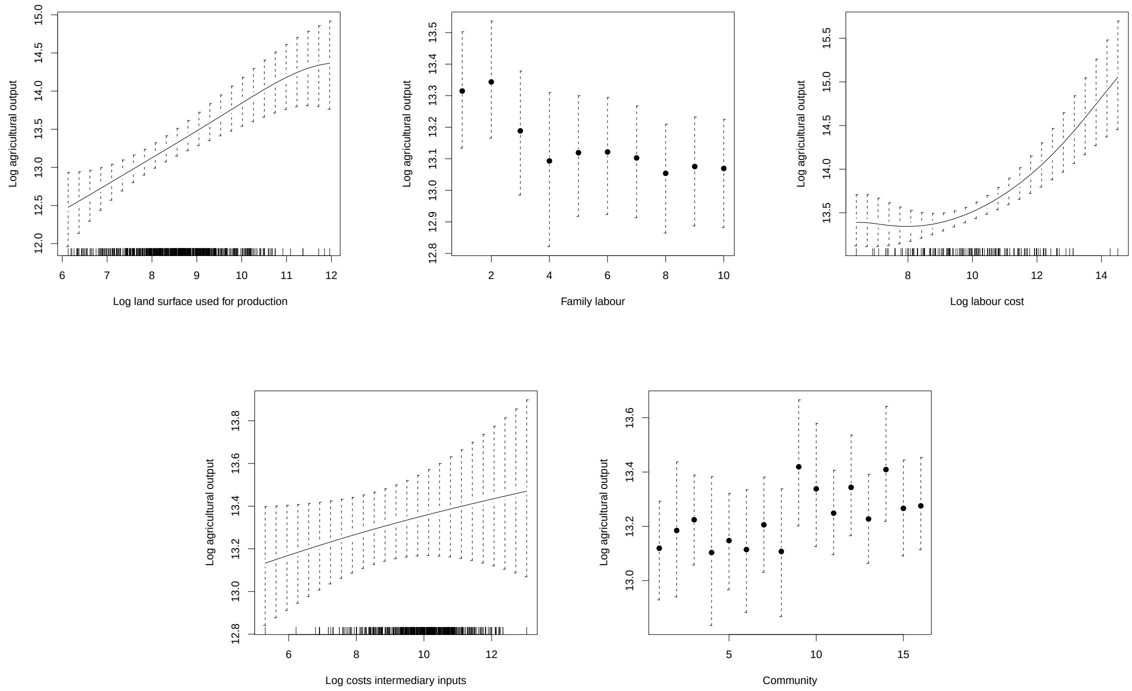


Figure 3: Model 2

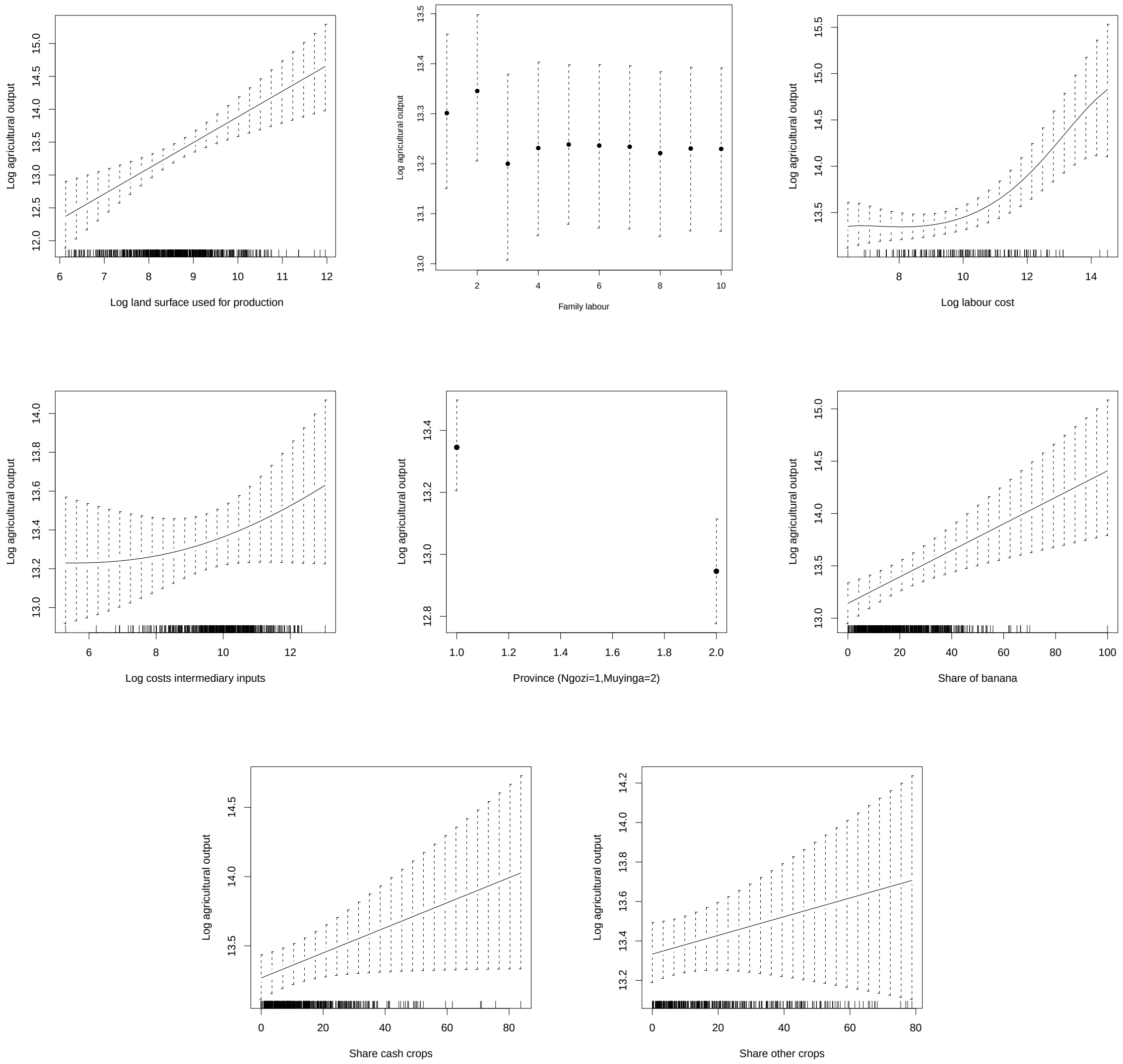


Figure 4: Model 3

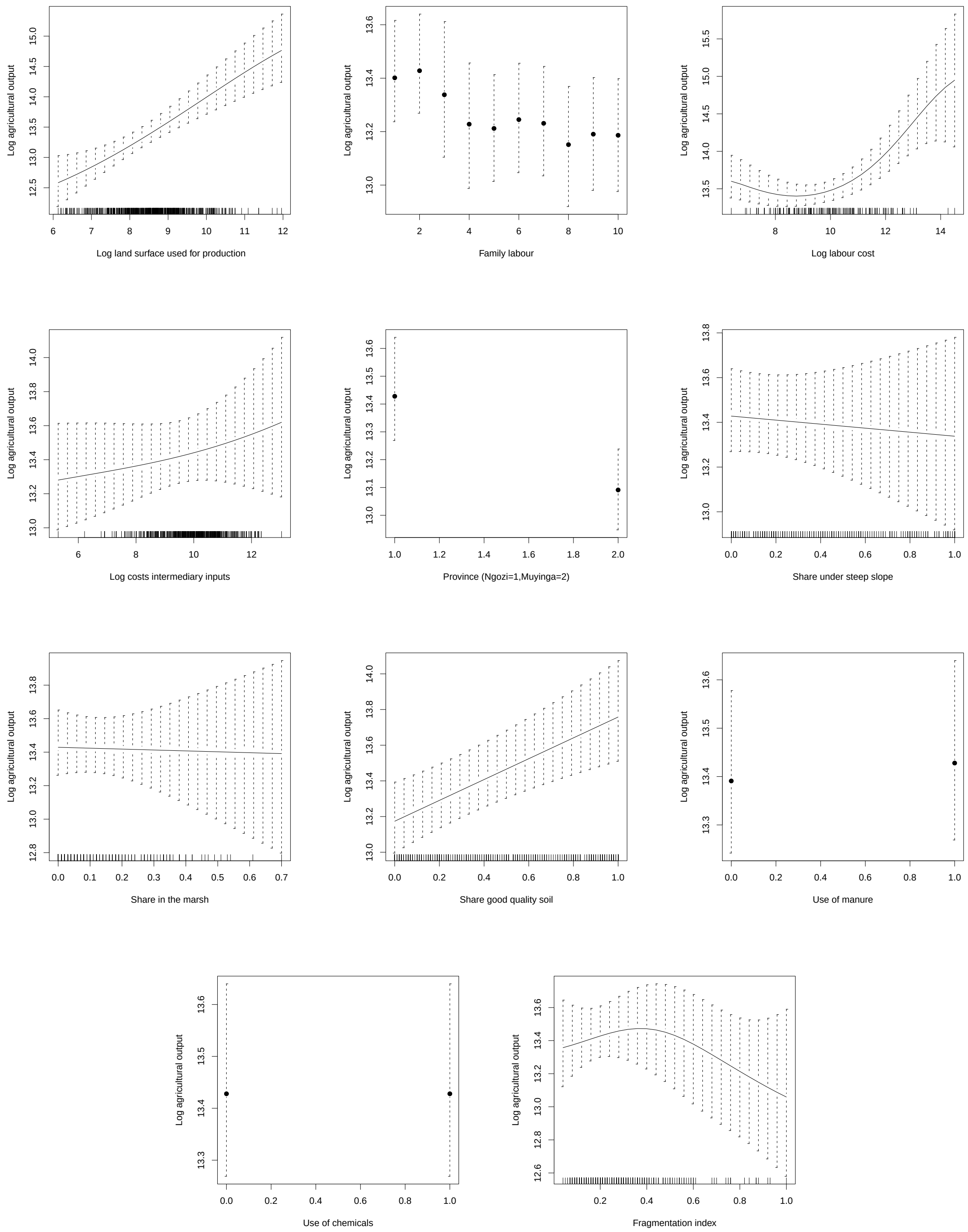


Figure 5: Model 4

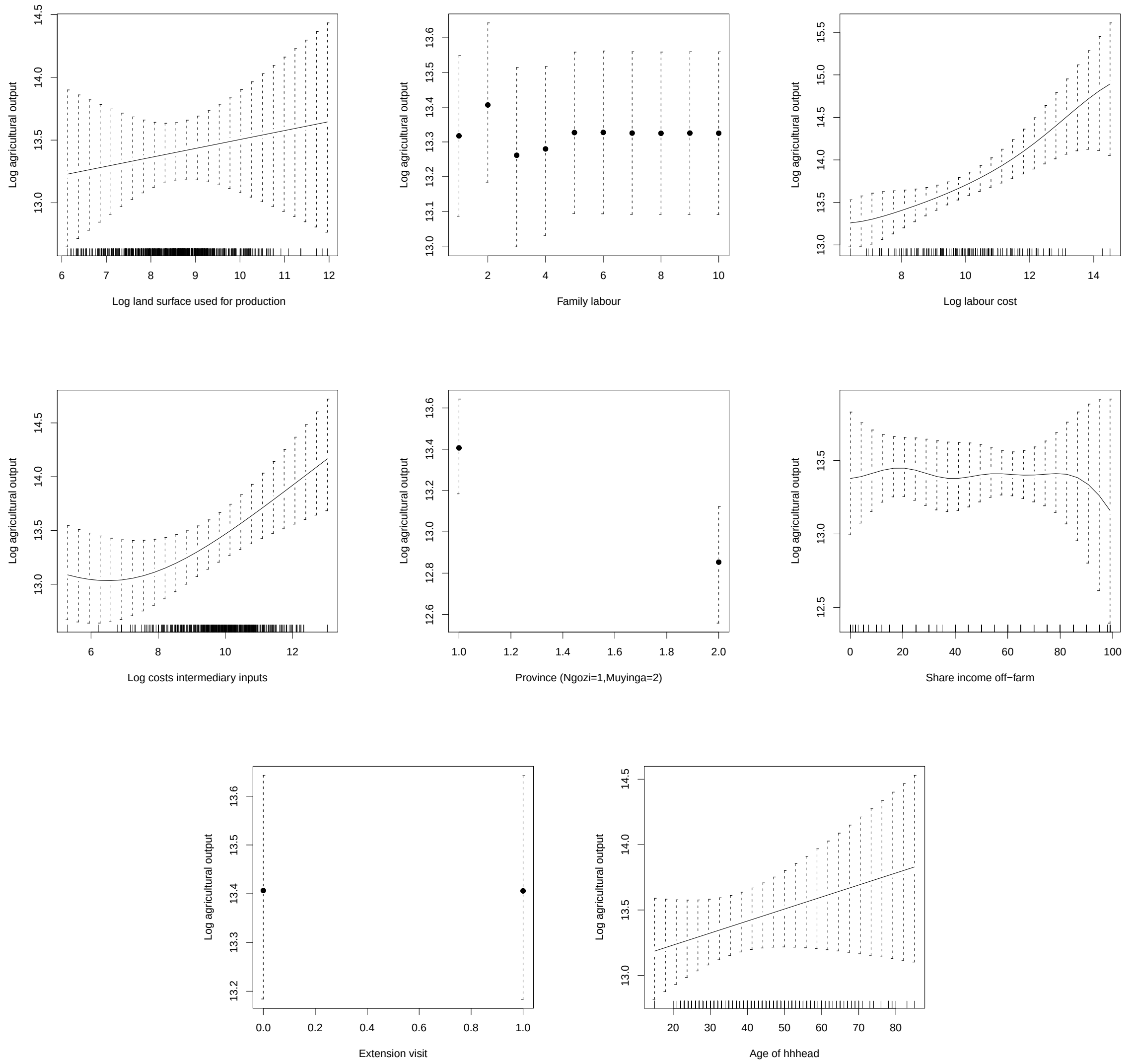
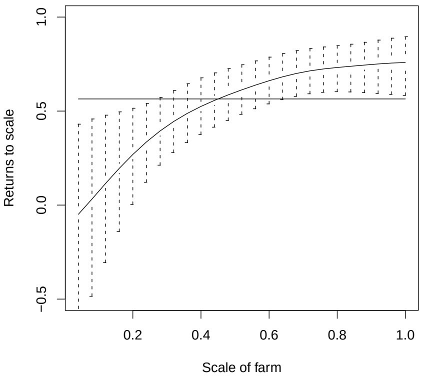
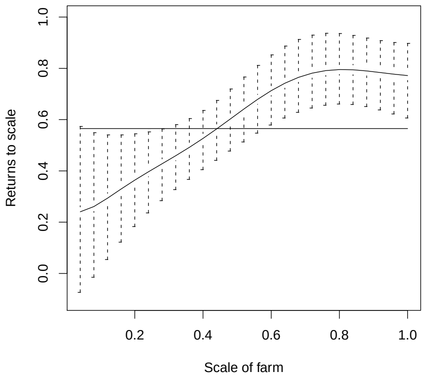


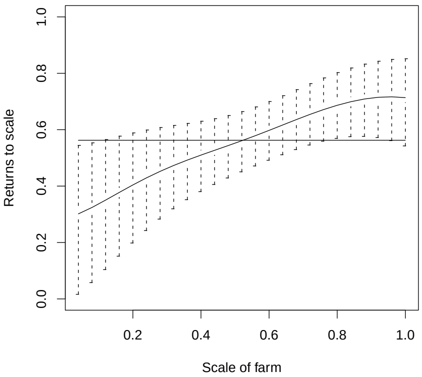
Figure 6: Model 5



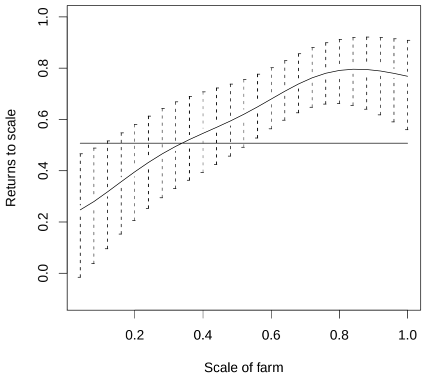
(a) Base model



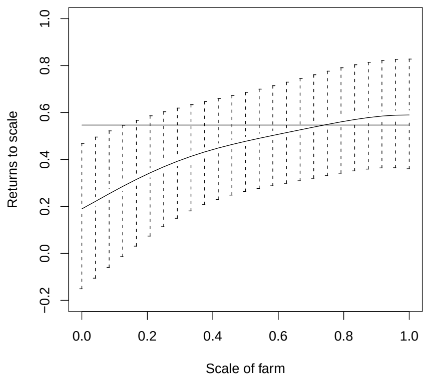
(b) Model2



(c) Model 3



(d) Model 4



(e) Model 5

Figure 7: Returns to scale in function of scale of farm

4 Conclusions

The possibly inverse relationship between farm size and land productivity is one of the most persistent puzzles in development economics, even more so as many potential determinants have been put forward and tested without being able to provide an explanation. Using data on small scale farm holding in two Northern provinces of Burundi, which included information on the missing variables to which is referred in the literature, we study whether returns to scale are non-constant and whether this could contribute to the explanation of the occurrence of the IR.

We used a nonparametric kernel estimation of the production function (solved with a local-linear estimator) to allow for non-linearities and interaction effects. Five different models were estimated controlling for inputs, household, farm and soil characteristics. In each model the effect of size of cultivated land, cost of intermediary inputs and of hired labour was consistent. We find a significant effect of land size and a non-linear effect of hired labour on agricultural output. In addition, crops choice and field characteristics matter. Coffee and banana production are found to yield higher returns compared to the other crops. Fragmentation and low perceived soil quality are associated with low agricultural productivity.

The models confirm that farm size itself matters for the relationship between its size and productivity. Scale elasticity varies between 0.2 for the smallest farms and 0.8 for the largest farms and the assumption of a constant scale elasticity over the whole size range is rejected. Our findings confirm the higher land productivity of the very small farms, but they also show that larger farms may exploit economies of scale. In this sense we qualify the occurrence of the inverse relationship between farm size and land productivity, yet without fully explaining it. However, as returns to scale are scale dependent, it is possible that unobserved very large farms with low field fragmentation and adapted crop production strategies realize constant or increasing returns to scale, as argued by Collier and Dercon (2009).

5 Appendix

	Model I	Model II	Model III	Model IV	Model V
Log cultivated land	1.26	1.43	$5.08e^5$	1.98	$2.89e^5$
Family labour	0.56	0.56	0.81	0.61	1.00
Log hired labour cost	1.62	1.75	1.66	1.61	2.03
Log costs intermediary inputs	4.93	5.92	5.40	5.05	4.04
Province	0.17		0.14	0.22	0.12
Community		0.69			
Share of banana			$1.70e^7$		
Share of cash crops			$2.13e^6$		
Share other crops			$2.17e^7$		
Share under steep slope				$2.55e^5$	
Share in the marsh				$2.83e^5$	
Share good quality soil				$1.26e^5$	
Use of manure				0.36	
Use of chemicals				0.50	
Fragmentation index				0.26	
Share income-off-farm					12.81
Extension visit					0.50
Age of hhhead					$1.47e^6$

Table 3: Bandwidths

	Model I	Model II	Model III	Model IV	Model V
	5% quantile				
Log cultivated land	0.26	0.25	0.38	0.39	0.06
	(0.06)	(0.04)	(0.00)	(0.04)	(0.00)
Log hired labour cost	-0.03	0.05	0.03	0.00	0.12
	(0.05)	(0.03)	(0.03)	(0.04)	(2.85)
Log costs intermediary inputs	0.11	0.07	0.06	0.05	0.20
	(0.02)	(0.01)	(0.01)	(0.01)	(1.43)
	25% quantile				
Log cultivated land	0.40	0.33	0.42	0.43	0.10
	(0.04)	(0.03)	(0.00)	(0.03)	(0.00)
Log hired labour cost	0.06	0.13	0.06	0.08	0.14
	(0.03)	(0.02)	(0.03)	(0.03)	(2.61)
Log costs intermediary inputs	0.09	0.04	0.05	0.04	0.19
	(0.01)	(0.01)	(0.01)	(0.01)	(1.31)
	50% quantile				
Log cultivated land	0.44	0.38	0.43	0.44	0.14
	(0.04)	(0.03)	(0.00)	(0.03)	(0.00)
Log hired labour cost	0.13	0.19	0.11	0.14	0.16
	(0.03)	(0.02)	(0.03)	(0.03)	(2.63)
Log costs intermediary inputs	0.05	0.01	0.04	0.02	0.16
	(0.01)	(0.01)	(0.01)	(0.01)	(1.32)
	75% quantile				
Log cultivated land	0.44	0.42	0.43	0.45	0.18
	(0.05)	(0.03)	(0.00)	(0.03)	(0.00)
Log hired labour cost	0.20	0.23	0.16	0.21	0.19
	(0.04)	(0.02)	(0.03)	(0.04)	(2.88)
Log costs intermediary inputs	-0.01	-0.02	0.01	-0.01	0.13
	(0.01)	(0.01)	(0.01)	(0.01)	(1.44)
	95% quantile				
Log cultivated land	0.39	0.44	0.43	0.46	0.23
	(0.07)	(0.04)	(0.00)	(0.04)	(0.00)
Log hired labour cost	0.32	0.29	0.23	0.30	0.23
	(0.05)	(0.03)	(0.04)	(0.05)	(3.51)
Log costs intermediary inputs	-0.08	-0.04	-0.03	-0.07	0.08
	(0.02)	(0.01)	(0.01)	(0.02)	(1.76)

Table 4: Coefficients in function of farm scale

6 Acknowledgments

We sincerely thank the participants of the 122nd EAAE Seminar in Ancona, Sietse Bracke, Mika Kortelainen and Klaas Mulier for their helpful suggestions and insightful comments on an earlier draft of this paper. Marijn Verschelde acknowledges financial support from the Fund for Scientific Research Flanders (FWO Vlaanderen).

References

Aitchison, J., Aitken, C. G. G., 1976. Multivariate binary discrimination by kernel method. Biometrika 63 (3), 413–420.

Assunção, J., Braido, L., 2007. Testing household-specific explanations for the inverse productivity relationship. American Journal of Agricultural Economics 89 (4), 980–990.

Assunção, J., Ghatak, M., 2003. Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity. Economics Letters 80, 189–194.

Barrett, C., 1996. On price risk and the inverse farm size-productivity relationship. Journal of Development Economics 51, 193–215.

Barrett, C., Bellemare, M., Hou, J., 2010. Reconsidering conventional explanations of the inverse productivity-size relationship. World Development 38, 88–97.

Benjamin, D., 1995. Can unobserved land quality explain the inverse productivity relationship? Journal of Development Economics 46, 51–85.

Byiringiro, F., Reardon, T., 1996. Farm productivity in rwanda: effects of farm size, erosion, and soil conservation investments. Agricultural Economics 15, 127–136.

CIA, 2010. The world factbook.
URL `https://www.cia.gov/library/publications/the-world-factbook/`

Cochet, H., 2004. Agrarian dynamics, population growth and resource management : The case of burundi. GeoJournal 60, 111–122.

Collier, P., Dercon, S., 2009. African agriculture in 50 years: Smallholders in a rapidly changing world? In: FAO, UN Economic and Social Development Department.

Dercon, S., Krishnan, P., 1996. Income portfolios in rural ethiopia and tanzania: Choices and constraints. Journal of Development Studies 32, 850–875.

FAO, 2009. Food insecurity in the world, economic crises impacts and lessons learned. FAO (Food and Agriculture Organisation): Rome.

FAOSTAT, 2005. Country profile.
URL `http://faostat.fao.org/`

Feder, G., 1985. The relation between farm size and farm productivity : The role of family labor, supervision and credit constraints. Journal of Development Economics 18, 297–313.

Hall, P., Li, Q., Racine, J. S., 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. Review of Economics and Statistics 89 (4), 784–789.

Hayfield, T., Racine, J. S., 2008. Nonparametric Econometrics: The np Package. Journal of Statistical Software 27 (5), 1–32.

Heltberg, R., 1998. Rural market imperfections and the farm size-productivity relationship: Evidence from pakistan. World Development 26, 1807–1826.

Henderson, D. J., Kumbhakar, S. C., 2006. Public and private capital productivity puzzle: A nonparametric approach. Southern Economic Journal 73 (1), 219–232.

Lamb, R., 2003. Inverse productivity: Land quality, labor markets, and measurement error. Journal of Development Economics 71, 71–95.

Li, Q., Racine, J., 2004. Cross-validated local linear nonparametric regression. Statistica Sinica 14 (2), 485–512.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: theory and practice. Princeton University Press.

Lipton, M., 2010. From policy aims and small-farm characteristics to farm science needs. World Development 10, 1399–1412.

Platteau, J., 1996. The evolutionary theory of land rights as applied to sub-saharan africa: A critical assessment. Development and Change 27, 29–86.

Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data. Journal of Econometrics 119 (1), 99–130.

Rishirumuhirwa, T., Roose, E., 1998. The contribution of banana farming systems to sustainable land use in burundi. Advances in GeoEcology 31, 1197–1204.

Schultz, T., 1964. Transforming Traditional Agriculture. Yale University Press: New Haven.

UNDP, 2010. Human Development Report 2010. The Real Wealth of Nations: Pathway to Human Development. United Nations Development Programme, New York, USA.

Wang, M. C., van Ryzin, J., 1981. A class of smooth estimators for discrete distributions. Biometrika 68 (1), 301–309.

Wiggins, S., Kisten, J., Llambi, L., 2010. The future of small farms. World Development 38, 1341–1348.

WorldBank, 2011. World development indicators.
URL `http:`
`www.worldbank.org`