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

MARIJN VERSCHELDE[1], MARIJKE D'HAESE[2], ELLEN VANDAMME[2], GLENN RAYP[1]

[1]SHERPPA, Department of General Economics, Ghent University, Tweekerkenstraat 2, 9000 Gent, Belgium. e-mail: marijn.verschelde@ugent.be

[2]Department of Agricultural Economics, Ghent University, Coupure links 653, 9000 Gent, Belgium, e-mail: marijke.dhaese@ugent.be

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

Abstract

We use a nonparametric estimation of the production function to investigate the relationship between farm productivity and farming scale in poor smallholder agricultural systems in the north of Burundi. Burundi is one of the poorest countries in the world, with a predominant small scale subsistence farming sector. 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. Four production models were specified each with different control variables. For the relatively small farms, we find clear evidence of an inverse relationship. The relatively large farms show a different pattern. Returns to scale are found to be farm scale dependent. Parametric Cobb-Douglass models tend to over-simplify the debate on returns to scale because of not accounting for the different effects of large farms. Other factors that significantly positively affect 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

1 Introduction

Literature points to the utmost importance of increasing land and labour productivity in the agricultural sector in order to achieve an increase of the African food production (Collier and Dercon, 2009; Piesse and Thirtle, 2010). The (possibly inverse) relationship between farm/plot size and land productivity has been heavily debated over decades now (see e.g. the introduction by Wiggins et al. (2010) of a special section in the November 2010 issue of World Development on the future of small farming). In particular, Collier and Dercon (2009) point to the need for increasing labour productivity on African smallholder farms. Agricultural labour productivity in small-scale farm systems is found to be very low, this is mainly due to the reported overallocation of (family) labour also referred to as hyper-exploitation of family labour, which is basically a problem of very low marginal labour productivity levels (Barrett, 1996).

Important policy issues that emerge are not only how productivity could be increased, but also whether the focus on small — family oriented — farms is the right vehicle for achieving productivity growth. Since Schultz (1964) small farms are considered to be efficient in what they do (Schultz, 1964), and support has been geared towards these smallholder producers. Yet, are they up to the challenge of feeding the growing population? (Wiggins et al., 2010) Are they currently productive enough to meet increasing food demand in the future? The contribution of our study to these questions is that we analyze the factors influencing productivity using a non-parametric estimation of the production function estimated for a unique dataset in the North of Burundi.

Several obvious and less obvious reasons and explanations for the inverse relation between farm productivity and scale (IR) have been tested and proven. A primary 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 less motivated and effective, it takes more productive family labour to supervise hired labour which decreases overall labour productivity at larger farms.
Recent research 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. A second set of missing variables are household specific characteristics such as household size, dependency ratio, and gender of the household head (Assunção and Braido, 2007; Barrett et al., 2010). However none of the studies cited up to now has proven household characteristics to solely explain the IR.
In this paper we try to address a number of important empirical issues. First, we account for mixed output by calculating the market values of all crops produced while allowing for mixed cropping systems. Secondly, by using a nonparametric approach we are able to track heterogeneity in productivity effects of increased access to production factors Thirdly, our rich dataset allows controlling for several of the missing variables mentioned above. The data collection and methodology is explained in the next section.

2 Methodology

2.1 Data

Household data on farm activities was gathered in 2007 in two densely populated provinces of 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.
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.
The farming system in Burundi consists of small peasant landholdings (of generally less than 1 ha per family), very small plots with double cropping, 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 (Cochet, 2004). The most important food crops produced and consumed in the study area were sweet potatoes, beans, cassava, banana and flour of maize (FAO STAT, country profile, 2005).

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 quantities.

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, namely: 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 one of the largest cities 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 a 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	(1,019.01)	2.99**
Size cultivated land (ha)	0.65	(1.1)	0.87	(1.1)	0.75	(1.11)	-2.44**
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**
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 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%

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, a multivariate regressor x and additive error ϵ.

$$y = g(x) + \epsilon \tag{1}$$

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

$$\ln y = \alpha + \sum_{k=1}^{q} \beta_k \ln x_k + \epsilon \tag{2}$$

However, if there are non-linearities or interactions in the true model, the empirical model is misspecified and coefficients are inconsistent (Henderson and Kumbhakar, 2006). A flexible parametric alternative is the Translog specification; quadratic effects and interaction effects are introduced in the empirical model.

$$\ln y = \alpha + \sum_{k=1}^{q} \beta_k \ln x_k + 0.5 \sum_{k=1}^{q} \sum_{l=1}^{q} \beta_{kl} \ln x_k \ln x_l + \epsilon \tag{3}$$

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), (2) the high multicollinearity or low degrees of freedom as result of the inclusion of quadratic effects and interactions.

To avoid imposing *'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 locally averaging those values of the dependent variable which have similar levels of the regressors (one could note it as $\hat{g}(x) = E[Y|\text{X close to x}]$).

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

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i w_i \tag{4}$$

We use Racine and Li (2004) generalized kernel weights to specify the weight function w_i for $x = [x^c, x^o, x^u]$, 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. Kernel functions (l^c, l^o, l^u) are used to be able to give more weight to observations near the observation point. Window widths $(\lambda^c, \lambda^o, \lambda^u)$ impose the window of local averaging. If the window width is large, the curve will be a smooth straight line (as in a linear regression). On the other hand, if the window width is small, non-linearities are allowed for and the curve becomes less smooth.[2] It is intuitively clear and shown in literature that the choice of weighting function is of far less importance than the choice of the window of localization - which we will discuss below.
To construct the weight function for the local averaging, we specify a standard normal kernel function l^c to weight the continuous variables x^c. An Aitchision and Aitken (1976) kernel l^u is specified to weight discrete unordered variables x^u (see (5)). To weight the ordered discrete values x^o, we use a Wang and van Ryzin (1981) kernel function (see (6)).

$$l(X^u_{il}, x^u_l, \lambda^u_l) = \begin{cases} 1 \text{ if } X^u_{il} = x^u_l, \\ \lambda_s \text{ otherwise} \end{cases} \tag{5}$$

$$l(X^o_{im}, x^o_m, \lambda^o_m) = \begin{cases} 1 \text{ if } X^o_{im} = x^o_m, \\ (\lambda^o_m)^{|X^o_{im} - x^o_m|} \text{ otherwise} \end{cases} \tag{6}$$

To allow for a multivariate regression, we use - as is common practice - product kernels. The product kernel of X^c is $W_{\lambda^c}(X^c_i, x^c) = \prod_{k=1}^{q}(\lambda^c_k)^{-1}K((X^c_{ik} - X^c_k)/\lambda^c_k)$. For X^u, the product kernel is defined as $L_{\lambda^u}(X^u_i, x^u) = \prod_{l=1}^{r} l^u(X^u_{il}, x^u_l, \lambda^u_l)$. The product kernel of X^o is $L_{\lambda^o}(X^o_i, x^o) = \prod_{m=1}^{s} l^o(X^o_{im}, x^o_m, \lambda^o_m)$. All together, we can specify a Li-Racine generalized kernel function as $\mathcal{K}_\gamma(X^c_i, X^o_i, X^u_i) = W_{\lambda^c}(X^c_i, x^c)L_{\lambda^u}(X^u_i, x^u)L_{\lambda^o}(X^o_i, x^o)$, with $\gamma = (\lambda^c, \lambda^u, \lambda^o)$.

We estimate $E(Y|X = x)$ by the use of a local-linear estimator. The local-constant (Nadaraya-Watson) estimator 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). The local-linear estimator estimates a local linear relation for each observation point by obtaining a and b in (9). If bandwidths are very large 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.

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

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

[2]Note the trade-off between bias and variance

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

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.[3]

3 Results

3.1 Parametric approach

We start the estimation of the production model with the Cobb-Douglass approach. As there is too few variation in family labour, we define family labour as an ordered discrete variable. As shown in Table 2, the four inputs (land use, family labour, hired labour and intermediary inputs) are found to have a positive and significant effect on output. However, the effect of increasing family labour from 1 or 2 to 3 or 4 persons was only significant at the 10% confidence level. We find no positive effect of increasing family labour above 4 persons. The fixed effect for province was significant with a higher output in the Ngozi province. In addition, the output elasticity for cultivated farm area was smaller than 1. Hence, an IR found between farm size and farm output per unit of land. As the sum of output elasticities of the regressors is significantly lower than 1, the Cobb-Douglass model finds diminishing returns to scale. However, as noted in Section 2, the Cobb-Douglass does not allow for quadratic effects and interactions between the log of the regressors.

To introduce interactions and quadratic effects, we test the proper working of the Translog model for this data set. The results are not reported because we did not find any significant effect any more from the inputs the farmers used. We only find a significant quadratic effect of cost of labour and a significant interaction effect between cost of labour and cost of intermediates. As these results are in sharp contrast to the Cobb-Douglass model, we have doubts on the value of these results. The variation in the model is probably too low to include all the quadratic and interactions effects. Instead of an iterative process of step-wise reduction of the parametric Translog model, we opt for an alternative approach: the nonparametric regression as described in Section 2.

[3] 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.

	Estimate	Std. Error	t-value	p-value
(Intercept)	9.53	0.32	30.20	0.00***
Log cultivated land	0.40	0.03	11.62	0.00***
Family labour: 3-4	0.13	0.08	1.72	0.09°
Family labour: 5 or more	-0.05	0.06	-0.87	0.39
Log hired labour cost	0.17	0.03	5.68	0.00***
Log costs intermediary inputs	0.07	0.03	2.31	0.02*
Province	-0.29	0.06	-5.08	0.00***
Use of hired labour	-1.33	0.29	-4.57	0.00***
Use of intermediary inputs	-0.33	0.31	-1.07	0.29
Adjusted R^2	0.47			
Observations	620			

Significance levels : ∘: 10% $*$: 5% $**$: 1% $***$: 0.1%

Table 2: Cobb Douglass Model

3.2 Nonparametric approach

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. 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 or modus. 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 agrictultural production. Please note that due to limitations of space in this paper we only show figures for the base model and returns to scale.

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 (see Figure 1). 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 size is therefore considered as a combination of both. In Figure 2(a), 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 2(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.

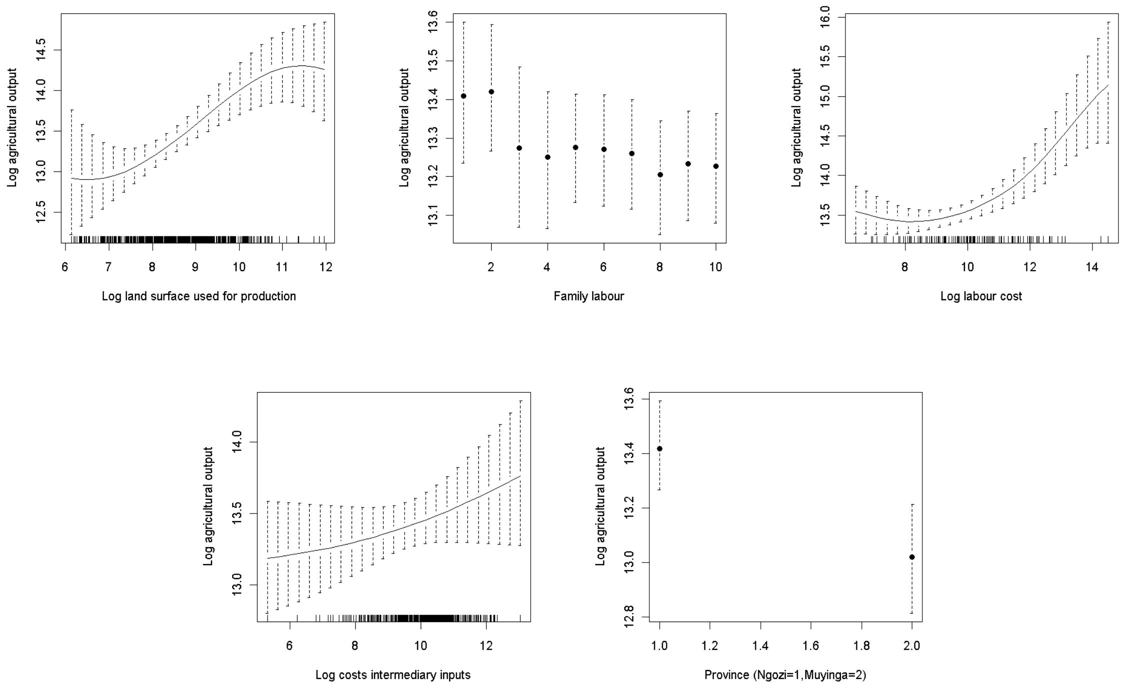
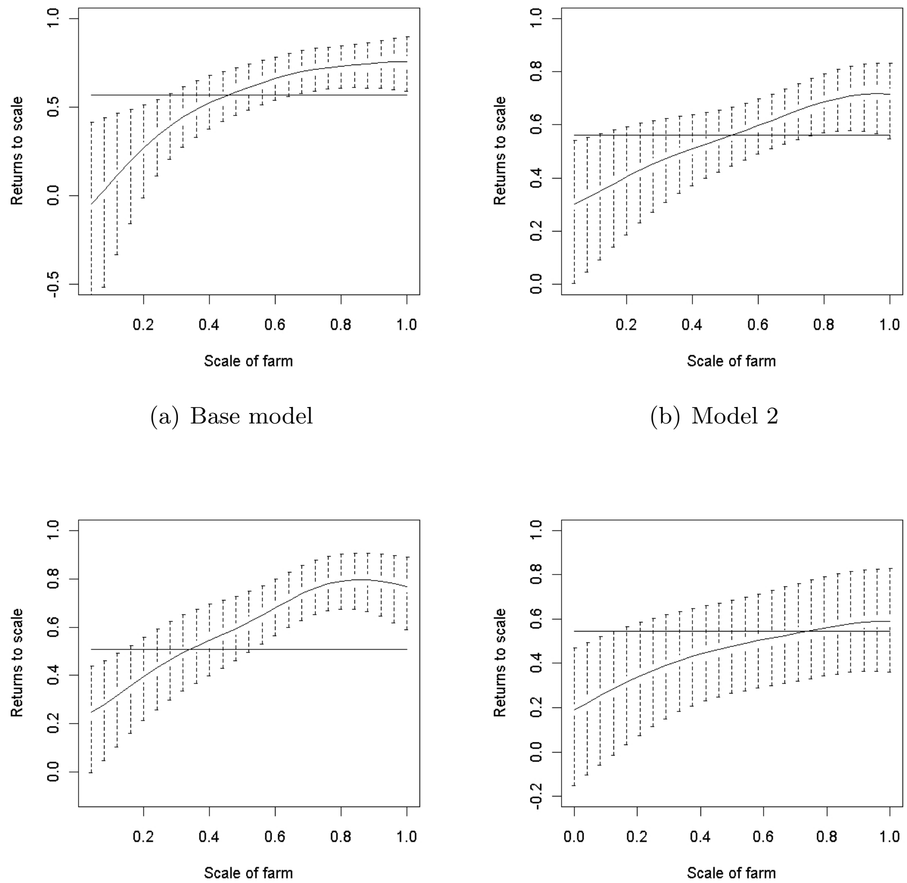


Figure 1: Base Model

In a second model, we control for land use. 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 the farm with banana are found to have a higher agricultural output. The share of coffee planted as the only cash crop positively contributes to production. Again, Figure 2(b) shows that returns to scale are scale dependent.

Model 3 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 . 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 2(c)).
Finally a fourth 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. 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 care. 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 2(d)).



(a) Base model

(b) Model 2

(c) Model 3

(d) Model 4

Figure 2: Returns to scale in function of scale of farm

4 Conclusions

Our results show that parametric models (Cobb-Douglass and Translog specifications) were not satisfactory to estimate the determinants of crop productivity in small-scale farming systems in Burundi. We used a nonparametric kernel estimation of the production function (solved with a local-linear estimator) to allow non-linearities and interaction effects. Four 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 model confirms that farm size itself matters for the relationship between its size and productivity. Our findings confirm both the relatively high productivity of the very small farms, but it also shows the economies of scale that larger farms may exploit. This is a confirmation of the comments made in Dercon and Collier (2009) on the farming scales that are compared in IR literature, namely that the range of farm sizes studied with parametric econometric models is not large enough to show the true relationship between size and productivity. Our results confirm that the effect of size on production is different over the size spectrum. Hence, the potential contribution of agriculture to the potential improvement of the households' livelihoods is different. The implication for policy makers should be to rethink their focus on smallholder agriculture. The options for diversification out of agriculture for these small farms are rather small and they are limited to low paid irregular jobs on other peoples farms or businesses. Yet exploring new better-paid and protected rural non-farm opportunities for the smallest farms is an area for further research. Another topic that we want to explore in the near future are the possible agricultural policy options for optimizing farm production. This includes possibilities for exploiting economies of scale by crop specialization and reducing land fragmentation.

References

Aitchision, 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.

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.

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., Nov. 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. Review Of Economics And Statistics 89 (4), 784–789.

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., Jul. 2006. Public and private capital productivity puzzle: A nonparametric approach. Southern Economic Journal 73 (1), 219–232.

Li, Q., Racine, J., Apr. 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.

Piesse, J., Thirtle, C., 2010. Agricultural r&d, technology and productivity. Philosophical transactions of the Royal Society B 365, 3035–3047.

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., Mar. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. Journal Of Econometrics 119 (1), 99–130.

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

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.