Economía Agraria y Recursos Naturales. ISSN: 1578-0732. e-ISSN: 2174-7350. Vol. 22,1. (2022). pp. 31-52

Farm size and productivity in Nicaragua: Parametric and nonparametric analyses for panel data

Alexandre Nunes Almeida[a], Boris E. Bravo-Ureta[b]

ABSTRACT: This paper examines the inverse relationship hypothesis (IR-H) between farm size and agricultural productivity in Nicaragua using parametric and nonparametric methods designed specifically for panel data. The data employed are from the World Bank's Living Standards Measurement Study for the years 1998, 2001 and 2005. The analysis reveals strong support for the IR-H based on a number of alternative parametric specifications. The results from the nonparametric models lend partial support to the IR-H, and such support is weaker than what is obtained from the parametric models, particularly among medium and large landholders.

Tamaño de la finca y productividad en Nicaragua: análisis paramétrico y no paramétrico con datos de panel

RESUMEN: Este artículo examina la hipótesis de la relación inversa (H-RI) entre el tamaño de la finca y la productividad agrícola en Nicaragua utilizando métodos paramétricos y no paramétricos diseñados específicamente para datos de panel. Los datos provienen del Living Standards Measurement Study del Banco Mundial para los años 1998, 2001 y 2005. El análisis revela un claro apoyo para la H-RI basado en una serie de especificaciones paramétricas. Los resultados de los modelos no paramétricos también apoyan la H-RI; sin embargo, estos últimos proporcionan soporte parcial y más débil particularmente para los propietarios medianos y grandes.

***KEYWORDS* / PALABRAS CLAVE:** Inverse relationship hypothesis, parametric and nonparametric models, farm productivity / *Hipótesis de relación inversa, modelos paramétricos y no paramétricos, productividad agrícola*.

***JEL classification* / Clasificación JEL:** Q12, Q15.

DOI: https://doi.org/10.7201/earn.2022.01.02

[a] Dept. of Economics, Administration and Sociology, Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil. E-mail: alex.almeida@usp.br

[b] Dept. of Agricultural and Resource Economics, University of Connecticut, USA. E-mail: boris.bravoureta@uconn.edu

Cite as: Almeida, A. N. & Bravo-Ureta, B. E. (2022). "Farm size and productivity in Nicaragua: Parametric and nonparametric analyses for panel data". *Economía Agraria y Recursos Naturales*, 22(1), 31-52. doi: https://doi.org/10.7201/earn.2022.01.02

Correspondence author: Alexandre Almeida.

Received on February 2021. Accepted on December 2021.

1. Introduction

The inverse relationship hypothesis (IR-H) states that farm productivity tends to be inversely related to farm size - that is, producers with smaller landholdings tend to be more productive than those with larger ones. The reasons for this inverse relationship have been the subject of considerable controversy for decades[1]. In short, as is often the case in economic analysis, the IR-H debate revolves around two key considerations related to land distribution: equity and efficiency (Berry & Cline, 1979; Binswanger *et al.,* 1995; Helfand & Levine, 2004; Olavarría *et al.,* 2004; Julien *et al.,* 2019). Some authors have argued that failures in land, credit, insurance, and labor markets, along with soil quality and other time-invariant farm characteristics (e.g., farmer skills), are also important determinants in explaining the IR-H (Sen, 1966; Benjamin, 1995; Heltberg, 1998; Assunção & Braido, 2007; Rada *et al.,* 2019).

The presence of possible measurement errors, particularly regarding land (Carletto *et al.,* 2013; Holden & Fischer, 2013; Desiere & Jollife, 2018), along with the view that the analyses should be extended from partial to total factor productivity have provided recent motivation to revisiting the IR-H (Henderson, 2015; Kagin *et al.,* 2016; Julien *et al.,* 2019; Rada & Fuglie, 2019). Moreover, some of the current work on productivity and farm size has shown that is not clear whether small farms are indeed more efficient than large ones (Rada *et al.,* 2019; Fuglie *et al.,* 2020).

On the other hand, Czekaj & Henningsen (2012) and Verschelde *et al*. (2013) argue that findings regarding the IR-H can be also sensitive to the methodology used. Specifically, if larger farmers use different technologies than smaller ones, the use of inputs changes with the scale of production. Thus, econometric models that do not consider the underlying structure of the data and impose inflexible specifications may fail to capture nonlinearities when samples are highly heterogeneous (Verschelde *et al.,* 2013; Henderson & Parmeter, 2015).

The novelty in our analysis is to address the IR-H by adopting recent developments in kernel regression procedures that account for farm heterogeneity due to time-invariant unobservable components, while also checking whether the misspecification of a functional form has a significant bearing on the results. Therefore, our major goal is to analyze nonparametrically the inverse relationship hypothesis in Nicaragua, using an unbalanced household panel dataset for 1998, 2001, and 2005 surveys from the Living Standards Measurement Study (LSMS) of the World Bank.

The main advantage of nonparametric kernel regression methods is that they allow the estimation of a model without defining its functional form *a priori*; that is, such methods "let the data speak for themselves as much as possible" (Barret, 1996; Eubank, 1999). Parametric estimators are considered global (using all data points), while nonparametric kernels use sub-samples of the data close to a point to adjust the estimation while a global estimator is then constructed from such adjustments (Henderson & Parmeter, 2015). In other words, the mean value of the dependent variable is calculated with respect to the specific values of each covariate adjusted

[1] See Lipton (2009); Eastwood *et al*. (2010); Fuglie *et al*. (2020) for a full review.

nonparametrically rather than using the covariate means, as in fully parametric methods (Li & Racine, 2007). The use of nonparametric models with cross-sectional datasets is quite extensive in the literature; however, applications to panel data are limited (Czekaj & Henningsen, 2013; Henderson & Parmeter, 2015). Narrowing this gap in the literature is an important motivation for our study.

Henderson (2015) investigated the IR-H in Nicaragua using the same data source for the same period as we have. After a careful analysis controlling for farm technical and allocative efficiency, he found that labor market imperfections are likely driving the IR-H. He also suggested that the relationship between size and farm productivity is likely nonlinear across different farm size classes. However, Henderson (2015) assumed linearity so here we seek to investigate whether the relationship between farm size and productivity for Nicaraguan farmers remains when nonparametric kernel methods are used.

The rest of this paper is organized as follows: The next section briefly reviews the literature related to the scope of the paper, while section 3 describes the data used. Section 4 presents the parametric and nonparametric approaches adopted, followed by the associated results. Section 6 concludes with a discussion of our major findings and evidence-based policy recommendations.

2. Overview of related literature

As mentioned, the major rationale for using nonparametric methods is to avoid limitations or misspecifications that can stem from incorrect functional forms, which can produce biased estimates (Yatchew, 2009). In general, economic theory does not provide much guidance regarding the choice of functional form for estimation purposes and, in order to determine the shape of a conditional mean relationship, a nonparametric regression is more appropriate than a linear regression (Blundell & Duncan, 1998).

Henderson & Parmeter (2015) argue that, for policy analysis, a fully parametric model is always desirable because it is easier to interpret than a nonparametric alternative. However, policy recommendations can be misguided when inappropriate empirical methods are taken into account (Czekaj & Henningsen, 2012). Hence, nonparametric models can be very helpful in this regard by identifying the true underlying structure of the data (Eubank, 1999).

Parametric methods have been extensively used in investigating the IR-H through the most common specifications of production technologies, such as the Cobb-Douglas and the Translog (Czekaj & Henningsen, 2012; Verschelde *et al.,* 2013; Julien *et al.,* 2019; 2021). In contrast, related studies using semiparametric or nonparametric models, despite their advantage of partially or fully avoiding functional form misspecification, are scarce. Alternatively, the use of the Data Envelopment Analysis (DEA) also avoids having to choose a parametric functional form; however, the natural randomness of agriculture argues for the adoption of stochastic methods (Czekaj & Henningsen, 2012).

Barrett (1996) seems to have been the first to apply a nonparametric kernel regression model to analyze the IR-H using cross sectional data for Madagascar. This author found that production falls as farm size increases and that most of the IR-H is explained by differences between households' marketable surplus and price uncertainty. Assunção & Braido (2007) investigated the IR-H based on longitudinal village-level data for India from 1975 to 1984. They ran nonparametric kernel regressions treating their data as a unique cross-section and found an inverse relationship between output per acre and cropped area, both at the plot and at the aggregate household levels. They suggested that this result can be explained by technological factors. They also suggested that other farm determinants, together with land size, such as land value, soil type, irrigation, village location, year, season, and the crop grown should be also controlled for.

Barrett *et al*. (2010), using data from 2002 for 300 households from Madagascar, found support for the inverse productivity-size relationship based on a nonparametric regression of the logarithm of rice yield on the logarithm of cultivated area. They extended their analysis using fully parametric regressions including other regressors and concluded that market imperfections explain about one-third of the IR-H for Madagascar. However, when the authors incorporated farm specific control variables, (e.g., quality indicators of land), the evidence supporting the IR-H vanished.

Ali & Deininger (2015) reported an inverse relationship using a kernel-weighted nonparametric regression for the logarithm of crop output value against farm or plot size from a 2010/2011 survey of 3,600 households randomly selected from villages of Rwanda. Then, results obtained from fully parametric models including additional regressors revealed that labor market imperfections seem to be the key reason for the inverse relationship between productivity and farm size.

It is worth noting that the nonparametric analyses mentioned above are partial because they relied mainly on cross-sectional estimates concerning a simple (univariate) nonparametric regression between agricultural productivity and land endowments using kernel or spline regressions (Verschelde *et al.,* 2013). Verschelde *et al*. (2013) are the only authors, to our knowledge, who have evaluated the inverse relationship using multivariate kernel regressions, as we do. These authors relied on a cross-sectional household survey from 640 households in 2007 in two Northern provinces of Burundi. They controlled for some time-invariant farm heterogeneity (e.g., soil quality), but not for other sources such as farmer skills or motivation. Their nonparametric results did not reject the inverse relationship between size and farm productivity among small-scale farm holdings.

3. LSMS surveys

The data used in this study are from the Living Standards Measurement Study (LSMS) surveys for the years 1998, 2001, and 2005. The LSMS is a nationwide household survey carried out by the Nicaraguan Statistical Service (INIDE), with technical assistance from the World Bank. The periods of information gathering for the 1998, 2001, and 2005 surveys were April–August, May–August, and July–October, respectively[2]. The LSMS covers a wide range of topics, such as household composition, health, education, income and expenditures, occupation, agricultural production, credit, and savings. The Nicaraguan LSMS is very useful for research purposes because it is designed to follow the same households and individuals over time.

To construct the dataset used in this study, we extracted observations for the three years of the LSMS surveys representing all farms that had non-zero values for: (1) cultivated land (owned, sharecropped, borrowed, or rented); (2) hours of males and females aged 15 or older working on farms; and (3) total farm output (from sales of crops and/or livestock). Moreover, in order to maintain a panel data structure and be able to control for farm/household time-invariant factors, only farm households surveyed for at least two of the three years were included. These conditions yielded an unbalanced panel consisting of 3,278 observations with 986, 1,136, and 1,156 farms for the years 1998, 2001, and 2005, respectively. Data points for households that did not meet the specified criteria, along with a few clear outliers, were excluded.

Table 1 shows descriptive statistics for all variables contained in the econometric analysis. All monetary values were converted from *C*órdoba (C$) to US dollars (US$), using the official nominal exchange rate, and then were converted to real US$ based on the Consumer Price Index (CPI 2005 = 100). The official exchange rate and the CPI were both extracted from the World Development Indicators (World Bank, 2020). The average annual (real) value of output (including sales and consumption) per farm, generated from crop and livestock activities, was low but rose from US $1,431 in 1998 to US $2,737 in 2005. The average reported landholding was about 17 hectares per farm, and did not vary significantly over the survey period. It is important to underscore that land distribution has remained highly unequal in Nicaragua, with a Gini coefficient ranging from 0.76 in 1998 to 0.75 in 2005[3]. Table 1 also shows a classification of the farmers at the 20th, 40th, 60th, and 80th percentiles according to their land size between 1998 and 2005. Our sample shows that 33 % of farmers were marginal with no more than 1.4 hectares, while 21 % were small with the size ranging from 1.4 to 3.5 hectares. Around 22 % and 24 % were medium (between 3.5 and 14 hectares) and large (above 14 hectares) holders, respectively. About 55 % of the large farm group cultivated 40 hectares or less during the three-year period under analysis.

[2] The data can be accessed at no cost at: www.worldbank.org/lsms. We are grateful to the World Bank and the *Instituto Nacional de Información de Desarollo* (INIDE) (www.inide.gob.ni) in Nicaragua for making these data available.

[3] A user-friendly procedure for the STATA software (*fastgini.do*) developed by Zurab Sajaia, from the World Bank, was used to calculate the Gini index for land endowments in our sample. The procedure is available at: http://fmwww.bc.edu/repec/bocode/f/fastgini.ado

TABLE 1

Sample statistics at the farm level: Variable definitions, means, median and standard deviations

Variable	Description	1998		2001		2005	
		Mean	SD	Mean	SD	Mean	SD
TVFO	Total value of farm output in US$	1,431.66	3,364.34	1,986.73	4,112.71	2,737.87	5,067.15
Land	Owned, rented, borrowed and sharecropped in hectares (ha)	18.79	49.38	16.52	43.89	16.51	44.09
Marginal land (20th)	Owned, rented, borrowed and sharecropped below 1.4 ha (%)	0.33	0.47	0.32	0.47	0.33	0.47
Small land (40th)	Owned, rented, borrowed and sharecropped between 1.4 and 3.5 ha (%)	0.21	0.41	0.22	0.41	0.20	0.4
Medium land (60th)	Owned, rented, borrowed and sharecropped between 3.5 and 14 ha (%)	0.22	0.41	0.22	0.42	0.23	0.42
Large land (80th)	Owned, rented, borrowed and sharecropped above 14 ha (%)	0.24	0.43	0.24	0.42	0.24	0.42
IExp	Total expenditure on seed, fertilizers, etc., in dollars/ha	29.51	68.27	32.85	68.59	28.11	63.27
DuIExp	= 1 if farmer spent on seed, fertilizers, etc.	0.81	0.39	0.84	0.37	0.85	0.36
FLabor	On-farm family labor in hours (in male units) per week[a]	148	171	168	208	174	208
HLabor	Hired labor in hours (in male units) per week[a]	8	69	11	45	9	30
DuHLabor	= 1 if used hired labor, and 0 otherwise	0.28	0.45	0.21	0.40	0.35	0.48
RLand	= 1 if used rented, borrowed and sharecropped, and 0 otherwise	0.39	0.49	0.36	0.48	0.40	0.49
Maize	= 1 if produced maize, and 0 otherwise	0.31	0.46	0.79	0.41	0.85	0.36
Livest	= 1 if raised livestock, and 0 otherwise	0.92	0.27	0.89	0.32	0.52	0.50
Title	= 1 if own title of land, and 0 otherwise	0.52	0.50	0.56	0.50	0.55	0.50
Off-farm	= 1 if there are young and adults working off-farm, and 0 otherwise	0.39	0.49	0.50	0.50	0.51	0.50
Age HHHead	Age of household head	46.39	15.40	47.51	15.36	50.04	14.71
Sex HHHead	= 1 if household head is male, and 0 otherwise	0.89	0.31	0.87	0.33	0.86	0.34
Educ HHHead	Education of household head (years of schooling)	2.10	2.50	2.29	2.59	2.30	2.62
HHsize	Household size	6.44	2.97	6.43	2.90	6.24	2.85
Teens	No. of infants and teens (age < 15) working on farm	1.62	3.51	1.61	3.18	1.50	3.67
Training	= 1 if farmers received training, and 0 otherwise	0.14	0.35	0.13	0.33	0.06	0.24
Organiz	= 1 if farmers participated organizations, and 0 otherwise	0.06	0.25	0.07	0.25	0.24	0.42
Credit	= 1 if farmers received rural credit, and 0 otherwise	0.11	0.31	0.09	0.28	0.26	0.44
Irrig	= 1 if irrigated the land, and 0 otherwise	0.01	0.10	0.01	0.11	0.01	0.10
Managua	= 1 if farm located in Managua region, and 0 otherwise	0.02	0.14	0.01	0.11	0.01	0.12
Pacific	= 1 if farm located in Pacífico region, and 0 otherwise	0.24	0.43	0.23	0.42	0.22	0.42
Central	= 1 if farm located in Central region, and 0 otherwise	0.48	0.50	0.52	0.50	0.52	0.50
Atlantic	= 1 if farm located in Atlántico region, and 0 otherwise	0.26	0.44	0.24	0.43	0.24	0.43

Notes: Hours of family and hired labor are in male equivalent units. The weights used are: 5- to 15-year-old male = 0.75 adult male; 5- to 15-year-old female child = 0.65 adult male; and female older than 16 years = 0.75 adult male.

Source: Own elaboration from 1998, 2001 and 2005 LSMS surveys.

On average, both male and female heads of household were in their late forties, their level of education was generally low (two years of schooling), and around 87 % of household heads were male. For all farmers, technical training decreased from 14 % in 1998 to 6 % in 2005 and, during the same period, the number of participants in farmer organizations fluctuated between 6 % and 24 %. The production of maize (above 30 %) and livestock (above 60 %) were the two major agricultural activities, while farms using hired labor varied between 28 % (in 1998), 21 % (in 2001) and 35 % (in 2005). Approximately 54 % of the farmers in the sample worked on land for which they held legal title, and among marginal and small landholders this figure dropped to 34 %.

The participation of male and female members aged 15 or older in the household in any type of off-farm activity (wage labor or self-employment taking place either on other people's farms or in other economic enterprises in rural or urban areas) ranged from 39 % to 51 %. Moreover, for this age group, the total average time devoted to on-farm activities hovered between 148 and 174 hours per week in male worker equivalent units, a measure that weighs the hours worked by females and teens according to FAO (1999) criteria (See at the bottom of Table 1). The LSMS collects detailed labor information for all individuals over five years old, asking whether these persons had worked at one or two jobs for the week prior to the survey, and if they had worked for a 12-month period prior to the survey. Data from these three questions were used in the analysis to compute the total time devoted to on-farm activities[4].

4. Methodology

Our main interest is to investigate the robustness of the association between farm size and the total value of farm output per unit of land while comparing parametric and nonparametric approaches. To illustrate the difference between these two approaches, we first consider the following fully parametric specification including time and farm fixed effects:

$$Ln\left(\frac{TVFO_{it}}{Land_{it}}\right) = \beta_0 + \beta_1 ln(Land_{it}) + \beta_2 ln\left(\frac{IExp_{it}}{Land_{it}}\right) + \beta_3 ln\left(\frac{FLabor_{it}}{Land_{it}}\right) + \beta_4 ln\left(\frac{HLabor_{it}}{Land_{it}}\right) \\ + \beta_5' D_{it} + \beta_6' Z_{it} + \beta_7 year_t + \sum_{i=1}^{n} \alpha_i F_i + \varepsilon_{it} \quad [1]$$

where $TVFO_{it}$ is the total value of farm output (maize, beans, coffee, other crops, and all livestock using constant prices) for farmer *i* at time *t* in US$; $Land_{it}$ is cultivated land in hectares; $IExp_{it}$ is expenditure on purchased inputs (fertilizers, seeds, seedlings, pesticides, etc.) in US$; $FLabor_{it}$ is on-farm family hours of male

[4] Labor activities performed on other people's farms are not included. For methodological details regarding the surveys, see World Bank (1998); World Bank (2002); World Bank (2006).

equivalent units; $HLabor_{it}$ is the use of hired labor in hours of male equivalent units; and D_{it} is a vector of four dummy variables equal to 1 if farmers use rented land, cultivate maize, produce livestock, and have a land title. The term Z_{it} represents a vector of variables capturing farm characteristics including: children and teenagers working on the farm, training, membership in labor associations, rural credit, farm irrigation, farm location (regions), household members working off the farm, and household head characteristics (gender, age, and education). Details are presented in Table 1. The model also includes a year dummy variable to account for technological progress, and farm fixed effects (F_i) to capture unobserved farm/farmer heterogeneity. The random error is assumed to be *iid* with a zero mean and finite variance.

An important limitation in estimating Equation 1 using a nonparametric regression approach is the "curse of dimensionality", which can lead to inconsistent estimates (Li & Racine, 2007; Henderson & Parmeter, 2015). In other words, as dimensionality increases it becomes more difficult to detect the real structure of the data without assuming *a priori* assumptions (e.g., linearity) as parametric models do (Henderson & Parmeter, 2015). Thus, to avoid the curse of dimensionality while accounting for observed and unobserved variables, we estimate a more parsimonious specification by omitting the vector Z_{it}, deemed less critical to our analysis. Therefore, the nonparametric version of Equation 1 is as follows:

$$Ln\left(\frac{TVFO_{it}}{Land_{it}}\right) = \varphi[\ln(Land_{it}), ln\left(\frac{IExp_{it}}{Land_{it}}\right), ln\left(\frac{FLabor_{it}}{Land_{it}}\right), Ln\left(\frac{HLabor_{it}}{Land_{it}}\right), D_{it}, year_t, \sum_{i=1}^{n} F_i] + \varepsilon_{it} \qquad [2]$$

where Equation 2 follows the same general specification as (1) except that now is an unknown function of all continuous and dummy variables including the farm fixed effects (F_i). To estimate Equation 2 while mitigating the curse of dimensionality (Czekaj & Henningsen, 2013), we apply the nonparametric regression method developed by Racine & Li (2004); Li & Racine (2004), which can handle both continuous and categorical (e.g., years and farm fixed effects) explanatory variables. This method is analogous to the least-squares with dummy-variable (LSDV) approach in parametric models (Henderson & Parmeter, 2015).

Finally, for any kernel nonparametric estimation, the choice of the bandwidth parameter, which regulates the trade-off between variance and bias in the estimates, is more important than the choice of the kernel functions such as Epanechinikov, Uniform or Gaussian (Li & Racine, 2007). The bandwidths are estimated via cross-validation based on Akaike information criteria (AICCV) using the Gaussian kernel for the continuous variables (Henderson & Parmeter, 2015), and kernel functions specifically designed to deal with categorical variables (Aitchison & Aitken, 1976; Wang & van Ryzin, 1981). The theoretical foundation and application details for the nonparametric estimators discussed above are omitted here but are available in Li & Racine (2007) and Henderson & Parmeter (2015). The latter authors also provide R software codes that allow users to replicate all the examples presented in their 2015 book, and the programing codes can be easily adapted to other cases. The codes can be accessed at http://www.the-smooth-operators.com/code.

5. Results and discussion

To investigate the IR-H, we start with results for a total of six fully parametric models presented in Table 2. The dependent variable is the total value of farm output (TVFO) in US$ divided by total cultivated land in hectares. Specifically, we have three pooled models (columns 1, 2 and 3), and three that incorporate fixed effects (columns 4, 5 and 6). Another major difference across the regressions is the number of observable controls embedded in the D_{it} and Z_{it} vectors (see Equation 1). A further consideration is that 16.4 % and 72.2 % of all observations in the 1998–2005 panel for the variables *IExp* and *Hlabor*, respectively, are reported as zeros. So, to facilitate the estimation of the parametric and nonparametric production functions in logarithmic terms, we used the Battese (1997) procedure. It is important to note that Hausman tests favored fixed over random effects at the 1 % level for all three specifications that include individual farm effects.

Table 2 reveals that the coefficients for land, of primary interest here, are highly significant and uniformly negative, lending support to the IR hypothesis. The first specification (column 1) shows a land coefficient of -0.25 and this value becomes -0.15 when the D_{it} vector (land rental, corn cultivation, livestock production and land tenure) is excluded (column 2). However, when the observable farm characteristics embedded in the Z_{it} vector, and the control dummies for years and regions are omitted, the four variables included in the D_{it} vector have a substantial influence in explaining farm productivity in our naïve cross-section analysis, revealing a more robust inverse relationship with a land elasticity of -0.20 (column 3).

A more comprehensive examination of the IR-H is possible when important and often unobservable or omitted variables, such as institutional features, managerial ability, imperfect factor markets, soil productivity and environmental resources, can be controlled for (Barrett *et al.,* 2010; Julien *et al.,* 2019). If panel data are available, as they are for our study, the researcher can conveniently control for the effects of such variables (Henderson, 2015). Our panel data evidence of the IR-H, exhibited in columns 4-6 of Table 2, is more robust than the results obtained when the data are treated as cross-sectional (columns 1, 2, 3 in Table 2). The coefficients for land of the three panel data models range between -0.53, -0.48 and -0.47, and all are statistically significant at the 1 % level. Consequently, the parametric results clearly indicate that the time-invariant and unobserved farm/farmer heterogeneity play an important role in explaining the IR-H for Nicaragua.

TABLE 2

Parametric regressions of farm productivity

	Dependent Variable: Ln TFVO per ha					
	Pooled			Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Land	-0.25*** (0.03)	-0.15*** (0.03)	-0.20*** (0.03)	-0.53*** (0.04)	-0.47*** (0.04)	-0.48*** (0.04)
D_1Ln Iexp per ha	0.28*** (0.02)	0.30*** (0.02)	0.31*** (0.02)	0.16*** (0.02)	0.16*** (0.02)	0.17*** (0.02)
D_1	-0.47*** (0.08)	-0.38*** (0.08)	-0.45*** (0.08)	-0.24*** (0.08)	-0.20** (0.08)	-0.24*** (0.08)
Ln Flabor per ha	0.13*** (0.03)	0.13*** (0.03)	0.11*** (0.02)	0.12*** (0.03)	0.13*** (0.03)	0.14*** (0.03)
D_2Ln Hlabor per ha	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.08*** (0.03)	0.09*** (0.03)	0.08*** (0.02)
D_2	0.38*** (0.05)	0.38*** (0.05)	0.45*** (0.05)	0.27*** (0.06)	0.28*** (0.06)	0.28*** (0.06)
Rland (dummy)	-0.08 (0.05)		-0.10* (0.06)	-0.05 (0.06)		-0.04 (0.06)
Maize (dummy)	0.29*** (0.05)		0.69*** (0.05)	0.48*** (0.06)		0.47*** (0.06)
Livest (dummy)	0.55*** (0.05)		0.12** (0.05)	0.29*** (0.06)		0.29*** (0.06)
Title (dummy)	0.24*** (0.05)		0.24*** (0.06)	0.26*** (0.07)		0.27*** (0.07)
Off-Farm	-0.30*** (0.04)	-0.32*** (0.04)		-0.19*** (0.05)	-0.19*** (0.05)	
Age HHHead	-0.00 (0.00)	0.00 (0.00)		0.01 (0.00)	0.01* (0.00)	
Sex HHHead	0.22*** (0.06)	0.22*** (0.06)		0.13 (0.13)	0.13 (0.14)	
Educ HHHead	0.03*** (0.01)	0.02*** (0.01)		0.06*** (0.02)	0.05*** (0.02)	
HHSize	-0.01 (0.01)	-0.00 (0.01)		0.02 (0.01)	0.01 (0.01)	
Teens	0.00 (0.01)	0.00 (0.01)		-0.00 (0.01)	-0.00 (0.01)	
Trainning (dummy)	0.06 (0.07)	0.08 (0.07)		0.20** (0.08)	0.20** (0.08)	
Organiz (dummy)	0.01 (0.06)	0.05 (0.06)		0.02 (0.07)	0.04 (0.07)	
Credit (dummy)	0.03 (0.05)	-0.01 (0.05)		0.14** (0.06)	0.13* (0.07)	
Irrig (dummy)	-0.24 (0.15)	-0.29* (0.16)		0.03 (0.23)	-0.00 (0.22)	

TABLE 2 (cont.)

Parametric regressions of farm productivity

	Dependent Variable: Ln TFVO per ha					
	Pooled			Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Regions (dummies)	Yes	Yes	No	Yes	Yes	No
Years (dummies)	Yes	Yes	No	Yes	Yes	No
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,278	3,278	3,278	3,278	3,278	3,278
R^2	0.45	0.42	0.37	0.50	0.47	0.49
F	90.49	97.90	164.98	78.03	83.51	138.45
Hausman χ^2				217.76***	168.86***	222.10***

Notes: Robust standard errors in parentheses.

***, **, * *Significant at the 1 %, 5 % and 10 %, respectively.*

Source: Own elaboration.

In contrast to the previous analysis, the nonparametric kernel methods relax parametric assumptions (e.g., linearity) in the production function (Henderson & Simar, 2005), and allow the data to have a greater role in determining the functional form underlying the models (Czekaj & Henningsen, 2013). Consequently, nonparametric approaches accommodate non-linear relationships between farm productivity and production factors which can have implications for the ensuing analysis (Verschelde *et al.,* 2013)[5].

Our fully nonparametric kernel results, corresponding to Equation 2, are displayed in Table 3 and consist of the following two models: (1) A fully nonparametric model designed for pooled datasets called Local Linear Least Squares (LLLS); and (2) A fixed effects nonparametric estimator developed by Li & Racine (2004) in order to account for the panel data structure. To make the results of these two kernel nonparametric models comparable with their parametric specifications (columns 3 and 6, Table 2), we summarize elasticities for each regressor at their overall means and at the mean of the 20th (Q1), 40th (Q2), 60th (Q3), and 80th (Q4) percentiles along with the corresponding robust standard errors (Henderson & Parmeter, 2015). Two additional specification tests between the nonparametric LLLS pooled model and the comparable parametric OLS pooled model (column 3, Table 2) show inconclusive results. As can be seen at the bottom of Table 3, p-value statistics indicate that the fully parametric model is rejected, favoring the nonparametric model based on the

[5] Appendix Figure 1A shows a much weaker support for the IR-H coming from the nonparametric model compared to the parametric regressions, suggesting the importance of a careful scrutiny of this hypothesis.

Ullah test (1985). However, the Horowitz & Hardle (1994) test does not reject the nonparametric model against the parametric counterpart[6].

Overall, both nonparametric models show higher R^2s (0.49 and 0.65) than the parametric counterparts (0.37 and 0.49). The LLLS pooled model generates elasticities for land that are similar to those obtained from the parametric models, exhibiting support for the IR-H. However, this finding holds for the marginal (20th), small (40th) and medium (60th) farm sizes, but not for the larger ones (80th). The statistically significant estimates for land range from -0.22 to -0.11 across percentiles, while the overall mean elasticity of land is also statistically significant with a value of -0.15.

The nonparametric estimates of the Li & Racine fixed effects model corroborate the relevance of the unobserved and time-invariant characteristics, as the panel data parametric models do, although the results are statistically robust in explaining the IR-H only for the marginal and small farms and at the overall mean. Specifically, the 20th percentile of the land estimate, i.e., farmers who cultivate up to 1.4 hectares, shows an elasticity of -0.29, while farmers at the 40th percentile show an estimate of -0.20. The mean land elasticity is -0.16, substantially lower than the -0.48 value obtained from the parametric counterpart (column 6, Table 2).

The IR-H evidence from the nonparametric findings across percentiles, observed in both the cross-section and panel data analyses, confirms that farm size matters in achieving higher productivity gains. Specifically, the kernel results reveal a weaker or no inverse relationship for medium and large farmers, but the relationship is supported for marginal and small farms. This is not a completely unexpected result for Nicaragua given that approximately 45 % of the total population lives in rural areas, and 70 % of the rural inhabitants are subsistence farmers below the poverty line (IFAD, 2017). Thus, under the presence of labor market failures and if landowners are in regions with poor soils, as off-farm market wages go up more time is allocated to off-farm work and less time and resources are devoted to on-farm activities, which can lower agricultural output (Almeida & Bravo-Ureta, 2019). This latter finding is consistent with Henderson (2015), who argues that frictions in labor market participation likely explain the inverse relationship particularly among small landholders in Nicaragua.

The pooled and fixed effects parametric models (Table 2) show that the coefficients of total expenditures on seed, fertilizers etc. (Iexp) and family labor (Flabor) are statistically significant with positive signs and are largely unaffected by the inclusion of the control variables inserted in the D_{it} and Z_{it} vectors. Similarly, the nonparametric estimates (Table 3) for Iexp also display positive and statistically significant values across percentiles; however, the same consistent results are not observed for Flabor. In both parametric and nonparametric estimations, hired labor (Hlabor) turned out to be positive and statistically significant, contrasting with some of the literature, which insinuates that the difficulties associated in supervising

[6] According to Henderson & Parmeter (2015), specification tests for nonparametric models have not been studied enough and are a promising avenue for future work particularly in the context of panel data models.

wage workers likely lead to lower labor and farm productivity (Feder, 1985; Ali & Deininger, 2015).

The parametric estimates reveal that renting land contributes negatively to agricultural productivity; however, this coefficient is statistically significant only in one of the pooled models (column 3, Table 2) but not in the fixed effects models (columns 4 and 6, Table 2). The pooled nonparametric model exhibits statistically significant estimates for rented land at the 20^{th} percentile (marginal farmer's) with a value of -0.35 and at the 40^{th} percentile (small farmer's) with a value of -0.23 in the Li & Racine model.

Corn and livestock production contribute positively to farm productivity according to both the parametric and nonparametric findings. In addition, the coefficients for land titling in the pooled and fixed effects parametric specifications (columns 3 and 6, Table 2) are statistically significant with values of 0.24 and 0.27, respectively. The LLLS pooled model displays robust estimates for land titling ranging from 0.20 at the 60^{th} percentile to 0.32 at the 80^{th} percentile, while for the Li & Racine approach a significant estimate (0.37) is observed only at the 80^{th} percentile.

The remaining parameters in the Z_{it} vector for the parametric pooled and fixed effects regressions (Table 2) indicate a positive statistical relationship between farm productivity and sex (in favor of males) and the education of household heads, and a negative association with household members working off farm[7] and with irrigation. Training and rural credit are also positively associated with farm productivity; however, statistically significant parameters for these two variables are observed only for the fixed effects estimates (columns 4 and 5).

As discussed in the methodology section, the performance of any kernel nonparametric regression relies critically on the selection of bandwidths. Therefore, conducting robustness checks using alternative bandwidths is strongly recommended (Henderson & Parmeter, 2015). A variety of methods, such as rule-of-thumb, plug-in and cross-validation, are available for obtaining optimal bandwidths. However, no final verdict has yet been reached on which of these methods is preferable (Henderson & Parmeter, 2015). Thus, to check the robustness of our findings, Tables A1 and A2 in the Appendix show nonparametric results for the LLLS and Li & Racine methods, using different bandwidths estimated via rule-of-thumb (ROT) and least squares cross-validation (LSCV)[8]. In summary, the results still support the IR-H for marginal (ROT and LSCV in the panel data models) and small (LSCV in the pooled model) farmers but not for medium and large farmers. Similar outcomes were observed previously for the Akaike information criteria cross-validation (AICCV) option.

[7] In contrast to our findings, evidence from other developing countries suggests that off-farm income can contribute positively to the purchase of inputs and to on-farm investments, leading to improved yields and more profitable farms (Oseni & Winters, 2009; Zeeshan & Giri, 2019).

[8] See Appendix Table A3.

TABLE 3

Elasticities of farm productivity for selected nonparametric estimators with bandwidths estimated via cross-validation based on Akaike information criteria (AICCV)

	Dependent Variable: Ln TFVO per ha									
	LLLS-Pooled					Li & Racine				
	Mean	Q1	Q2	Q3	Q4	Mean	Q1	Q2	Q3	Q4
Ln Land	-0.15***	-0.22***	-0.15*	-0.11***	-0.06	-0.16***	-0.29**	-0.20*	-0.13	-0.03
	(0.05)	(0.07)	(0.09)	(0.05)	(0.06)	(0.06)	(0.13)	(0.12)	(0.08)	(0.06)
D_1Ln Iexp/ha	0.36***	0.30***	0.34***	0.37***	0.40***	0.35***	0.26***	0.32***	0.37***	0.43***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)	(0.08)
D_1	-0.50***	-0.96***	-0.77***	-0.53***	0.05	-0.55***	-1.12***	-0.75***	-0.40***	0.07
	(0.13)	(0.14)	(0.13)	(0.13)	(0.15)	(0.20)	(0.21)	(0.22)	(0.12)	(0.15)
Ln Flabor/ha	0.10**	0.04	0.07*	0.10**	0.16	0.09***	0.01	0.05	0.09	0.16***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.10)	(0.03)	(0.04)	(0.04)	(0.06)	(0.06)
D_2Ln Hlabor/ha	0.19***	0.12	0.17*	0.21***	0.24*	0.23***	0.10	0.17***	0.23***	0.32***
	(0.06)	(0.08)	(0.09)	(0.06)	(0.13)	(0.06)	(0.09)	(0.05)	(0.06)	(0.10)
D_2	0.28***	0.15*	0.24***	0.30***	0.39***	0.23*	0.05	0.13	0.21*	0.42**
	(0.09)	(0.08)	(0.09)	(0.11)	(0.09)	(0.12)	(0.08)	(0.09)	(0.11)	(0.21)
Rland (dummy)	-0.19	-0.35**	-0.21	-0.14	-0.05	-0.17	-0.37	-0.23***	-0.10	0.06
	(0.12)	(0.17)	(0.20)	(0.10)	(0.11)	(0.21)	(0.12)	(0.16)	(0.08)	(0.25)
Maize (dummy)	0.72***	0.51***	0.71***	0.81***	0.94***	0.41**	0.12	0.29**	0.55***	0.71***
	(0.13)	(0.08)	(0.11)	(0.11)	(0.18)	(0.20)	(0.25)	(0.12)	(0.15)	(0.17)
Livest (dummy)	0.18**	-0.06	0.05	0.18*	0.44***	0.51***	0.29**	0.41***	0.52***	0.75***
	(0.07)	(0.08)	(0.07)	(0.10)	(0.12)	(0.12)	(0.13)	(0.09)	(0.06)	(0.12)
Title (dummy)	0.18*	0.06	0.12	0.20**	0.32*	0.16	-0.06	0.05	0.16	0.37***
	(0.10)	(0.13)	(0.11)	(0.09)	(0.17)	(0.13)	(0.07)	(0.09)	(0.11)	(0.16)
Years (dummies)			No							
Farm Fixed Effects			No							
Observations			3,278					3,278		
R^2			0.49					0.65		
Specification Tests			Parametric OLS Pooled vs. Nonparametric LLLS Pooled							
Ullah (1985) p-value			0.00							
Horowitz & Hardle (1994) p-value			0.50							

This Table reports elasticities at the mean, 20th (Q1), 40th (Q2), 60th (Q3) and 80th (Q4) percentiles of each regressor along with bootstrapped standard errors (400 replications) in parentheses.

***, **, * *Significant at the 1 %, 5 % and 10 %, respectively.*

Source: Own elaboration.

6. Concluding remarks

Nicaragua has experienced gains in economic growth and poverty reduction over the past decade but remains one of the poorest countries in Central America (IFAD, 2017). Moreover, historically, subsistence farmers have been neglected socially either for cultural or ethnic reasons, or for living in places where their assets are very limited, especially regarding land access, social services and insufficient infrastructure (Piccioni, 2015). Even though our data are a few years old, the analysis contains valuable insights regarding the Nicaraguan rural sector while using methods that are novel in applied production economics research.

The objective of this study was to test the IR-H between productivity and farm size in Nicaragua, using panel data along with parametric and nonparametric methods. We also explored whether functional form misspecification plays a significant role in the IR-H. To the best of our knowledge, panel data nonparametric regressions have been largely ignored in applied work and thus warrant further attention in future studies. Our parametric analysis revealed consistent results in support of the IR-H; however, the evidence stemming from the nonparametric findings is weaker, suggesting that closer attention is needed when deciding which methodological approach to use.

Parametric and nonparametric panel data estimations also provided more robust support for the IR-H compared to cross-sectional analysis. An interesting feature of our nonparametric findings is that productivity decreased as farm size increased for marginal and small operations, but not so for medium and large landholdings. Thus, in regions where farmland is insufficient or less productive, and labor market imperfections are significant, off-farm work is likely to be more attractive. Therefore, producers are likely to devote more efforts off-farm with potentially adverse effects on agricultural output (Almeida & Bravo-Ureta, 2019).

A major implication of this study is that Nicaraguan farmers with limited resources would find it difficult to adjust their inputs and production practices without agricultural policies seeking to correct market imperfections, particularly in terms of labor, and to enhance managerial capacities and soil fertility aiming at increasing farm productivity (Henderson, 2015). Another key strategy for increasing productivity is the implementation of agricultural policies that facilitate access to improved inputs and credit (Sibande *et al.,* 2017).

Our parametric and nonparametric analyses show that, between 1998 and 2005, having legal land ownership contributed positively to farm productivity. In fact, we observed that 66 % of our sample comprised marginal and small landholders who had no legal title. As a result, their access to credit was limited; thus, they were constrained in terms of the capacity to invest in their farms with the purpose of boosting productivity. Consequently, land tenure policies can also contribute to improve access to credit and greater farm income (Boucher *et al.,* 2005). Renting land as well as maize cultivation and cattle raising also played an important role in explaining the inverse relationship between farm size and productivity.

In sum, transferring land from large to small farmers in Nicaragua might be important for promoting agricultural growth (Henderson, 2015). However, such transfers, to be effective, should be accompanied by policy interventions designed not only to enhance productivity and social and human capital, but also to correct factor market imperfections (Henderson, 2015; Piccioni, 2015), along with promoting the adoption of new technologies that are friendly to the environment (De los Santos-Montero & Bravo-Ureta, 2017). Finally, as our dataset is based on self-reported survey data rather than on GPS-based measurements, there is a potential bias if land area, land quality or output are measured with error (Desiere & Jolliffe, 2018; Julien *et al.,* 2019). Of course, this caveat affects most of the related literature, so the ongoing work to reduce measurement error deserves continued support.

References

Almeida, A.N. & Bravo-Ureta, B.E. (2019). “Agricultural productivity, shadow wages and off-farm labor decisions in Nicaragua”. *Economic Systems,* 43, 99-110. https://dx.doi.org/10.1016/j.ecosys.2018.09.002

Aitchison, J. & Aitken, C.G.G. (1976). “Multivariate binary discrimination by the kernel method”. *Biometrika,* 63(3), 413-420. https://dx.doi.org/10.2307/2335719

Ali, D.A. & Deininger, K. (2015). “Is there a farm size-productivity relationship in African agriculture? Evidence from Rwanda”. *Land Economics,* 91(2), 317-343. Retrieved from: https://www.jstor.org/stable/24773531

Assunção, J.J. & Braido, L.H.B. (2007). “Testing household-specific explanations for the inverse productivity relationship”. *American Journal of Agricultural Economics,* 89(4), 980-990. https://dx.doi.org/10.1111/j.1467-8276.2007.01032.x

Barrett, C.R. (1996). “On price risk and the inverse farm size-productivity relationship”. *Journal of Development Economics,* 51(2), 193-215. https://dx.doi.org/10.1016/S0304-3878(96)00412-9

Barrett, C.R., Bellemare, M.R. & Hou, J.Y. (2010). “Reconsidering conventional explanations of the inverse productivity-size relationship”. *World Development,* 38(1), 88-97. https://dx.doi.org/10.1016/j.worlddev.2009.06.00

Battese, G.E. (1997). “A note on the estimation of Cobb-Douglas production function when some explanatory variables when some explanatory variables have zero values”. *Journal of Agricultural Economics,* 48(2), 250-252. https://dx.doi.org/10.1111/j.1477-9552.1997.tb01149.x

Benjamin, D. (1995). “Can unobserved land quality explain the inverse productivity relationship?” *Journal of Development Economics,* 46(1), 51-84. https://dx.doi.org/10.1016/0304-3878(94)00048-H

Berry, R.A. & Cline, W.R. (1979). *Agrarian Structure and Productivity in Developing Countries*. Baltimore: The John Hopkins University Press.

Binswanger, H.P., Deininger, K. & Feder, G. (1995). "Power, distortions, revolt and reforming agricultural land relations". In Behrman, J. & T.N. Srinivasan (Eds.): *Handbook of Development Economics, Volume 3* (pp. 2659-2772). North Holland, Amsterdam: ScienceDirect.

Blundell, R & Duncan, A. (1998). "Kernel regression in empirical economics". *Journal of Human Resources,* 13(1), 435-461. https://dx.doi.org/10.2307/146315

Boucher, S.R., Barham, B.L. & Carter, M.R. (2005). "The impact of market-friendly reforms on credit and land markets in Honduras and Nicaragua". *World Development,* 33(1), 107-128. https://dx.doi.org/10.1016/j.worlddev.2004.09.004

Carletto, C., Savastano, S. & Zezza, A. (2013). "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship". *Journal of Development Economics,* 103, 254-261. https://dx.doi.org/10.1016/j.jdeveco.2013.03.004oi.org/10.1111/1477-9552.12086

Henderson, D.J. & Parmeter, C.F. (2015). *Applied Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Henderson, D. & Simar, L. (2005). "A fully nonparametric stochastic frontier model for panel data." *STAT Discussion Paper 052.* Retrieved from: https://dial.uclouvain.be/pr/boreal/object/boreal:122235

Holden, S. & Fisher, M. (2013). "Can area measurement error explain the inverse farm size productivity relationship?" *Centre for Land Tenure Working Paper* 12/13. Retrieved from: https://www.nmbu.no/download/file/fid/40190

Horowitz, J.L. & Härdle, W. (1994). "Testing a parametric model against a semiparametric alternative", *Econometric Theory,* 10(5), 821-48. Retrieved from: https://www.jstor.org/stable/3532855

International Fund for Agricultural Development (IFAD). (2017). "Investing in rural people in Nicaragua". Technical Report. Rome: IFAD. Retrieved from: https://www.ifad.org/en/web/knowledge/-/publication/investing-in-rural-people-in-nicaragua

Julien, J.C., Bravo-Ureta, B.E. & Rada, N.E. (2019). "Assessing farm performance by size in Malawi, Tanzania, and Uganda". *Food Policy,* 84, 153-164. https://dx.doi.org/10.1016/j.foodpol.2018.03.016

Julien, J.C., Bravo-Ureta, B.E. & Rada, N.E. (2021). "Productive efficiency and farm size in East Africa". *Agrekon,* 60, 209-226. https://dx.doi.org/10.1080/03031853.2021.1960176

Kagin, J., Taylor, E.J. & Yúnes-Naude, A. (2016). "Inverse productivity or inverse efficiency? Evidence from Mexico". *Journal of Development Studies,* 52(3), 396-411. https://dx.doi.org/10.1080/00220388.2015.1041515

Li, Q. & Racine, J.S. (2004). "Cross-validated local linear nonparametric regression". *Statistica Sinica,* 14, 485-512. Retrieved from: https://www.jstor.org/stable/24307205

Li, Q. & Racine, J.S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

Lipton, M. (2009). *Land Reform in Developing Countries: Property Rights and Property Wrongs*. New York: Routledge.

Olavarría, J.A., Bravo-Ureta, B.E. & Cocchi, H. (2004). "Productividad total de los factores en la agricultura chilena: 1961-1996". *Economía Agraria y Recursos Naturales,* 4(8), 121-132. https://dx.doi.org/10.22004/ag.econ.28774

Oseni, G. & Winters, P. (2009). "Rural nonfarm activities and agricultural crop production in Nigeria". *Agricultural Economics,* 40(2), 189-201. https://dx.doi.org/10.1111/j.1574-0862.2009.00369.x

Piccioni, N.B. (2015). *Agriculture in Nicaragua: Performance, Challenges, and Options*. Washington D.C.: The World Bank. Retrieved from: https://openknowledge.worldbank.org/handle/10986/25978

Racine, J.S. & Li, Q. (2004). "Nonparametric estimation of regression functions with both categorical and continuous data". *Journal of Econometrics,* 119(1), 99-130. https://dx.doi.org/10.1016/S0304-4076(03)00157-X

Rada, N. & Fuglie, K.O. (2019). "New perspectives on farm size and productivity". *Food Policy,* 84, 147-152. https://dx.doi.org/10.1016/j.foodpol.2018.03.015

Rada, N., Helfand, S. & Magalhães, M. (2019). "Agricultural productivity growth in Brazil, large and small farmers excel". *Food Policy,* 84, 176-185. https://dx.doi.org/10.1016/j.foodpol.2018.03.014

Sen, A. (1966). "Peasants and dualism with or without surplus labor". *Journal of Political Economy,* 74(5), 425-450. Retrieved from: https://www.jstor.org/stable/1829592

Sibande, L., Bailey, A. & Davidova, S. (2017). "The impact of farm input subsidies on maize marketing in Malawi". *Food Policy,* 69, 190-206. https://dx.doi.org/10.1016/j.foodpol.2017.04.001

Ullah, A. (1985). "Specification analysis and economics models". *Journal of Quantitative Economics,* 2, 187-209.

Verschelde, M., D'Haise, M., Rayp, G. & Vandamme, E. (2013). "Challenging small-scale farming: A nonparametric analysis of the (inverse) relationship between farm productivity and farm size in Burundi". *Journal of Agricultural Economics,* 64(2), 319-342. https://dx.doi.org/10.1111/j.1477-9552.2012.00373.x

Wang, M.C. & van Ryzin, J. (1981). "A class of smooth estimators for discrete distributions". *Biometrika,* 68(1), 301-309. https://dx.doi.org/10.2307/2335831

World Bank. (1998). *Nicaragua: Living standards measurement study 1998*. Basic Information Document. Washington D.C.: The World Bank.

World Bank. (2002). *Nicaragua: Living standards measurement study 2001*. Ficha Técnica. Washington D.C.: The World Bank.

World Bank. (2006). *Nicaragua: Encuesta Nacional De Hogares Sobre Medición De Nivel De Vida 2006*. Informe de Metodología y Operaciones. Washington D.C.: The World Bank.

World Bank. (2020). *World Bank Development Indicators*. Washington D.C.: World Bank. Retrieved from: http://data.worldbank.org/data-catalog/world-development-indicators

Yatchew, A. (2009). *Semiparametric regression for the applied econometrician*. Cambridge: Cambridge University Press.

Zeeshan, G.M. & Giri, A.K. (2019). "The effects of non-farm enterprises on farm households' income and consumption expenditure in Rural India". *Economía Agraria y Recursos Naturales*, 19(1), 195-222. https://dx.doi.org/10.7201/earn.2019.01.10

Appendix

TABLE A1

Elasticities of farm productivity for selected nonparametric estimators with bandwidths estimated via rule-of-thumb (ROT)

	Dependent Variable: Ln TFVO per ha									
	LLLS-Pooled					Li & Racine				
	Mean	Q1	Q2	Q3	Q4	Mean	Q1	Q2	Q3	Q4
Ln Land	-0.12	-0.42	-0.22	-0.06	0.12	-0.15**	-0.40***	-0.22	-0.09	0.11
	(0.15)	(0.37)	(0.18)	(0.13)	(0.11)	(0.06)	(0.13)	(0.16)	(0.20)	(0.19)
D_1Ln IExp/ha	0.32***	0.15***	0.30***	0.40***	0.51***	0.31	0.11	0.27**	0.38***	0.54***
	(0.06)	(0.07)	(0.11)	(0.11)	(0.16)	(0.04)	(0.07)	(0.12)	(0.10)	(0.13)
D_1	-0.43	-1.24*	-0.66*	-0.20	0.38**	-0.37	-1.07**	-0.53**	-0.06	0.23*
	(0.41)	(0.64)	(0.36)	(0.37)	(0.19)	(0.32)	(0.46)	(0.25)	(0.31)	(0.12)
Ln FLabor/ha	0.12	-0.08	0.04	0.14	0.31	0.08	-0.08	0.03	0.10	0.26**
	(0.17)	(0.21)	(0.12)	(0.11)	(0.19)	(0.03)	(0.07)	(0.07)	(0.11)	(0.11)
D_2Ln HLabor/ha	0.28*	-0.21*	0.07	0.34	0.69***	0.29	0.00	0.15**	0.30**	0.48***
	(0.16)	(0.11)	(0.14)	(0.29)	(0.12)	(0.17)	(0.00)	(0.07)	(0.14)	(0.02)
D_2	0.30**	-0.16	0.20	0.36**	0.69***	0.22	-0.04	0.09	0.26***	0.54***
	(0.12)	(0.30)	(0.24)	(0.18)	(0.16)	(0.19)	(0.13)	(0.17)	(0.13)	(0.17)
RLand (dummy)	-0.15	-0.57*	-0.26*	-0.08	0.27	-0.17	-0.56**	-0.25***	-0.05	0.19
	(0.15)	(0.31)	(0.14)	(0.18)	(0.24)	(0.22)	(0.27)	(0.09)	(0.14)	(0.29)
Maize (dummy)	0.68	0.19	0.63***	0.84***	1.10***	0.41	0.01	0.30*	0.56***	0.81***
	(0.42)	(0.28)	(0.21)	(0.22)	(0.26)	(0.08)	(0.16)	(0.16)	(0.15)	(0.15)
Livest (dummy)	0.23	-0.24**	0.04	0.29*	0.66**	0.49	0.04	0.34***	0.52***	0.81***
	(0.69)	(0.12)	(0.16)	(0.15)	(0.25)	(0.11)	(0.13)	(0.10)	(0.17)	(0.14)
Title (dummy)	0.19	-0.16	0.06	0.26	0.58	0.13	-0.23	-0.01	0.17	0.45***
	(0.12)	(0.11)	(0.15)	(0.42)	(0.37)	(0.45)	(0.21)	(0.26)	(0.18)	(0.17)
Years (dummies)			No					Yes		
Farm Fixed Effects			No					Yes		
Observations			3,278					3,278		
R^2			0.61					0.76		

Notes: This Table reports elasticities at the mean, 20th (Q1), 40th (Q2), 60th (Q3) and 80th (Q4) percentiles of each regressor along with bootstrapped standard errors (400 replications) in parentheses.

***, **, * *Significant at the 1 %, 5 % and 10 %, respectively.*

Source: Own elaboration.

TABLE A2

Elasticities of farm productivity for selected nonparametric estimators with bandwidths estimated via least-squares cross validation (LSCV)

	Dependent Variable: Ln TFVO per ha									
	LLLS-Pooled					Li & Racine				
	Mean	Q1	Q2	Q3	Q4	Mean	Q1	Q2	Q3	Q4
Ln Land	-0.14 (0.39)	-0.34 (0.22)	-0.19^{*} (0.11)	-0.07 (0.11)	0.05 (0.18)	-0.13 (0.09)	-0.49^{***} (0.11)	-0.24 (0.63)	-0.07 (0.15)	0.20 (0.40)
D_1Ln IExp/ha	0.33^{**} (0.13)	0.18 (0.17)	0.30^{**} (0.15)	0.39^{***} (0.13)	0.51^{***} (0.14)	0.32^{***} (0.04)	0.09 (0.08)	0.27^{**} (0.11)	0.40^{***} (0.12)	0.57^{***} (0.09)
D_1	-0.35^{*} (0.19)	-0.99 (0.71)	-0.50 (0.37)	-0.11 (0.17)	0.22 (0.18)	-0.44 (0.31)	-1.43^{**} (0.73)	-0.69^{***} (0.23)	-0.13 (0.17)	0.45^{**} (0.20)
Ln FLabor/ha	0.11 (0.07)	-0.03 (0.11)	0.06 (0.16)	0.12 (0.17)	0.23^{***} (0.08)	0.10 (0.13)	-0.15 (0.11)	0.00 (0.00)	0.13 (0.11)	0.36 (0.28)
D_2Ln HLabor/ha	0.22^{***} (0.06	-0.01 (0.15)	0.12 (0.11)	0.25^{***} (0.08)	0.42^{*} (0.26)	0.32^{***} (0.08)	-0.23 (0.15)	0.09 (0.11)	0.38^{***} (0.12)	0.83^{***} (0.05)
D_2	0.23^{***} (0.06)	0.00 (0.00)	0.18^{*} (0.11)	0.31^{***} (0.11)	0.52^{***} (0.19)	0.27^{***} (0.10)	-0.22 (0.15)	0.12 (0.15)	0.35^{**} (0.17)	0.75^{***} (0.24)
RLand (dummy)	-0.16 (0.12)	-0.47^{**} (0.23)	-0.25^{**} (0.11)	-0.10 (0.46)	0.17 (0.22)	-0.18 (0.11)	-0.66^{***} (0.17)	-0.30 (0.24)	-0.04 (0.32)	0.30 (0.29)
Maize (dummy)	0.66^{**} (0.28)	0.23 (0.21)	0.63^{***} (0.19)	0.82^{***} (0.22)	1.03^{***} (0.21)	0.45^{***} (0.14)	-0.03 (0.10)	0.33 (0.26)	0.60^{***} (0.22)	0.89^{***} (0.24)
Livest (dummy)	0.21^{**} (0.11)	-0.11 (0.11)	0.02 (0.05)	0.19 (0.21)	0.56^{***} (0.21)	0.49^{***} (0.10)	-0.05 (0.27)	0.32^{***} (0.09)	0.58^{***} (0.15)	0.95^{**} (0.37)
Title (dummy)	0.16 (0.15)	-0.12 (0.13)	0.05 (0.29)	0.21 (0.13)	0.48^{***} (0.20)	0.13 (0.34)	-0.35^{**} (0.14)	-0.01 (0.27)	0.23^{**} (0.11)	0.61 (0.43)
Years (dummies)			No					Yes		
Farm Fixed Effects			No					Yes		
Observations			3,278					3,278		
R^2			0.59					0.83		

Notes: This Table reports elasticities at the mean, 20th (Q1), 40th (Q2), 60th (Q3) and 80th (Q4) percentiles of each regressor along with bootstrapped standard errors (400 replications) in parentheses.

***, **, * *Significant at the 1 %, 5 % and 10 %, respectively.*

Source: Own elaboration.

TABLE A3

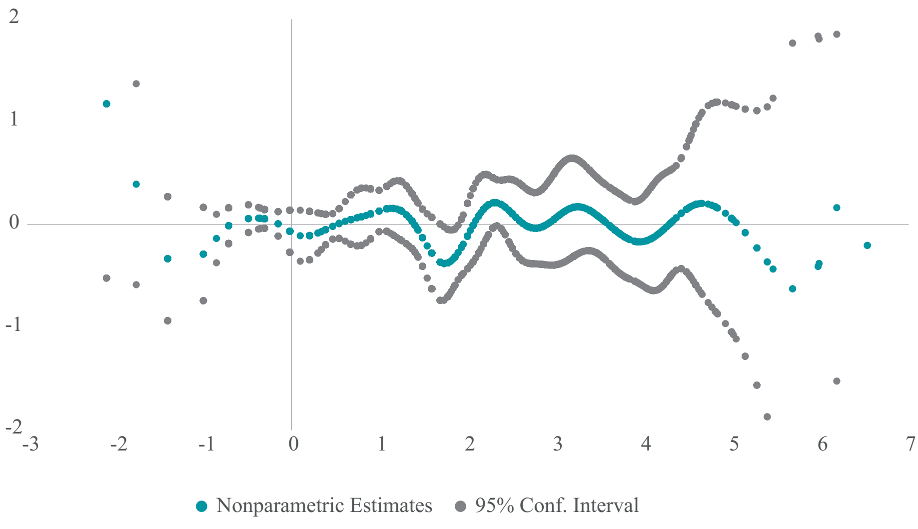
Bandwidths for covariates according to three selected methods

	ROT	AICCV	LSCV
Ln Land	0.91	1.06	0.91
D_1Ln IExp per ha	1.04	1.58	1.04
D_1	0.22	0.50	0.19
Ln FLabor per ha	0.97	1.71	1.39
D_2Ln HLabor per ha	0.56	0.97	0.80
D_2	0.27	0.66	0.28
RLand	0.29	0.92	0.87
Maize	0.28	0.89	0.92
Livest	0.25	0.78	0.78
Title	0.30	0.91	0.91

Source: Own elaboration.

FIGURE 1A

Nonparametric kernel relationship between the natural logarithm of total value of farm output per hectare and the natural logarithm of land



Note: Gray lines represent confidence intervals obtained from 400 bootstrap replications.

Source: Own elaboration.