Are users of market information efficient? A stochastic production frontier model corrected by sample selection.

J. Kamiche Zegarra[1]; B. Bravo-Ureta[2]

1: Universidad del Pacifico / University of Connecticut, Economics Department / Agricultural and Resource Economics, Peru, 2: UCONN, Ag and Resource Economics, United States of America

Corresponding author email: kamiche_jn@up.edu.pe

Abstract:

This article analyzes how information use affect farm productivity and efficiency. Our hypothesis is that farmers make better decisions when they use information (for example, choosing a high value crop combination or selling the products at higher prices) and that will enhance on productivity and efficiency. We use two techniques to mitigate the possible biases generated by observable and unobservable variables: Propensity Score Matching (PSM) for the first one and the stochastic production function (SPF) approach corrected by sample selection for the second one. We take advantage of the underused Peruvian National Agricultural Survey (ENA) which includes information about 12 877 farmers located in the Andean region. Our results show that farmers who use information are systematically nearer to their frontier than those who do not use information (0.50 vs. 0.47, on average). The analysis by plot size and age suggest that farmers with smaller plots and those who are middle age are more efficient in the users group; however, the relation is not clear among the nonusers of information. Thus, more research is needed about the complementarity of the agricultural inputs and information use. These results can contribute to the design of a cost-effectiveness evaluation of information extension programs.

Acknowledgement:

JEL Codes: Q16, O13

#2668

Title: Are users of market information efficient? A stochastic production frontier model corrected by sample selection

Abstract

This article analyzes how information use affect farm productivity and efficiency. Our hypothesis is that farmers make better decisions when they use information (for example, choosing a high value crop combination or selling the products at higher prices) and that will enhance on productivity and efficiency. We use two techniques to mitigate the possible biases generated by observable and unobservable variables: Propensity Score Matching (PSM) for the first one and the stochastic production function (SPF) approach corrected by sample selection for the second one. We take advantage of the underused Peruvian National Agricultural Survey (ENA) which includes information about 12 877 farmers located in the Andean region. Our results show that farmers who use information are systematically nearer to their frontier than those who do not use information (0.50 vs. 0.47, on average).

The analysis by plot size and age suggest that farmers with smaller plots and those who are middle age are more efficient in the users group; however, the relation is not clear among the nonusers of information. Thus, more research is needed about the complementarity of the agricultural inputs and information use. These results can contribute to the design of a cost-effectiveness evaluation of information extension programs.

Introduction

Agricultural development has an important role in poverty reduction and contributes to food security. Evidence suggests that GDP growth generated in the agricultural sector is at least twice as effective in reducing poverty as GDP growth originated in other sectors (World Bank, 2007; de Janvry & Sadoulet, 2010). Higher agricultural productivity is needed to provide food for a growing population (World Bank, 2007), and the search for mechanisms to increase agricultural productivity is an ongoing pursuit. Although information is a necessary condition to increase agricultural productivity (Anderson & Feder, 2007), few studies focus on this relationship.

Market information can enhance investing, farming and marketing decisions in agriculture and many government institutions have made efforts and huge investments to provide it with extension services (Rivera, 2011). However, there is little evidence that show how information contributes to agricultural outcomes.

Farmers' access to market information has been facilitated in the last decade by the development of Information and Communication Technologies (ICTs), and it is expected that this greater access will contribute to the growth of the agricultural sector (Aker, 2011; Ali & Kumar, 2011; Nakasone & Torero, 2016; World Bank, 2012; World Bank, 2011). The idea is that more information provided via ICTs (cellphone, Internet) will enable farmers to identify potential markets and sell their products at higher prices (Aker, 2011; Ali & Kumar, 2011). The latter assertion assumes a straightforward relationship among access to information, its use and the generation of an outcome; however, the literature does not provide strong evidence to support that claim. For example, Chong (2011) reviewed a sample of 41 ICT randomized control trial studies in Latin America and found that only 39% of the studies show a strong link between the ICT component and the outcome. Similarly, Aker, Ghosh, and Burrell (2016) highlight that the literature on the impact of information transmitted via ICTs on agricultural outcomes (e.g., output, profits) is limited, and the available evidence shows mixed results. Our research contributes to fill the gap about the effect of information use on productivity and efficiency.

Most empirical studies of the role of information have focused on the use of price data to make selling decisions (Camacho & Conover, 2011; Labonne & Chase, 2009; Nakasone, 2014). Moreover, Aker, Ghosh, and Burrell (2016) highlight that the literature on the impact of ICTs on agricultural outcomes (output, profits) is limited and shows mixed results.

The main objective of this essay is to analyze how the use of market information affects farmer productivity. We use the term "market information" to refer to a market information system that encompasses prices of inputs and outputs, and data on quantities traded (Shepherd & Shalke, 1995). Moreover, we differentiate "market information" from "marketing information," where the latter can include potential market channels, payment requirements, packaging, and other types of specifications (Shepherd, 1997).

Our hypothesis is that when farmers use market information,[1] they make better decisions in their agricultural process because they can select better crop combinations (e.g., high-value crops like fruits and vegetables); choose a better mix of inputs, given current prices; and/or decide on the best time to harvest, given output prices. We expect that farmers who use market information will have a higher level of efficiency and overall productivity than those who do not use it.

[1] Hereafter, when we use the term "information," we will be referring to the concept of market information.

Our empirical strategy has two parts, considering the potential sample selection on observable and unobservable variables that arises when we try to measure the impact of information use on farm´s productivity (Heckman, 1973; Greene, 2010). First, we use the propensity score matching (PSM) method to create a comparable sample of farms on observable variables between the users and nonusers of information. Then, we apply the stochastic frontier approach corrected by sample selection (Greene, 2010) to estimate the impact of information use on productivity, using a Cobb–Douglas (CD) formulation. This approach corrects for sample selection on unobservable variables. The results allow us to estimate the Technical Efficiency (TE) of users and nonusers of information. Our database is the Peruvian Agricultural Survey (ENA) 2015, which provides information about the farmer and farm´s characteristics and also very detailed information about agricultural production.

This essay makes two contributions: (1) it adds to the empirical literature on the productivity of small farmers in Latin America, a group that has received limited attention in the productivity literature despite its socioeconomic importance; and (2) it analyzes the effect of information on agricultural productivity, which has been a neglected topic in agricultural economics. In terms of public policy, a better understanding of how information use increases agricultural productivity can contribute to the design of effective extension programs.

The reminder of the paper is structured as follow. Section 2 describes the literature about the relation between information and efficiency and in Section 3 we present the methodological framework. Section 4 describes the empirical strategy and Section 5 presents the data. Section 6 discusses the results and we finish with our conclusions.

Information and Efficiency

Since the seminal work by Griliches (1957), who analyzed the adoption of hybrid seeds as part of the process of technological change in US agriculture, many studies have examined how information affect technology adoption (Anderson & Feder 2007; Feder, Just, & Zilberman, 1985; Feder & Slade, 1984; Genius, Pantzios, & Tzouvelekas, 2006; Rahm & Huffman, 1984). Recently, several articles study the role of ICTs (in particular, mobile phones) in the relationship between information and the agricultural sector (Aker, Ghosh and Burrell, 2016; Aker, 2011; Nakasone and Torero, 2016; Labonne and Chase, 2009; Camacho and Conover, 2011; Aker and Fafchamps, 2010, Aker, 2010; Fafchamps and Minten, 2012; Jensen, 2007; Mittal, Gandhi and Tripathi, 2010; Nakasone, 2013). Chandra Babu et al. (2012) cite several articles that identify the factors that explain the access/use of information by farmers, which include age, education, experience in farming, and farm size, among others. Most empirical studies of the role of information focus on the use of price data to make selling decisions (Aker, 2010; Camacho and Conover, 2011; Nakasone, 2013, Labonne and Chase, 2009) and much less in other stages of the agricultural cycle or in outcomes such as productivity or efficiency. Aker, Ghosh and Burrell (2016) highlight that the literature on the impact of information transmitted via ICTs on agricultural outcomes (e.g. output, profits) is limited and shows mixed results.

Few articles discuss the effect of information use on efficiency. Lio and Liu (2006) present empirical evidence about the relationship between information use and agricultural productivity, using panel data for 81 countries. These authors define use of information as the number of Internet users per 100 people, number of personal computers per 100 people, and number of telephone landlines per 100 people. Their results show that the returns from ICT in agricultural production on richer countries is two times higher than those of the poorer countries. The lack of complementary factors (e.g. human capital) is a possible explanation.

In a similar study, but using farm household data, Abdul-Salam and Phimister (2017) study the effects of access to information, measured by access to electricity, on efficiency among farmers in Uganda. The authors find a positive impact between access to information and farm efficiency, but they highlight that access to information is different from effective use, so there is a need to analyze the impact of information use on productivity. Our research aims to fill that gap.

Methodological Framework

To measure the effect of information use on productivity, we employ a Stochastic Production Frontier (SPF) approach (Aigner, Lovell, and Schmidt, 1977). In this framework, the output for each farmer *i* follows a stochastic process defined as:

$$y_i = f(X_i, \beta) + v_i - u_i \qquad (1)$$

where f generates the maximum possible output as a function of a vector X of nonstochastic inputs and β is a vector of unknown parameters. The error term is composed of two parts: v_i is associated with external events beyond the control of the farmer (the classical error term) and captures the statistical noise on observed output (Fried, Lovell and Schmidt, 2008); and u_i reflects whether the farmer is on the frontier or not. A value of u_i different from zero indicates a certain level of inefficiency. TE ranges between 0 and 1, with a value of 1 when the farm is perfectly efficient (Kumbhakar, Wang, & Horncastle, 2015; Coelli et al., 2005). The TE of each farm *i* can be measured as the ratio between observed output y_i and the output that could be produced by an efficient farm using the same set of inputs (Jondrow et al., 1982):

$$TE_i = \frac{y_i}{\exp(x_i'\beta + v_i)} = \frac{\exp(x_i'\beta + v_i - u_i)}{\exp(x_i'\beta + v_i)} = \exp(-u_i) \quad (2)$$

A shortcoming of this procedure relates to possible sample selection problem between users and nonusers of information. Farmers who use information may have skills that make them more prone to use the information and at the same time, more likely to make better administrative and technical decisions, even in the absence of information [2]. Thus, these farmers can be expected to outperform their peers and generate higher productivity and profits, independent of information usage. In his case, farmers who use information will differ from those farmers who do not use it. Not dealing with this problem could overestimate the impact of information use.

Farmer characteristics that can generate the sample selection bias can be observable (e.g., education) or unobservable (e.g., ability). We deal with these sources of sample selection using two methods: (a) Propensity Score Matching (PSM) to mitigate the sample selection bias generated by observable variables (Rosenbaum & Rubin, 1983) and then, we use (b) stochastic frontier framework corrected by sample selection (Greene, 2010) to cope with the possible sample selection generated by unobservable variables. A few recent articles have used this combination of techniques (Abdulai and Abdulai, 2017; Bravo-Ureta, Greene and Solis, 2012; De los Santos-Montero & Bravo-Ureta, 2017; Gonzalez-Flores et al., 2014, Villano et al., 2015); however, none of these papers analyze the impact of information use on agricultural productivity.

[2] The use of information implies a process of decoding and the adaptation to specific circumstances; those skills are unevenly distributed in the population.

To mitigate the problem of sample selection with respect to observable variables, the PSM strategy (Rosenbaum & Rubin, 1983) allows for the selection of two comparable groups based on observable variables between the treatment and the control groups, using a balancing score called the propensity score (PS).[3] The PS is defined as the probability that an individual will participate in the treatment, given its observed characteristics (Caliendo & Kopeinig, 2008). It is calculated as a function of observed covariates and is commonly estimated using a probit or logit model, where the outcome variable is equal to one if treated and zero otherwise (Caliendo & Kopeinig, 2008; Garrido, et al., 2014).

According to Imbens (2004), to identify the impact of the treatment on the treated (ATT) in observational studies, two conditions need to be met: (1) *Unconfoundedness,* which means that potential outcomes are independent of treatment assignment, given a group of observable covariates X[4]; and (2) *Overlap*, which means that persons with the same x values have a positive probability of being either a participant or a nonparticipant.[5] Caliendo & Kopeinig (2008) provide specific guidance and criteria for the implementation of the PSM, in particular in the selection of variables to construct the PS and in the choosing of the matching method.

After the selection of a matched sample using the PS, is it necessary to test the "balancing property" (Becker & Ichino 2002). A commonly used approach is to test if the means of all the covariates used in the estimation of the PS do not differ between treatment and control groups in the matched sample, using equally spaced intervals of the PS for the

---

[3] Rosenbaum and Rubin (1983) propose using "balancing scores" to match individuals between the treated and the control groups, based on their scores. The PS is a possible "balancing score," and it is the most used in the literature.

[4] In formal terms, $(r_1, r_0)\ ||\ D|x$, where || denotes independence and D the condition of treatment or control group. The term *unconfoundedness* used by Rosenbaum and Rubin (1983) is similar to the terms *selection on observables* and *conditional independence assumption (CIA)* (Caliendo and Kopeinig 2008).

[5] The formal representation of this condition is $0 < Prob(D = 1 \mid x) < 1$.

comparison. Within each interval, we test that the mean PS of treated and control groups does not differ. If the test fail for one interval, such interval is divided in half and test again. The process ends when the average PS for all intervals does not differ between the treated and control groups (Becker and Ichino, 2002; Garrido et al. 2014).

For the possible sample selection problem generated by unobserved characteristics, Greene (2010) specifies that the unobservable variables are correlated with the noise in the stochastic frontier model. He proposes the following models (Equations 3 and 4), as an extension of Heckman's sample selection model (Heckman 1979):

(3) Sample selection model:

$$d_i = 1[\boldsymbol{\alpha}' \boldsymbol{z}_i + w_i > 0], \qquad w_i \sim N[0,1]$$

(4) Stochastic Frontier Model:

$$y_i = \boldsymbol{\beta}' \boldsymbol{x}_i + \varepsilon_i$$

where (y_i, x_i) are observed only when $d_i = 1$; and the error term is composed by:

$\varepsilon_i = v_i - u_i$, with:

$$u_i = |\sigma_u U_i| = \sigma_u |U_i| \quad \text{where} \quad U_i \sim N[0,1],$$

$$v_i = \sigma_v V_i \quad \text{where} \quad V_i \sim N[0,1], \quad \text{and,}$$

$$(w_i, v_i) \sim N_2[(0,0), (1, \rho\sigma_v, \sigma_v^2)]$$

Equation (3) shows that d_i (e.g. the decision to use information) is a binary variable equal to one if farmer *i* uses information and zero otherwise; $\mathbf{z_i}$ is a vector of explanatory variables for the sample selection model; and w_i is the classical error tem. Equation (4) shows the typical stochastic frontier model, similar to Equation (1), but specifies that w_i and v_i follow a bivariate normal distribution where $\rho\sigma_v$ is the covariance between both variables. The statistical significance of ρ indicates the presence of a selectivity problem.

Empirical Strategy

As we mention before, we combine two methodologies to estimate the effect of information use on efficiency and overall productivity: PSM to mitigate the possible bias on observable variables and the SPF corrected by sample selection (Greene 2010) in order to mitigate bias from unobservable variables[6]. In this section, we provide details about the functional form and the variables used.

a) Propensity Score Matching

Our empirical strategy begins with the estimation of a probit model to generate a propensity score (PS) for use of information for all the observations in our database. To calculate the PS, we require a set of characteristics that influence the decision to use information but are unaffected by the treatment condition (Rosenbaum & Rubin, 1983; Caliendo & Kopeinig, 2008). These variables, based on the empirical literature and data availability,[7] include characteristics about the head of the household (age, gender, and education), the farm (land area) and geographical conditions (altitude, and location within a political region).[8] Using the estimated PS, we determine a comparable group of information users (treatment group) and nonusers (control group) based on observable characteristics. To match the observations between the treatment and the control group, we use the 1-to-1 nearest neighbor (NN) matching method and a maximum propensity score distance (caliper) of 0.25 standard deviations of the PS, which are standard methods and assumptions in the evaluation literature (Guo & Fraser, 2010; Rosenbaum & Rubin, 1985). Other studies suggest

[6] The model proposed by Greene (2010) can correct for sample selection on observable and unobservable variables at the same time, so we do a second group of estimations without the PSM step for comparison purposes. The results are very similar and are available upon request to the authors.

[7] Chandra Babu (2012) cited numerous articles that identify factors that explain the access and use of information by farmers.

[8] Peru is divided into 3 natural regions (one of them is the Andes) and 25 political regions. To deal with potential fixed effects related to the characteristics of the political regions, we include dummy variables for seven political regions that compose the Mountains (the eighth political region was the base category).

more rigorous values of the caliper such as 0.01 (Bellemare and Novak 2016) or 0.001 (Bellemare & Novak 2016, Gonzalez-Flores et al., 2014) and other methods of matching such as kernel or Mahalanobis (Calindo and Kopeinig, 2010). We use some of these methods to determine the robustness of our analysis.

b) SPF corrected by sample selection, in a comparable group

After we have a matched sample of users and nonusers of information, we apply the SPF model corrected by sample selection by estimating Equations 3 and 4. Thus, we model the decision to use information (Equation 3) by the farmer *i* as a function of the characteristics of the farmer, the farm, and geographical conditions:

$$d_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{ji} z_{ji} + \sum_{k=1}^{K} \alpha_{ki} z_{ki} + \sum_{m=1}^{M} \alpha_{mi} z_{mi} + w_i \qquad (5)$$

where z_j represents the characteristics of the farmer (e.g., age, gender, education, household size, access to credit); z_k are the characteristics of the farm (e.g., farm size); z_m represents geographical conditions (e.g., altitude); and w_i is the error term. As we mentioned before, we include these variables based on the literature and data available.

Then, we use a CD functional form of Equation (4) to estimate the effect of information use on productivity:

$$Ln(Y_i) = \alpha + \sum_{j=1}^{J} \beta_j Ln\left(X_{ji}\right) + \theta ALT + \sum_{s=1}^{S} \gamma_i HHH_i + v_i - u_i \quad \text{iff } d_i = 1 \qquad (6)$$

where Y_i is the total value of production of farm i; X_{ji} are the conventional inputs used by farm *i* measured in quantities (e.g., land size, family labor) or in values at constant prices (e.g., expenses on hired labor, fertilizers, or seeds); geographical conditions represented by the variable ALT which is the altitude where the farmer's plot(s) are located; and HHH represents characteristics of the head of household (e.g., age, gender, education). Given the

CD specification, the parameters for the inputs are partial elasticities for each of the inputs, and in the case of the dummy variables (e.g., gender), we need to calculate $[\exp(\delta) - 1] * 100$ to get the percentage effect of gender on Y_i (Halvorsen & Palmquist, 1980). Following Abdulai and Abdulai (2017), Gonzalez-Flores et al. (2014), Greene (2010) and Villano et al. (2015), we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) approach for the estimation of the parameters and we obtain the asymptotic standard errors using the Berndt–Hall–Hall–Hausman (BHHH) estimator (Gonzalez-Flores et al. 2014, Greene, 2010).

Finally, we estimate the TE for the users and nonusers of information applying the Jondrow et al. (1982) procedure. We estimate the TE for users and nonusers of information separately and with respect to different size of the land (e.g. all the matched sample, farmers who have less than 100 hectares and those who have less than 50 hectares). We compare the TE estimates within each the group, to determine how far or near they are from their optimum output. We analyze our TE results considering subgroups by gender, age and land size[9].

<u>Data</u>

We use the Peruvian National Agricultural Survey (ENA) for 2015, a new cross-sectional dataset that provides information about 12,887 agricultural units located in the Andean region. The database provides detailed information about farmers' characteristics (e.g. age, gender, education), characteristics of the farm (e.g. plot size, number of crops grown), amounts of inputs used (land, labor, seeds, fertilizer), economic variables about agricultural activity (e.g. cost of inputs, prices and quantities of products sold, among others) and the decision to use information (USEINF) or not. As we mentioned before, our

[9] Recent papers have highlighted that the comparison of TE between groups needs to be done within the groups (in relative terms) and not in an absolute perspective, because it is very likely that the frontier of each group differs (Gonzales-Flores et al. 2014; De los Santos-Montero & Bravo-Ureta 2017).

first step is to estimate a PSM to determine a comparable group on observable variables of user and nonusers of information. Following the literature, we use the variables LAND (in hectares), AGE, GENDER, education (ELEM, HIGH, HIGHER) and ALTITUDE to determine the comparable group of users and nonusers. The original sample includes 12 877 observations, where 4 705 farmers use information. One of the characteristics of the farmers located in the Andean region of Peru is that they have very small plots (Trivelli, Escobal and Revesz 2009)[10]. Escobal and Armas (2009, p. 61) define the small agriculture in Peru as the one where farmers have up to 50 hectares of land for agriculture activity. In the database ENA 2015, only 118 of the total sample have 50 hectares or more, so we do our analysis using the information of the farmers who have up to 50 hectares. We apply the PSM procedure to find a comparable group of users and nonusers, and we get a sample of 8000 observations, evenly distributed between users and nonusers of information. As we mentioned in the section about the methodology, we do several estimations varying the size of the caliper, the matching method (e.g. number of neighbors in the nearest neighbor method) and the upper limit of land size[11]. However, the descriptive statistics and the results presented on the following tables is focused on the matched sample of 8000 observations. On Table 1 we present (in bold) the mean of the variables used to estimate the PSM, for the unmatched sample and for the matched sample[12].

Table 1: Mean of the Variables for the Probit Model, Unmatched and Matched Sample

.[10] Trivelli, Escobal and Revesz (2009: p.207) shows that on average a family member that works in a plot located in the Andean region have 0.16 hectares per person, in comparison to the 0.43 hectares of a farmer located in the Coast.

[11] For determining the robustness of our analysis, we do additional estimations using the observations that have less or up to 100 hectares and also with the total sample. The results are very similar and are available upon request.

[12] On Appendix 1 we present the mean of the variables of an additional matched sample (less or up to 100 hectares=.

Variable	Description	Total Sample – Unmatched				Matched Sample (		
		All	Users	Nonusers	Diff	Users	Nonusers	Diff
Farmer characteristics								
AGE [1/.]	**HH age in years**	**52.54**	**50.55**	**53.69**	***	**50.60**	**50.08**	
GENDER [1/.]	**= 1 if HH male**	**0.69**	**0.73**	**0.67**	***	**0.72**	**0.73**	
ELEM [1/.]	**= 1 if HH finished elementary**	**0.53**	**0.54**	**0.53**		**0.53**	**0.53**	
HIGH [1/.]	**= 1 if HH finished highschool**	**0.24**	**0.27**	**0.23**	***	**0.28**	**0.28**	
HIGHER [1/.]	**=1 if HH has upper level studies**	**0.07**	**0.07**	**0.06**	*	**0.08**	**0.08**	
SPANISH	**= 1 if HH native language is spanish**	**0.39**	**0.37**	**0.41**	***	**0.40**	**0.37**	
QUECHUA	**= 1 if HH native language is quechua**	**0.55**	**0.60**	**0.52**	***	**0.57**	**0.60**	
TRAINING	= 1 if HH received training	0.11	0.16	0.09		0.15	0.10	
OTHER INC	= 1 if HH has other income	0.48	0.53	0.45		0.52	0.50	
LOAN	= 1 if HH asked for a loan	0.10	0.13	0.08		0.14	0.09	
INSURANCE	= 1 if HH had insurance	0.02	0.02	0.02		0.02	0.02	
SAVINGS	= 1 if HH had savings	0.21	0.23	0.19		0.23	0.20	
SIZE	Household size	3.71	3.91	3.59		3.87	3.77	
Farm characteristics								
LAND	Land size, in hectares	3.67	4.35	3.28	***	2.98	2.93	
YEARS	years in the farm activity	27.05	25.76	27.79		25.75	25.11	
CROPS	number of crops	4.46	4.74	4.30		4.69	4.08	
SOIL	= 1 if soil affected by salinity	0.67	0.71	0.65		0.68	0.64	
PLAGUE	= 1 If soil affected by plague	0.10	0.09	0.10		0.10	0.11	
Geographical Conditions								
ALT	Altitude in 1000 mts	3.24	3.27	3.22	***	3.22	3.23	
DISTANCE	distance to downtown, in hours	1.73	1.64	1.78	***	1.66	1.66	
Number of observations		**12877**	**4705**	**8172**		**4000**	**4000**	

1/. These variables are used in the Stochastic Frontier Model too.
Note: ***, **, ** Significant at 1%, 5% and 10% levels, respectively.
Source: National Agricultural Survey (ENA), 2015.

The second step of our methodology is the estimation of the SPF model corrected by sample selection. In this case, first we need to estimate the determinants of the decision of using information (Equation 5) and then we estimate the SPF model (Equation 6). For Equation (5), Table 1 includes the mean of the additional variables (not in bold). After the estimation of Equation (5), we need to test for the balancing condition which means that there is no significant difference between the users and nonusers of information in observable variables in the matched sample. Table 1 shows that before the matching there were significant differences between the groups in most of the variables, but after the matching process, the balancing condition is accomplished.

In Equation 6, we estimate the production equation using the variable total value of production as dependent variable[13] and the explanatory variables are the traditional inputs for the agriculture production and some of the farmers and farm characteristics and geographical conditions included in Equation 5. In Table 2, we present the mean of the variables used in the estimation of the production model[14], for the unmatched and matched sample. In general, the users of information apply more inputs in their agricultural production than the farmers who do not use information, in both samples.

Table 2: Variables for the Unmatched and Matched Sample, SPF Model

Variable	**Description**	**All**	**Unmatched Sample**		**Matched Sample**	
			Users	**Non-Users**	**Users**	**Non-Users**
Dependent Variable:						
PROD	Value of production, in soles	6,598.86	8,485.63	5,512.56	8,770.19	6,133.98
Traditional Inputs						
HIRE	Value of labor, in soles	756.65	1,048.50	588.62	1,082.92	646.08
FAM	Number of workers in the family	2.56	2.69	2.49	2.66	2.58
SEED	Value of seed, in soles	267.25	343.19	223.52	351.82	253.39
FERT	Value of fertilizer, in soles	209.60	288.50	164.17	301.43	195.25
PEST	Value of pesticide, in soles	130.80	174.11	105.85	188.39	124.53
PSEED	Proportion of production used as seeds	0.06	0.06	0.06	0.06	0.06
PCERT	Proportion of certified seeds used	0.03	0.03	0.02	0.04	0.02
Number of Observations		**12877**	**4705**	**8172**	**4000**	**4000**

Source: National Agricultural Survey (ENA), 2015.

Results and Discussion

In this section, we discuss the results for selection model (Equation 5) in the SPF approach corrected by sample selection and the results of the production model (Equation 6). In the

[13] We use a constant price for each of the product for all the observation in the sample, for avoiding the price effect in our dependent variable.
[14] The SPF model corrected by sample selection, also includes the variables age, gender, education (elem, high, higher) and education of the spouse; the mean of those variables is included in Table 1.

last part, we discuss the technical efficiency results for the users and nonusers of information.

Chandra Babu et al. (2012) cited numerous articles that identify the factors that explain the use of information by farmers; some of the variables are: age, education, experience in farming, business characteristics, farm size, type of farm enterprise, debt level, ownership status, geographical characteristics such as distance to market centers, among others[15]. Following the literature, Table 3 shows the results of the estimation of the probit model with the matched sample, using the variables presented in Table 1. The results show that variables such as having a female head of the household, receiving training about agricultural activities, getting another income and having financial activities (asking for loans and/or having savings), size of land, number of crops, quality of soil and being affected by a plague, increase the probability of using information. As expected, those farmers who have received training in agricultural activities have 0.22 points higher probability to use agricultural information, in comparison to those who do not receive that training. This variable has the largest effect as a determinant of information use. Also, those farmers who have financial activities, such as asking for a loan and/or having savings increase the probability of using information in 19 points and 8 points, respectively. About the farm characteristics, having more land or a greater number of crops increase the probability of using information. This can be explained because more land available can imply the need of selecting more crops and for that, more information is needed (e.g. to choose crops with higher prices or those who have a secure market). For the same reason, a more diversified bundle of crops (e.g. a greater number of crops) requires the use of more information (e.g.

[15] Some of these articles are: Carter and Batte (1993), Waller et al. (1998), Schnitkey et al. (1992), Ngathou, Bukenya and Chembezi (2006), Solano et al. (2003), Alvarez and Nuthall (2006), Llewellyn (2007).

for trying to find higher prices for selling). It is interesting to find that if the farm has been affected by soil quality problems (salinity or plague), more information is used. A possible explanation is that farmers search for solutions for their soil quality problems; however, this hypothesis needs to be tested in future research.

Table 3: Results for the Selection Model (Equation 5)

Variable	Coefficient		Standard Error
Farmer´s characteristics			
AGE	-0.0001		0.0061
AGE -SQ	0.0000		0.0001
GENDER	-0.0697	*	0.0359
ELEM	-0.0500		0.0502
HIGH	-0.0564		0.0590
HIGHER	-0.0827		0.0750
SPOUSE	-0.0470		0.0438
SIZE	0.0094		0.0079
SPANISH	-0.0769		0.0965
QUECHUA	-0.1111		0.0953
TRAINING	0.2286	***	0.0448
OTHER INC	0.0906	***	0.0317
LOAN	0.1926	***	0.0464
INSURANCE	-0.0685		0.1123
SAVINGS	0.0816	**	0.0363
Farm´s characteristics			
LAND	0.0590	***	0.0107
YEARS	0.0010		0.0018
CROPS	0.0276	***	0.0048
SOIL	0.1455	***	0.0358
PLAGUE	0.0934	*	0.0537
Geographical Conditions			
ALT	0.2717		0.1761
ALT - SQ	-0.0432		0.0299
DISTANCE	0.0007		0.0067
Constant	-0.6645		0.3167
Regional Controls		YES	
Log – likelihood function			-5,424.35
Chi – squared test statistic			241.65
Number of observations			8000

Source: National Agricultural Survey (ENA), 2015.

On the other side, contrary to previous studies (Chandra Babu et al. 2012), education variables, age, household size and geographical variables, do not have any effect on the decision of using information. This result can be explained because in the first step of the methodology we applied the PSM on those variables for finding a matched sample; therefore, the are no differences between the users and nonusers on that variables and they cannot explain the decision to use information in the second step. One unexpected result is that the gender variable has a statistical significant but negative effect: those farms that have a female as head of the household have a higher probability to use information in comparison to those who are headed by a male. This result needs more study for proposing an explanation.

After we estimate the selection equation and continuing with our second step, we estimate the production equation (Equation 6). Because we use a CD formulation for agriculture production, the estimates of the coefficients are partial elasticities of the traditional inputs on production.

For comparison purposes, Table 4 presents the estimation of four SPF Models. The first two columns are estimation of conventional SPF models, ignoring the possible bias on sample selection, using the unmatched (Column A) and matched (Column B) samples. The last two column present the results of SPF models corrected by sample selection, for the users (Column C) and nonusers of information (Column D).

Table 4: Results for the SPF Model (Equation 6)

Models / Number of observations	Conventional SPF				SPF corrected by Sample Selection			
	All sample – Unmatched (A)		All sample – Matched (B)		Users - Matched (C)		Nonusers – Matched (D)	
	N = 12877		N = 8000		N = 4000		N = 4000	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
LAND	0.4583 ***	0.0064	0.4643 ***	0.0080	0.4314 ***	0.0095	0.4060 ***	0.0120
HIRE	0.0306 ***	0.0016	0.0318 ***	0.0020	0.0347 ***	0.0026	0.0315 ***	0.0027
FAM	0.0617 ***	0.0169	0.0378 *	0.0216	0.0486	0.0299	0.0236	0.0332
SEED	0.1465 ***	0.0041	0.1524 ***	0.0052	0.1509 ***	0.0046	0.1802 ***	0.0054
FERT	0.0248 ***	0.0017	0.0245 ***	0.0022	0.0315 ***	0.0029	0.0199 ***	0.0034
PEST	0.0371 ***	0.0018	0.0355 ***	0.0022	0.0294 ***	0.0031	0.0408 ***	0.0036
PSEED	-0.1722 ***	0.0044	-0.1771 ***	0.0056	-0.1548 ***	0.0072	-0.1713 ***	0.0085
PCERT	0.0290 ***	0.0066	0.0230 ***	0.0078	0.0207 **	0.0090	0.0442 ***	0.0121
ALT	-0.8409 ***	0.1003	-1.0703 ***	0.1246	-1.7111 ***	0.1648	-0.9712 ***	0.1555
ALT-SQ	0.0987 ***	0.0169	0.1362 ***	0.0211	0.2252 ***	0.0283	0.1299 ***	0.0269
SOIL	-0.0606 ***	0.0177	-0.0674 ***	0.0220	-0.1293 ***	0.0322	-0.0870 **	0.0345
AGE	0.0164 ***	0.0033	0.0161 ***	0.0042	0.0123 **	0.0057	0.0173 **	0.0063
AGE - SQ	-0.0002 ***	0.0000	-0.0002 ***	0.0000	-0.0001 ***	0.0001	-0.0002 ***	0.0001
GENDER	0.1150 ***	0.0195	0.1014 ***	0.0248	0.1149 ***	0.0360	0.1081 ***	0.0397
ELEM	0.0669 ***	0.0253	0.0915 ***	0.0351	0.1089	0.0507	0.0695	0.0573
HIGH	0.2170 ***	0.0309	0.2350 ***	0.0409	0.1949 ***	0.0571	0.2343 ***	0.0668
HIGHER	0.2940 ***	0.0409	0.3096 ***	0.0515	0.3436 ***	0.0710	0.2085 ***	0.0794
SPOUSE	-0.0472 **	0.0241	-0.0114	0.0308	0.0287	0.0434	-0.0258	0.0521
USEINF	0.1052 ***	0.0180	0.1167 ***	0.0205	-		-	
Constant	7.9970	0.1770	8.2387 ***	0.2204	10.0830 ***	0.2996	7.8528 ***	0.2988
Log - Likelihood	-17 248.65		-10 638.57					
λ	0.8354 ***	0.0228	0.8295 ***	0.0290	-		-	
σ	0.8908 ***	0.0095	0.8684 ***	0.0121	-		-	
Sigma(u)	-		-		0.9236 ***	0.0455	1.1562 ***	0.0413
Sigma(v)	-		-		0.7269 ***	0.0328	0.8365 ***	0.0294
ρ (w, v)	-		-		-0.4842 ***	0.1021	0.7953 ***	0.0387

Significance at 1% (***), 5% (**) and 10% (*).

The conventional SPF model in Column A ignores possible bias of sample selection and uses all the sample available (N = 12 877). We include a dummy variable named USEINF that captures the effect of the decision to use information (=1) and the results suggest that information use has a positive effect on agricultural productivity; however, the possible bias on sample selection prevents to conclude a causal relationship with these results. In Column B we estimate the same model but in this case we use the matched sample (N = 8 000), meaning that the users and nonusers of information have no differences on observable variables. The result confirms the previous estimation and it suggests that there is a strong positive effect of the use of information on agricultural productivity; but again this results can be affected by the sample selectivity issue.

In the estimation of the sample selection models, the last row in Table 4 for the last two columns shows that the selectivity coefficient (ρ) is statistically significant different from zero for the users and nonusers of information. This result justifies the estimation of two separated SPF models, one for each group. The partial elasticities presented in Columns C and D for the traditional inputs are all positive for LAND, LABOR (HIRE, FAM), SEED, FERTILIZER and PESTICIDE, which are common results in SPF models (Abdul-Salam and Phimister, 2017; Bravo-Ureta, Greene and Solis 2012; Villano et al. 2015). For example, an increase of 10% of the available land increases the productivity in 4.1%. Thus, 10% of more labor (hire or familiar) generates an increment of 2% - 3% on productivity (depending on the specification of the model).

The elasticity for the proportion of certified seeds used (PCERT) is positive but small (Villano et al. 2015) and the estimate for the proportion of production used as seed (PSEED) is negative and different from zero, and that can be explained as an effect of seed quality[16]. In case of geographical variables, if the plot is located in a higher altitude (ALT), the effect is negative, meaning that the higher areas of the mountains are less productive; moreover, the effect intensifies as the location is higher (ALT-SQ). With respect to socioeconomic variables such as age, gender and education, they have the same sign found in previous studies. In the case of education, it is interesting to notice that for users, the effect of having more studies than high school (HIGHER) is greater than having only high school (HIGH), but that is not the case for nonusers, where the effect is opposite. In either case, having elementary studies has any effect on productivity. All our results are robust for the different specifications and samples.

Finally, we calculate the TE for users and nonusers of information, using the definition by Jondrow et al. (1982) which is specified in Equation 2; Table 5 shows our results. For both groups, the TE calculated using SPF Conventional estimation is higher than all of SPF models corrected by sample selection, showing that ignoring the issue implies an overestimation of the parameters. In the case of users, the TE estimations are very similar using the matched sample and the other samples; however, in the case of the nonusers of information, the farmers that have smaller plots show slightly smaller TE than the complete matched sample. It is not possible to compare the TE between the users and nonusers of information because it is expected that each of the groups has its own reference frontier. However, it is possible to say that the users of information are nearer to their own frontier than the nonusers of information (all their TE values are greater than 0.50), thus, they are relatively more efficient.

---

[16] Seed quality is an important issue in the agricultural activity in Peru, because one of the reasons of the low productivity is the use of a proportion of last year production as the seed for the next season (Trivelli, Escobal and Revesz 2009) and the limited use of certified seeds.

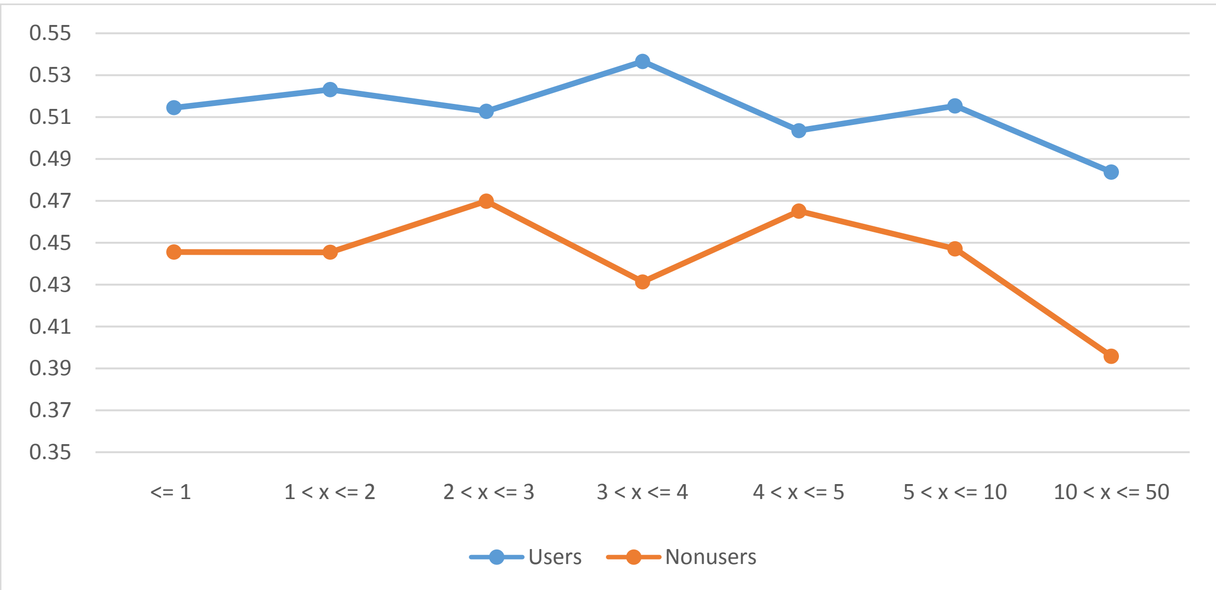
Table 5: Technical Efficiency estimates for Users and Nonusers of information

Variable		Number of Obs.	Mean	Standard Deviation	Minimum	Maximum
Users of information						
TE_User	SPF Conventional Model	4,705	0.6016	0.1621	0.0035	0.8887
TE_User	All the matched sample	4,023	0.5021	0.1537	0.0303	0.8643
TE_User	Matched & <100 hectares	4,019	0.4987	0.1570	0.0333	0.8721
TE_User	Matched & <50 hectares	4,000	0.5149	0.1497	0.0422	0.8679
Nonusers of information						
TE_Nonuser	SPF Conventional Model	8,172	0.5794	0.1624	0.0080	0.8877
TE_Nonuser	All the matched sample	4,023	0.4731	0.1584	0.0262	0.8408
TE_Nonuser	Matched & <100 hectares	4,019	0.4636	0.1639	0.0281	0.8448
TE_Nonuser	Matched & <50 hectares	4,000	0.4442	0.1725	0.0194	0.8401

Source: Authors´elaboration.

Finally, we analyze the TE estimates in each group of users and nonusers, considering the age, gender and land distribution. On Figure 1, we present the TE estimates by each group considering 7 groups of land size, between 0 to 50 hectares. As we can observe, the farmers that have more land (more than 5 hectares) are further from their own frontier than those who have less land available and that can be related to the inverse relationship between land and productivity (Carletto, Savastano & Zezza, 2013; Eastwood, Lipton & Newell, 2010).

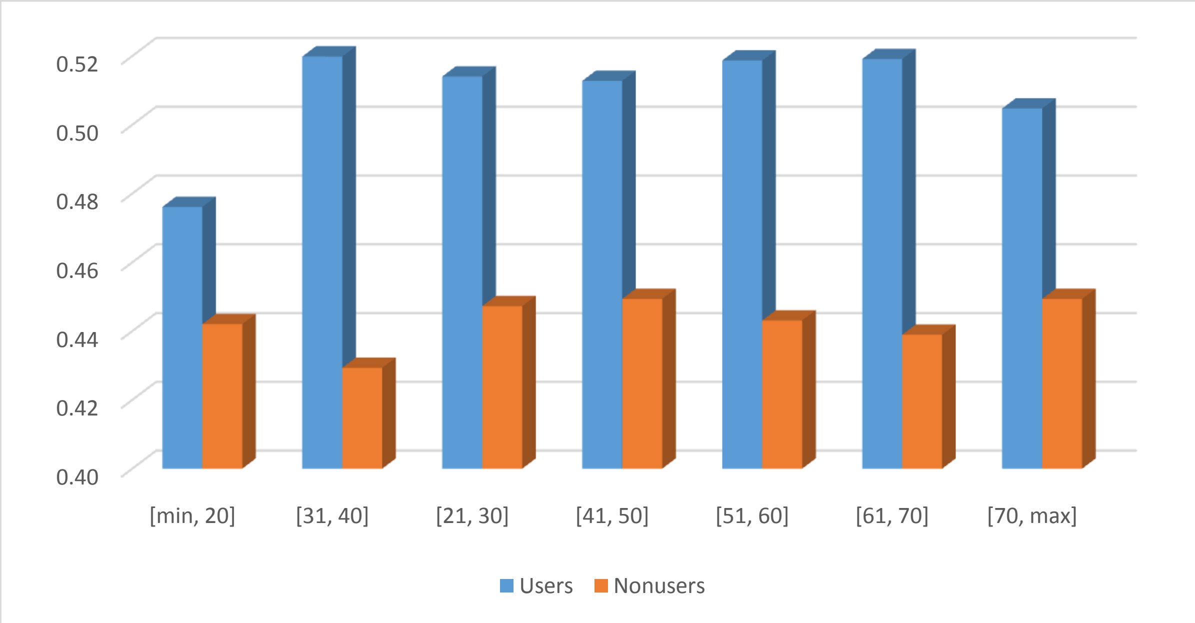
Figure 1: Users and Nonusers TE estimates by land size.



Source: Authors´elaboration.

In the case of gender, there are no differences in the TE estimates for the users and nonusers of information. The last part of our analysis is focused on the age of the head of the household.

Figure 2: Users and Nonusers TE estimates, by age of Head of the Household



Source: Authors´elaboration.

Figure 2 shows that the TE in the nonusers is very similar among the age groups, except the case of the head of the households who are in the 31-40 years old. Those farmers have the lowest TE among the nonusers of the information (0.43). In the case, of the user of information, the youngest group of farmers (less than 20 years old) shows the lowest TE estimate (0.47) because all the other groups of users show an average of 0.50 or more. This result can be explain because with age farmers gain more experience in the activity and that will probably allow them to make better decisions when they use the information available. Again, it is necessary to analyze more deeply the relations of complementarity between agricultural factors and the use of information.

Conclusions

This article analyzes how information use affect farm productivity and efficiency. Our hypothesis was that farmers make better decisions when they use information (for example, choosing a high value crop combination or selling the products at higher prices) and that will enhance on productivity and efficiency. We use two techniques to mitigate the possible biases generated by observable and unobservable variables: Propensity Score Matching (PSM) for the first one and the stochastic production function (SPF) approach corrected by sample selection for the second one. We use the Peruvian National Agricultural Survey (ENA) which includes information about 12 877 farmers located in the Andean region and our results show that farmers who use information are systematically nearer to their frontier than those who do not use information (0.50 vs. 0.47, on average). The results are robust when we analyze different samples (e.g. differentiated by land size). For example, those farmers who have 50 or less hectares and use information, have a TE estimate of 0.51 and those who do not use information and have the same range of hectares, have TE estimates of 0.44.

The analysis by plot size and age suggest that farmers with smaller plots and those who are middle age are more efficient in the users group; however, the relation is not clear among the nonusers of information. Thus, more research is needed about the complementarity of the agricultural inputs and information use.

These results are important because they can contribute to the design of extension programs that deliver agricultural information, because they provide a quantitative measure of its impact on specific outcomes (productivity and efficiency). These results can be used in public policy evaluation as an indicator for the estimation of the cost-effectiveness of extension programs.

<u>References</u>

Abdulai, A.N., and A. Abdulai. 2017. "Examining the impact of conservation agriculture on environmental efficiency among maize farmers in Zambia." Environment and Development Economics 22: 117-201.

Abdul-Salam, Y. and Phimister, E. (2017). Efficiency effect to access to information on small-scale agriculture: Empirical evidence from Uganda using stochastic frontier and IRT models. *Journal of Agricultural Economics*, 68 (2): 494-517.

Aigner, D., Lovell, C. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21-37.

Aker, J., Ghosh, I., and Burrell, J. (2016). The promise (and pitfalls) of ICT for agriculture initiatives. *Agricultural Economics* 47 (51): 35-48.

Aker, J. (2010). Information from markets near and far: Mobile phones and agricultural markets in Niger. *American Economic Journal: Applied Economics* 2 (3): 46-59.

Aker, J. (2011). Dial “A” for agriculture: A review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics* 42: 631-647.

Aker, J. and Fafchamps, M. (2010). How does mobile phones coverage affect farm-gate prices? Evidence from West -Africa. *World Bank Economic Review* 29 (2): 262 -292.

Ali, J., and Kumar, S. 2011. "Information and communication technologies (ICTs) and farmers' decision - making across the agricultural supply chain." International Journal of Information Management (31): 149-159.

Anderson, J. and Feder, G. (2007). Agricultural Extension. In Evenson, R. and Pingali, P. (eds) *Handbook of Agricultural Economics, Volume 3* . Elsevier B .V., 2344-2378.

Becker, S., and A. Ichino. 2002 . "Estimation of average treatment effects based on propensity scores." The Stata Journal 2 (4): 358 -377.

Bellemare, M., and L. Novak. 2016. "Contract farming and food security." American Journal of Agricultural Economics 99 (2): 357-378.

Bravo-Ureta, B., W. Greene, and D. Solis. 2012. "Technical efficiency analysis correcting for biases from observed and unobserved variables: An application to a natural resource management project." Empirical Economics 43: 55-72.

Caliendo, M and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching *Journal of Economic Surveys*. 22 (1): 31-72.

Camacho, A. and Conover, E. (2011). *The Impact of Receiving Price and Climate Information in the Agricultural Sector.* IDB Working papers Series N. IDB WP 220.

Carletto, C., S. Salvastano and A. Zezza (2013). Fact or artifact: The impact of measurement errors on the farm size-productivity relationship. *Journal of Development Economics*. 103: 254-264.

Chandra Babu, S., Glendenning, C., Asenso-Okyere, K. and Kumar Govindarajan, S. (2011). Farmers' Information Needs and Search Behaviors: Case Study in Tamil, Nadu, India. IFPRI. http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/126836.

Chong, A. 2011. Development Connections: Unveiling the impact of new information technologies. New York: Inter-American Development Bank.

Coelli, T., Prasada, D., O'Donnell, C. and Battese, G. (2005). *An Introduction to Efficiency and Productivity Analysis .* New York: Springer.

de Janvry, A. and Sadoulet, E. (2010). Agricultural growth and poverty reduction: Additional evidence. *The World Bank Research Observer* 25 (1): 1-20.

De los Santos-Montero, L., and B. Bravo-Ureta. 2017. "Productivity effects and natural resource management: Econometric evidence from POSAF-II in Nicaragua." Natural Resources Forum (NJ) 1-13.

Eastwood, R., M. Lipton and A. Newell (2010). Farm size. In: Evenson, Robert, Pingali and Prabhu (Eds). Handbook of Agricultural Economics, Vol 4. Elsevier.

Escobal, J., and C. Armas. 2015. "El uso de encuestas y censos agropecuarios para desarrollar una tipologia de la pequeña y mediana agricultura familiar en el Peru." In Agricultura peruana: Nuevas miradas desde el censo agropecuario, by J. Escobal, R. Fort and E. Zegarra, 15-86. Lima: GRADE.

Fafchamps, M. and Minten, B. (2012). Impact of SMS-based agricultural information on Indian farmers. *The World Bank Economic Review* 26 (3): 383-414.

Feder, G. and Slade, R. . (1984). The Acquisition of information and the adoption of new technology. *American Journal of Agricultural Economics* 66 (3): 312-320.

Feder, G., Just, R. and Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change* 33 (2): 255-298.

Fried, H., C.A. Lovell, and S. Schmidt. 2008. "Efficiency and productivity." In The measurement of productive efficiency and productivity growth, by H. Fried, C.A. Lovell and S. Schmidt, 3-91. New York: Oxford University Press.

Garrido, M., A. Kelley, J. Paris, K. Roza, D. Meier, R. Morrison, and M. Aldridge. 2014. "Methods for constructing and assessing propensity scores." Health Services Research 49 (5): 1701-1720.

Genius, M., Pantzios, C., and Tzouvelekas, V. (2006). Information acquisition and adoption of organic farming practices. *Journal of Agricultural and Resource Economics* 31 (1): 93-113.

Gonzalez-Flores, M., B. Bravo-Ureta, D. Solis, and P. Winters. 2014. "The impact of high value markets on smallholder productivity in the Ecuadorean Sierra: A Stochastic Production Frontier Approach correcting for selectivity bias." Food Policy 44 : 237-247.

Greene, W. (2010). A stochastic frontier model with correction for sample selection. *Journal of Productivity Analysis* 34: 15 -24.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica* 25 (4): 501-522.

Guo, S., and M. Fraser. 2010. Propensity score analysis, statistical methods and applications. California: Advanced Quantitative Techniques in the Social Sciences, SAGE Publications.

Halvorsen, R. and Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* 70 (3): 474-475.

Heckman, J. 1979. "Sample selection bias as a specification error." Econometrica 47 (1): 153-161.

Imbens, G. 2004. "Nonparametric estimation of average treatment effects under exogeneity: a Review." The Review of Economics and Statistics 86 (1): 4-29.

Jensen, R. (2007). The digital provide: Information (technology), market performance and welfare in the South Indian fisheries sector. *The Quarterly Journal of Economics* 122 (3): 879-924.

Jondrow, J., Lovell, C. A., Materov, I. and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19: 233-238

Kumbhakar, S., Wang, H. and Horncastle, A. (2015). *A Practitioner's Guide to Stochastic Frontier Analysis using Stata* . New York: Cambridge University Press.

Labonne, J. and Chase, R. (2009). *The Power of Information: The Impact of Mobile Phones on Farmers' Welfare in the Philippines* . Washington DC: World Bank .

Lio, M. and Liu, M. (2006). ICT and agricultural productivity: Evidence from cross-country data. *Agricultural Economics* 34: 221-228.

Mittal, S., Gandhi, S. and Tripathi, G. (2010). *Socio-economic Impact of Mobile Phones on Indian Agriculture.* New Delhi: Indian Council for Research on International Economic Relations.

Nakasone, E. (2013). The Role of Price Information in Agricultural Markets: Experimental Evidence from Rural Peru. Paper presented at the 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013. Available at: http://ebrary.ifpri.org/cdm/ref/collection/p15738coll5/id/4039

Nakasone, E. and Torero, M. (2016). A text message away: ICTs as a tool to improve food security. *Agricultural Economics* 47 (51): 49-59.

Rahm, M. and Huffman, W. (1984). The adoption of reduced tillage: The role of human capital and other variables. *American Journal of Agricultural Economics* 66 (4): 405-413.

Rivera, W. 2011. "Public sector agricultural extension system reform and the challenges ahead." The Journal of Agricultural Education and Extension 17(2): 165-180.

Rosenbaum, P., and D. Rubin. 1985. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistian* 39 (1): 33- 38.

Rosembaum, P. and Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41-55.

Shepherd, A. (1997). *Market Information Services: Theory and Practice.* Rome: FAO - Food and Agriculture Organization of the United Nations.

Shepherd, A. and Shalke, A. (1995). *An Assessment of the Indonesian Horticultural Market Information Service.* Rome: FAO-AGSM Occasional Paper N. 8.

Trivelli, Escobal, and Revesz. 2009. Desarrollo rural en la sierra: aportes para el debate. Primera Edición. Lima: CIPCA, GRADE, IEP, CIES.

Villano, R., B. Bravo-Ureta, D. Solis, and E. Fleming. 2015. "Modern rice technologies and productivity in the Philippines: Disentangling technology from managerial gaps." Journal of Agricultural Economics 66 (1): 129 - 154.

World Bank. 2011. ICT in agriculture: Connecting smallholders to knowledge, networks and institutions. Washington DC: World Bank.

World Bank. (2007). *World Development Report 2008: Agriculture for Development.* New York: World Bank.

Appendix 1: Mean of the variables for Matching, using different land size.

Variable	Description	Matched Sample - All		Matched Sample (Land < = 100 hectares)		Matched Sample (Land < = 50 hectares)	
		Users	Nonusers	Users	Nonusers	Users	Nonusers
Farmer´s characteristics							
AGE [1/..]	HH age in years	50.55	50.22	50.60	50.14	50.60	50.08
GENDER [1/.]	HH gender	0.72	0.73	0.72	0.73	0.72	0.73
ELEM [1/.]	HH finished elementary	0.53	0.54	0.53	0.52	0.53	0.53
HIGH [1/.]	HH finished highschool	0.28	0.28	0.28	0.28	0.28	0.28
HIGHER [1/.]	HH has upper level studies	0.07	0.08	0.08	0.08	0.08	0.08
SPOUSE	spouse has no education	0.14	0.16	0.14	0.16	0.15	0.15
SIZE	Household size	3.89	3.78	3.87	3.79	3.87	3.77
SPANISH	= 1 if native language is spanish	0.40	0.37	0.40	0.38	0.40	0.37
QUECHUA	= 1 if native language is quechua	0.57	0.60	0.57	0.60	0.57	0.60
TRAINING	HH received training	0.16	0.09	0.15	0.10	0.15	0.10
OTHER INC	HH has other income	0.52	0.50	0.52	0.51	0.52	0.50
LOAN	= 1 if asked for a loan	0.14	0.09	0.14	0.10	0.14	0.09
INSURANCE	= 1 if had insurance	0.02	0.01	0.02	0.02	0.02	0.02
SAVINGS	= 1 if had savings	0.22	0.19	0.22	0.19	0.23	0.20
Farm´s characteristics							
LAND	**Land size, in hectares**	**4.68**	**3.97**	**3.48**	**3.45**	**2.98**	**2.93**
YEARS	years in the farm activity	25.75	25.04	25.76	25.02	25.75	25.11
CROPS	number of crops	4.71	4.09	4.71	4.09	4.69	4.08
SOIL	= 1 if soil affected by salinity	0.68	0.64	0.68	0.65	0.68	0.64
PLAGUE	= 1 If soil affected by plague	0.10	0.11	0.10	0.11	0.10	0.11
Environmental Conditions							
ALT	Altitude in 1000 mts	3.23	3.24	3.23	3.23	3.22	3.23
DISTANCE	distance to downtown, in hours	1.68	1.67	1.67	1.69	1.66	1.66
Number of observations		**4023**	**4023**	**4019**	**4019**	**4000**	**4000**