TECHNICAL EFFICIENCY ANALYSIS CORRECTING FOR BIASES FROM OBSERVED AND UNOBSERVED VARIABLES: AN APPLICATION TO A NATURAL RESOURCE MANAGEMENT PROJECT

Boris E. Bravo-Ureta*, William Greene and Daniel Solís*****

ABSTRACT

This paper brings together the stochastic frontier framework with impact evaluation methodology to compare technical efficiency (TE) across treatment and control groups using cross sectional data associated with the MARENA Program in Honduras. A matched group of beneficiaries and control farmers is determined using Propensity Score Matching techniques to mitigate biases stemming from observed variables. In addition, possible self-selection arising from unobserved variables is addressed using a selectivity correction model for stochastic frontiers recently introduced by Greene (2010). The results reveal that average TE is consistently higher for beneficiary farmers than the control group while the presence of selectivity bias cannot be rejected. TE ranges from 0.67 to 0.75 for beneficiaries and from 0.40 to 0.65 for the control depending on whether biases were controlled or not. The TE gap between beneficiaries and control farmers decreases by implementing the matching technique and the sample selection framework decreases this gap even further. The analysis also suggests that beneficiaries do not only exhibit higher TE but also higher frontier output.

Keywords: Stochastic frontiers, Technical efficiency, Propensity Score Matching, Sample selection, Honduras

JEL Classification: D24, Q2, Q12, Q16

* Department of Agricultural and Resource Economics, University of Connecticut, US, and Department of Agricultural Economics, University of Talca, Chile. email: boris.bravoureta@uconn.edu

** Stern School of Business, New York University, US.

*** Division of Marine Affairs and Policy, Rosenstiel School of Marine and Atmospheric Science, University of Miami, US.

TECHNICAL EFFICIENCY ANALYSIS CORRECTING FOR BIASES FROM OBSERVED AND UNOBSERVED VARIABLES: AN APPLICATION TO A NATURAL RESOURCE MANAGEMENT PROJECT

1 Introduction

Measuring the impact of development projects has gained considerable prominence in the literature as donors are increasingly interested on quantitative evidence of the effects that development assistance has on the lives of poor people (Ravallion 2008). Calls for additional funding to meet the Millennium Development Goals need to be supported by convincing analysis showing that current spending is indeed contributing toward the attainment of such goals (Khandker et al. 2010; World Bank 2006; Pearson 2009; Radelet 2004).

There is growing evidence and adherence to the use of randomized experiments to undertake impact evaluation (Duflo et al. 2008). However, it is often the case in development projects that experimental designs are costly and difficult to implement and one needs to rely on quasi-experimental methods. A core issue in impact evaluation is that to isolate a project's impact one would ideally calculate the difference between the outcome for project beneficiaries and the outcomes from this same group had they not been part of the project. Clearly, both states of nature are not possible and such missing data are referred to as the 'counterfactual' in the impact evaluation literature (World Bank 2006). Thus, the counterfactual situation is what would have happened to beneficiaries had they not participated in the intervention.

A common evaluation technique is the before *versus* after approach where mean outcomes for the treatment group are compared before and after the intervention. This approach, usually referred to as a reflexive evaluation, yields information on the trend for the treatment group but does not allow for the attribution of the observed changes to the intervention since other external factors can be responsible, at least in part, for such changes. A better approach is to select a comparison group like the treatment group in every way, except that it was not subject to the intervention. This can be done using Propensity Score Matching (PSM), which is a well established practice to account for biases stemming from observed variables (World Bank 2006).

Another common issue that is prevalent in many development projects is that very often beneficiaries self-select into participation. If self-selection is based on unobserved variables (e.g., managerial ability) and panel data are available, then fixed effects estimators along with PSM can be used to deal with the problem provided that the unobserved variables are time invariant (Angrist and Pischke 2009). Thus, the generation of a counterfactual along with the mitigation of biases from observed and unobserved variables can be addressed in quasi-experimental designs provided one has samples for both treatment and control groups for the baseline and then a subsequent measure at the end or near the end of the project for the indicators of interest. Under these circumstances, PSM along with fixed effects makes it possible to derive suitable impact measures (e.g., Bravo-Ureta et al. 2011; Rodriguez et al. 2007).

The issue we want to tackle in this article is the comparison of technical efficiency (TE) across treatment and control groups using cross sectional data collected at the end of the implementation of the project. This type of data configuration is not uncommon in evaluation work in developing country projects. Thus, we first need to establish a group of beneficiaries and a control group that should have been very similar at the baseline, according to a vector of time invariant observable attributes. In addition, we need to address possible self-selection in the context of a stochastic production frontier (SPF) model, an issue we deal with by using the model recently introduced by Greene (2010). A contribution of this paper is to narrow the gap between significant methodological advances that have been made in the estimation of SPF models (Greene 2008), which has lead to a sizable number of farm level TE studies (Bravo-Ureta et al. 2007), and the rapidly evolving impact evaluation literature (Khandker et al. 2010).

The analysis relies on data available from the MARENA Program which was implemented in Honduras between 2004 and 2009 with funding from the Inter-American Development Bank. MARENA belongs to a set of development efforts that have been implemented by national governments in Central America during the last two decades focusing on improving environmental conditions, increasing agricultural productivity and reducing poverty among peasant farmers. The intent of MARENA was to increase productivity and alleviate poverty by strengthening natural resource management, at both local and regional levels, in an area of influence covering 13,721 km^2 and about 930,000 inhabitants (Bravo-Ureta 2009).

MARENA was organized into three components: Component I, addressed institutional, strategic, regulatory, and management capacity needs of key public agencies; Component II, financed investments in priority sub-basins through three complementary modules; and Component III dealt with the overall coordination of the program. The data used in this study relates to Module 3 within Component II, which promoted productivity growth by providing managerial training to beneficiaries and by fostering investments in sustainable agricultural production systems with a budget of US $7.6 million for that purpose (Bravo-Ureta 2009). An important feature of MARENA, as is the case with many programs of this type, is that once the beneficiary eligibility criteria are set and the Program is promoted throughout the intervention area, farmers decide whether or not to participate. Thus, self-selection plays an important role in any analysis that relies on data collected to evaluate the impact of these programs.

The remainder of the paper is organized into four additional sections. The second section presents a brief overview of related literature followed by the conceptual framework used. The next section presents the data used followed by the empirical model and the results. The paper ends with some concluding remarks.

2 Related literature

As already indicated, this paper seeks to narrow the gap between two large bodies of economic literature, SPF modeling on the one side and impact evaluation on the other. To our knowledge, the only paper that has made an explicit attempt in this direction is by Dinar et al. (2007) where an SPF model is used to evaluate the impact of agricultural extension on the performance of farmers in Crete. A major shortcoming of the Dinar et al. (2007) paper is that selectivity bias is not addressed; therefore, the reported SPF parameter estimates and associated TE scores are likely to be biased.

The combination of efficiency estimation and sample selection appears in a few studies which have generally dealt with selectivity bias by relying on the Heckman approach, a procedure that is unsuitable for nonlinear models such as the SPF (Greene 2010). Bradford et al. (2001) studied patient specific costs for cardiac revascularization in a large hospital. According to these authors, "… the patients in this sample were not randomly assigned to each treatment group. Statistically, this implies that the data are subject to sample selection bias. Therefore, we utilize a standard Heckman two-stage sample-selection process, creating an IMR [Inverse Mills Ratio] from a first-stage Probit estimator … and … this variable is included in the frontier estimate…" (p. 306).

Sipiläinen and Oude Lansink (2005) utilized a translog stochastic distance frontier model to analyze TE for organic and conventional farms. These authors state that "[p]ossible selection bias between organic and conventional production can be taken into account [by] applying Heckman's (1979) two step procedure" (p. 169.). In this case, the inefficiency component is distributed as the truncation at zero with a heterogeneous mean[1] and the IMR is added to the deterministic part of the frontier function.

Solís et al. (2007) analyze TE levels for hillside farmers under two different levels of adoption of soil conservation in El Salvador and Honduras applying the Switching Regression Approach (SRA) to a SPF. The authors examine potential selectivity bias for high and low level adopters, and separate stochastic production frontiers, corrected for selectivity bias, were estimated for each group. SRA also relies on the introduction of the IMR into the specification of the frontier.

Other authors have acknowledged the sample selection issue in stochastic frontier studies. Kaparakis et al. (1994), in an analysis of commercial banks, and Collins and Harris (2005), in their study of UK chemical plants, suggest that sample selection was a potential issue in their analysis. However, neither study modified the stochastic frontier models to address this issue.

Mayen et al. (2010) used an alternative approach to address self-selection into organic farming by using PSM to compare organic farms to otherwise similar conventional farms. The authors found small differences in TE between organic and conventional farms when TE is measured against the appropriate technology. Although this study corrected for biases from observed variables the authors did not account for biases stemming from unobserved factors.

In a recent paper, Rahman et al. (2009) used the methods described below and applied in this paper to analyze production efficiency for a sample of rice producers in Thailand. The authors analyzed the switch from lower quality rice varieties to Jasmine rice which is a higher quality product. Their sample included 207 farmers with lower quality rice and 141 in the other

[1] See Battese and Coelli (1995).

group. Their results indicate that the correction for adoption of the higher quality variety produced marked differences in the estimated production frontier and a highly significant ‘selection effect’. However, Rahman et al. (2009) did not use any matching techniques to ensure that the control and treated groups had similar observed characteristics.

Two other approaches have been introduced recently to modeling sample selection in the stochastic frontier model. Kumbhakar et al. (2009) developed a model where the selection mechanism is assumed to operate through the one-sided error in the frontier and they apply this model to examine organic versus conventional dairy farming in Finland. The other paper is by Lai et al. (2009) who formulate a wage equation in which the selection mechanism is correlated, through a copula function, with the composed error in the frontier instead of being correlated specifically with either the two sided or the one sided terms. In both the Kumbhakar et al. (2009) and Lai et al. (2009) papers the log likelihood is substantially more computationally demanding than the one used here. More importantly, the difference in the assumption of the impact of the selection effect is substantive.

Consequently, the current study adds to the literature by implementing an empirical framework which corrects for biases arising from both observed and unobserved variables, and applies this method to the impact evaluation literature.

3 Conceptual framework

To evaluate the impact of MARENA on the TE levels of beneficiaries we implement a multi-step framework where we first generate a group of comparable control farmers and then account for potential self-selection in the estimation of an SPF model. PSM is commonly used when quasi-experimental data are available, as is the case here, to generate a control group with observed characteristics that are as similar as possible as those for the treated group, a condition that is necessary to get an accurate measure of impact (Monteiro 2010). In other words, it is an approach that can be used to create the counterfactual situation while mitigating potential biases associated with observed characteristics (Rosenbaum and Rubin 1983). A binary choice model is used to generate a 'score' which is equal to the probability of receiving treatment, considering both treated and nontreated (control) groups based on a given set of predetermined covariates (Cameron and Trivedi 2005; Becker and Ichino 2002; Imbens and Wooldridge 2008). Several recent studies have applied PSM within the impact evaluation literature (e.g., Bravo-Ureta et al. 2011; Cerdán-Infantes et al. 2008; Cavatassi et al. 2009).

To deal with biases from unobserved variables (e.g., managerial ability) within an SPF formulation we use the model recently introduced by Greene (2010). This model assumes that the unobserved characteristics in the selection equation are correlated with the noise in the stochastic frontier model; hence, Greene's contribution can be seen as a significant improvement of Heckman's self selection specification for the linear regression model. The sample selection and SPF models, along with their error structures, can be expressed as:[2]

$$
\begin{array}{ll}
\text{Sample Selection:} & d_i = 1[\boldsymbol{\alpha}'\mathbf{z}_i + w_i > 0],\ w_i \sim N[0,1] \qquad (1)\\
\text{SPF:} & y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \varepsilon_i \sim \mathrm{N}[0,\sigma_\varepsilon^2]\\
 & (y_i,\mathbf{x}_i) \text{ observed only when } d_i = 1.\\
 & \\
\text{Error Structure:} & \varepsilon_i = v_i - u_i\\
 & u_i = |\sigma_u U_i| = \sigma_u |U_i| \text{ where } U_i \sim N[0,1]\\
 & v_i = \sigma_v V_i \text{ where } V_i \sim \mathrm{N}[0,1].\\
 & (w_i,v_i) \sim N_2[(0,1), (1, \rho\sigma_v, \sigma_v^2)]
\end{array}
$$

In the set of equations above, d is a binary variable equal to one for beneficiaries and zero for control, y is output, z is a vector of covariates included in the sample selection equation, and x is a vector of inputs in the production frontier. The Greek characters α and β are parameters to

[2] For details on model specification see Greene (2010).

be estimated while the characters in the error structure correspond to the typical characterization of a stochastic frontier model. It is useful to underscore that the parameter ρ captures the presence or absence of selectivity bias.

The log likelihood for the model in (1) is formed by integrating out the unobserved $|U_i|$ and then maximizing with respect to the unknown parameters. Thus,

$$\log L(\boldsymbol{\beta},\sigma_u,\sigma_v,\boldsymbol{\alpha},\rho) = \sum_{i=1}^{N}\log\int_{|U_i|} f(y_i \mid \mathbf{x}_i,\mathbf{z}_i,d_i,|U_i|)\,p(|U_i|)\,d\,|U_i|. \qquad (2)$$

The integral in (2) is not known and must be approximated. To simplify the estimation, Greene (2010) uses a two step approach. The single equation MLE of $\boldsymbol{\alpha}$ in the Probit equation in (1) is consistent but inefficient. However, for the estimation of the parameters of the SPF it is not necessary to reestimate $\boldsymbol{\alpha}$ and the estimates of $\boldsymbol{\alpha}$ are taken as given in the simulated log likelihood. The Murphy and Topel (2002) correction is used to adjust the standard errors in essentially the same fashion as Heckman's correction of the canonical selection model.

Greene (2010) goes on to argue that the non-selected observations (i.e., when $d_i = 0$) do not contribute information about the parameters to the simulated log likelihood and thus the function to be maximized becomes:

$$\log L_{S,C}(\boldsymbol{\beta},\sigma_u,\sigma_v,\rho) = \sum_{d_i=1}\log\frac{1}{R}\sum_{r=1}^{R}\left[\begin{array}{c}\dfrac{\exp\left(-\frac{1}{2}(y_i-\boldsymbol{\beta}'\mathbf{x}_i+\sigma_u|U_{ir}|)^2/\sigma_v^2)\right)}{\sigma_v\sqrt{2\pi}}\times \\ \Phi\left(\dfrac{\rho(y_i-\boldsymbol{\beta}'\mathbf{x}_i+\sigma_u|U_{ir}|)/\sigma_\varepsilon+a_i}{\sqrt{1-\rho^2}}\right)\end{array}\right] \qquad (3)$$

where $a_i = \hat{\alpha}' z_i$.

The parameters of the model are estimated using a conventional gradient based approach, the BFGS method, and use the BHHH estimator to obtain the asymptotic standard errors. The maximand reduces to that of the maximum simulated likelihood estimator of the basic frontier model when ρ equals zero. This provides us with a method of testing the specification of the selectivity model against the simpler model using a (simulated) likelihood ratio test.

The end objective of the estimation process is to characterize the inefficiency in the sample, u_i, or the efficiency, $\exp(-u_i)$. Aggregate summary measures, such as the sample mean and variance are often provided (e.g., Bradford, et al. (2001) for hospital costs). Researchers also compute individual specific estimates of the conditional means based on the Jondrow et al. (1982) (JLMS) result given by

$$E[u_i \mid \varepsilon_i] = \frac{\sigma\lambda}{1+\lambda^2}\left[\mu_i + \frac{\phi(\mu_i)}{\Phi(\mu_i)}\right],\ \mu_i = \frac{-\lambda\varepsilon_i}{\sigma},\ \varepsilon_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i. \tag{4}$$

In the standard approach this function is computed using the maximum likelihood estimates. In principle, we could repeat this computation with the maximum simulated likelihood estimates. However, the alternative approach used here takes advantage of the simulation of the values of u_i during estimation. It should be noted that the approach gives a strikingly similar answer to the JLMS plug in result (see Greene (2010) Section 2.4 for details).

4 Data and empirical model

In this study we combine PSM, to correct for biases from observed characteristics, with the Greene (2010) model to correct for selectivity bias arising from unobserved variables and then measure and compare TE scores resulting from various combinations of these correction procedures. We use cross sectional data collected for a total of 371 farm households located in MARENA’s general area of intervention for the agricultural year 2007-08. Of this total, 109 are beneficiaries of the Program and the other 262 are non-beneficiaries. This last set provides the basis for constructing the control group. The beneficiaries were randomly selected from a comprehensive list of farmers participating in the program while non-beneficiaries were randomly selected from a list of farmers living either in intervened villages, but that were not part of the project, or in intervened municipalities but from villages not participating in MARENA. More details on the sampling procedure can be found in ESA (2008) and in Bravo-Ureta et al (2009).

As indicated earlier, to accurately measure the impact of a project such as MARENA it is necessary to obtain a counterfactual group of farmers who display time-invariant characteristics that are similar to those associated with the project. The PSM technique is often used to generate such group (Cameron and Trivedi 2005). The implementation of PSM first requires the estimation of the probability that a farmer in the sample will become associated with the project. Then, control and beneficiary groups are generated which can be done using various criteria. Here we utilize the '1-to-1 nearest neighbor without replacement' criterion where every beneficiary is matched with a non-beneficiary farmer imposing the common support condition (Sianesi 2001). Although PSM does not completely eliminate biases that might stem from observed characteristics across the treated and non-treated groups, Imbens and Wooldridge (2008) argue that this method generally yields reasonable results.

We should point out that several alternative matching criteria have been developed and applied in the literature (Cameron and Trivedi 2005). The decision to choose the '1-to-1 nearest neighbor without replacement' criterion is based on the fact that it is easy to implement (Caliendo and Kopeinig 2008; Rosembaum and Ruben 1985) and has the most intuitive interpretation of all the alternatives available. In addition, this criterion has become a popular choice in applied economic analysis published recently (e.g., Bravo-Ureta et al 2011; Dillon 2011; Rodriguez et al 2007). Moreover, as discussed below, the evidence shows that a suitable match between control and beneficiaries is achieved, and that such match is found for all beneficiaries.

The matching procedure yielded a total of 109 pairs (i.e., all 109 beneficiaries were paired with 109 non-beneficiaries out of 262). Following Leuven and Sianesi (2003), *t-tests* were conducted before and after matching to evaluate the null hypotheses that the means of observed characteristics of beneficiaries and non-beneficiaries are equal. The results show that the mean of most of the observed characteristics are not statistically different suggesting that the balancing property of the covariates is satisfied (Leuven and Sianesi 2003). Table 1 defines all the variables included in the empirical analysis, while Table 2 presents the descriptive statistics for the unmatched (N=371) and matched samples (N=218) and the results of the tests.

Once the matched samples are constructed, we estimate the SPF model with correction for sample selection. In doing so, it is necessary to model the decision of the ith farmer to participate in MARENA or not. This behavior can be described by a criterion function, which is postulated to be associated with exogenous household socioeconomic variables as follows:

$$B_i = \alpha_0 + \sum_{j=1}^{7} \alpha_j Z_{ji} + w_i \qquad (5)$$

where B is a dichotomous variable reflecting the farmer's decision to participate in the project (i.e., 1 for beneficiary and 0 otherwise); Z is a vector exogenous variables (i.e., linear and quadratic terms for age and education, family size, total farm land and the possession of legal title on the land); α are the unknown parameters; and w is the disturbance term distributed as $N(0,\sigma^2)$.

Then, the production frontier for beneficiaries is estimated using a Translog (TL) specification as follows: [3]

$$\ln(Y_i) = \beta_0 + \sum_{j=1}^{4} \beta_j lnX_{ji} + \frac{1}{2}\sum_{j=1}^{4}\sum_{k=1}^{4} \beta_{jk} lnX_{jki} + v_i - u_i \quad \text{iff } B=1 \qquad (6)$$

[3] Preliminary comparisons led to the rejection of the Cobb-Douglas functional form.The same framework is used to estimate the SPF model with correction for Sample Selection for the control group. In this case the dependent variable in equation (5) equals 1 for non-beneficiaries and 0 for beneficiaries.

where Y_i represents output, X are inputs, β are the unknown parameters, and v and u are the elements of the composed error term, ε. The dependent variable for the SPF model is the total value of agricultural production (TVAP) measured in Lempiras (US $1 = NL19.3). The explanatory variables include: production expenditures on purchased inputs; value of hired and family labor; and cultivated area. To account for environmental conditions the altitude at which the farm is located is also included. A different set of variables is included in the Probit and SPF models to satisfy the identification criterion stated by Maddala (1983).

5 Results and analysis

Before turning to the results we find it convenient to summarize the steps that are implemented sequentially to arrive at the TE measures that come from models that have been corrected for biases from both observables and unobservables. These steps are:

1. All available data are used to estimate a pooled unmatched SPF model (P-U) where the binary variable BENEF (0 for control, 1 for beneficiaries) is included as a regressor to account for Program participation. Thus, this model ignores any type of biases.

2. Two separate SPF models are estimated with unmatched data, one for beneficiaries (B-U) and the other for control (C-U) farmers again ignoring any biases.

3. Two separate SPF models are reestimated with correction for selectivity bias based on Greene (2010), one for beneficiaries (B-U-S) and the other for control (C-U-S) farmers.

4. All available data are used to implement the PSM which provides the basis for correcting for biases from observed characteristics by matching beneficiaries and control farmers.

5. The pooled SPF model is reestimated but using only the matched subgroups and the BENEF dummy variable (P-M) is included as a regressor to account for Program participation.

6. Two separate SPF models are estimated using the matched subsamples, one for beneficiaries (B-M) and the second for the control (C-M) group without correction for selectivity bias. Thus these models correct only for biases from observables.

7. Finally, two separate SPFs are estimated using the matched subsamples, one for beneficiaries (B-M-S) and the other for the control (C-M-S) group, correcting for

selectivity bias. Thus, the models in this step incorporate corrections for both biases (from observed and unobserved variables).

Tables 3 and 4 present the maximum likelihood estimates for the whole unmatched and matched samples, respectively. Following common practice, all variables in the TL models were normalized by their geometric mean (GM). Thus, the first-order coefficients can be interpreted as partial production elasticities at the GM. As expected, all estimated models present positive partial production elasticities; however, their magnitudes and statistical significance differ across models. Consistently, cultivated land (ALAND) and purchased inputs (EXPENSE) contribute the most to farm production. This result is consistent with Kalirajan (1991) who argues that budget restrictions and land availability are the main production constrains for small scale farmers in developing societies. The sum of all partial production elasticities is consistently less than 1 revealing decreasing returns to scale in all models, a result that is consistent with previous research on small scale farmers in less favorable areas (e.g., González and López 2007; Solís et al. 2009; Chavas et al. 2005).

The values for the σ^2 and γ parameters are also reported at the end of Tables 3 and 4. The null hypothesis that $\gamma = 0$ is rejected in all cases which suggests that technical inefficiency (TI) is indeed stochastic and that inefficiency is an important contributor to observed output variability.

The main goal of this study is to measure potential efficiency differences among beneficiary and control farmers and the effect of controlling for biases from observed and unobserved variables. First, the pooled models (P-U and P-M) suggest that there are no significant differences between the two studied groups of farmers based on the lack of statistical significance of the parameter for BENEF. These results are however dismissed by a likelihood ratio test (LR) that offers evidence favoring the estimation of separate technologies for beneficiaries and control farmers. Specifically, the estimated LR test is:

$$LR = 2*(lnL_P - (lnL_B + lnL_C)) \qquad (7)$$

where lnL_P, lnL_B and lnL_C represent the log-likelihood function obtain from the pooled model (P-U or P-M), and the Beneficiary (B-U or B-M) and Control (C-U or C-M) subsamples (restricted models), respectively (Greene 2007). The estimated LR test rejects, in both cases (unmatched and matched samples), the null hypothesis for equality, confirming that the parameters for the production frontiers differ across the two groups of farmers.

The impact of correcting for self-selection is analyzed next. Table 5 shows the empirical estimates for the self-selection Probit model for both the unmatched and the matched samples. The results of these two Probit models are, in general, compatible. Specifically, the null hypothesis that all coefficients are simultaneously zero is rejected in both cases and the estimated coefficients exhibit comparable values. However, the number of statistically significant variables is lower for the matched sample. This reduction of the statistical significance of the parameters can be explained by the fact that PSM reduces the variability between the two samples (beneficiary and control farmers) which affects the significance of the estimates.[4] The empirical results suggest that both age and education display nonlinear effects on the choice to become a beneficiary of MARENA. Specifically, age and education display, correspondingly, a U and an inverted-U shape relationship with respect to the propensity to be a beneficiary farmer. On the other hand, the estimates for total land and legal ownership of the land display non significant effects.

The estimation of the sample selection SPF models reveals that the coefficient for the selectivity variable $RHO_{(w,v)}$ is statistically different from zero for the Beneficiary group using both samples, Unmatched and Matched, and for the Control group using the Matched sample.

[4] Following Caliendo and Kopeinig (2008), we used a set of t-tests to examine if, at the mean of the sample, both control and treated farmers display similar observed characteristics after matching. The results of these tests are reported in Table 2.

This result suggests the presence of selection bias, thus lending support to the use of a sample selection framework to estimate separate SPFs for the beneficiaries and control groups. The presence of selection bias also indicates that the estimates from the conventional SPF model yield biased frontier estimates which affect the TE scores. It is important to indicate that Rahman et al. (2009) also found selection bias among rice farmers in Thailand. However, selection bias was not an issue for Greene (2010).

Table 6 presents average TE for all estimated models using the conventional and sample selection SPF models for both the unmatched and matched samples. In addition, the table presents the differential, in percentage terms, between the TE for beneficiaries and control groups. Beneficiaries present an average TE ranging from 67% (B-U) to 75% (B-U-S). By contrast, the average TE for control farmers ranges from 40% (C-U) to 66% (C-U-S). Moreover, beneficiaries exhibit a higher average TE in all cases. These results clearly show that the efficiency gap between beneficiaries and control farmers decreases by implementing the matching technique, which is consistent with findings reported by Mayen et al. (2010). This outcome is expected since PSM makes both samples comparable. In addition, the sample selection correction decreases the TE gap even further.

To get a better understanding of how correcting for biases from observed and unobserved variables affects TE levels, Figure 1 presents the distribution of TE scores from the set of models with and without correction for both type of biases (extreme results); namely, the B-U and C-U SPFs (unmatched traditional SPF), and B-M-S and C-M-S SPFs (matched with Sample Selection). The results exhibit significant differences between these two sets of models. For instance, 7% of the beneficiaries have a TE level that is 81% or higher using the traditional SPF method and the unmatched sample (B-U); however, this percentage increases to 10% of beneficiaries with the B-M-S model. The effect of controlling for these biases is more significant for the control group. Specifically, Figure 1 shows that 34% of the control farmers operate at an efficiency level below 50% when using the C-U model, while not a single observation is found at this level once the biased correction is implemented (C-M-S model).

Finally, we investigate which of the two groups (control *versus* beneficiaries) has higher output after controlling for biases from observed and unobserved variables. For this purpose, we compare the predicted frontier output at three different input levels: 1) at the average for the smallest matched pair of farms; 2) at the average for the entire sample; and 3) at the average for the largest matched pair. As shown in Table 7, the total output gap is 16%, 14% and 9%, respectively in favor of beneficiaries. Thus, the analysis suggests that beneficiaries do not only exhibit higher efficiency but also higher total output.

6 Concluding remarks

This paper compares technical efficiency (TE) across treatment and control groups using cross sectional data associated with the MARENA Program in Honduras. A matched group of beneficiaries and control farmers is generated using Propensity Score Matching (PSM) techniques to mitigate biases associated with observed variables. In addition, we deal with possible self-selection arising from unobserved variables using a selectivity correction model for stochastic frontiers recently introduced by Greene (2010).

The results do reveal that average TE is consistently higher for beneficiary farmers than the control group while the presence of selectivity bias cannot be rejected. TE ranged from 0.67 to 0.75 for beneficiaries and from 0.40 to 0.65 for the control. It is worth noting that the TE gap between beneficiaries and control farmers decreases after the samples are matched. This result is expected since PSM makes both samples comparable on observables. In addition, the sample selection framework decreases this gap even further. Our empirical results also suggest that the frontier for beneficiary farmers is located above the one for the control group. These differences highlight the value of exploring extensions to the model used in this study to allow for more comprehensive analyses of the impact of development projects. Thus, an extended methodological framework that accommodates panel data would make it possible to decompose the impact of development projects on productivity growth by separating the effects of technological change, technical efficiency change as well as changes in scale or size.

Acknowledgements: The authors would like to acknowledge the helpful comments received from an anonymous reviewer and the Editor. We also are very thankful for the support provided by Ricardo Quiroga (IADB), Aarón Inestroza and the MARENA Team in Honduras during the original work. We also thank Alexandre Nunes Almeida for his suggestions and review comments on previous drafts.

Table 1 Definition of variables

Variable	Parameter	Unit	Definition
TVAP	Y	Lempiras†	Total value of agricultural production
EXPENSE	β_1	Lempiras	Farm production expenditures
LABOR	β_2	Lempiras	Total expenditure on hired and family labor
ALAND	β_3	Hectares	Total land devoted to agricultural production
ALTITUDE	β_{ALT}	100 Meters	Altitude at which the farm is located
BENEF	β_{BEN}	Dummy	1 if the household is a beneficiary of MARENA
AGE	α_1	Years	Age of the household head
EDUC	α_2	Years	Level of education of the household head
FAMILY	α_3	Number	Number of people in the household
TLAND	α_4	Hectares	Total farm land
TITLE	α_5	Dummy	1 if the household has legal title to at least some of the land farmed

† US $1=19.3 Lempiras (Lps)

Table 2 Descriptive statistics

Variable	POOLED		BENEFICIARIES		CONTROL		Test of means†
	Mean	SD	Mean	SD	Mean	SD	
UNMATCHED SAMPLE							
TVAP	46,729.6	74,025.1	43,006.7	84,563.0	48,278.4	69,289.1	0.624
EXPENSE	20,825.8	47,778.1	13,545.7	28,663.4	23,854.6	53,523.5	1.899*
LABOR	36,627.1	35,557.7	38,286.4	43,490.5	35,936.8	31,750.3	0.579
ALAND	2.24	2.42	1.86	1.61	2.40	2.67	1.967**
ALTITUDE	9.46	3.96	9.17	3.28	9.57	4.20	0.887
AGE	50.69	13.47	48.95	14.34	51.72	13.04	1.809*
EDUC	3.38	3.08	3.57	2.80	3.30	3.20	0.767
FAMILY	5.67	2.40	6.06	2.50	5.50	2.35	2.051**
TLAND	7.84	26.43	5.95	19.25	8.62	28.89	0.886
TITLE	0.82	0.39	0.80	0.40	0.83	0.38	0.682
Observations	***371***		***109***		***262***		
MATCHED SAMPLE							
TVAP	39,708.6	67,846.8	43,006.7	84,563.0	36,440.5	45,819.3	0.712
EXPENSE	15,864.5	32,001.1	13,545.7	28,663.4	18,162.2	34,976.9	1.066
LABOR	37,052.1	37,351.5	38,286.4	43,490.5	35,829.0	30,218.6	0.484
ALAND	2.03	1.79	1.86	1.61	2.20	1.94	1.408
ALTITUDE	9.16	3.70	9.18	3.28	9.14	4.09	0.079
AGE	48.73	12.96	48.95	14.34	48.50	11.50	0.255
EDUC	3.49	2.81	3.57	2.80	3.42	2.83	0.393
FAMILY	6.11	2.46	6.06	2.50	6.15	2.44	0.269
TLAND	5.10	14.21	5.95	19.25	4.26	5.94	0.876
TITLE	0.82	0.39	0.80	0.40	0.84	0.37	0.766
Observations	**218**		**109**		**109**		

† A t-test is use to determine if the sample means are significantly different between the beneficiaries and control groups.
*, P < 0.10; **, P< 0.05; ***, P< 0.01.

Table 3 Parameter estimates for the conventional and sample selection SPF models: Unmatched sample

Variables	Conventional SPF						Sample Selection SPF			
	P-U		B-U		C-U		B-U-S		C-U-S	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
CONSTANT	10.245***	0.142	10.252***	0.183	9.672***	0.224	10.199***	0.824	9.738***	0.714
β_1	0.413***	0.046	0.492***	0.078	0.367***	0.056	0 .396***	0.082	0.491***	0.127
β_2	0.115***	0.036	0.041	0.051	0.133***	0.046	0 .158**	0.066	0.044	0.076
β_3	0.229***	0.085	0.220	0.137	0.252**	0.106	0.195	0.148	0.222	0.193
β_{11}	0.074***	0.015	0.105***	0.025	0.062***	0.018	0 .070***	0.026	0.105***	0.037
β_{22}	0.027***	0.010	0.011	0.014	0.031**	0.014	0 .035**	0.018	0.013	0.021
β_{33}	0.188	0.123	0.353*	0.194	0.149	0.158	0.036	0.213	0.370*	0.204
β_{12}	0.007	0.006	0.041*	0.022	0.006	0.007	0.003	0.010	0.044	0.049
β_{13}	-0.038	0.034	-.141**	0.060	0.002	0.042	0.006	0.060	-0.147*	0.076
β_{23}	-0.026	0.021	-0.051	0.034	-0.040	0.027	-0.026	0.046	-0.056	0.078
β_{ATL}	0.020	0.013	0.044**	0.021	0.016	0.017	0.011	0.021	0.048**	0.023
β_{BEN}	0.027	0.108	--		--				--	
RTS	0.757		0.753		0.752		0.749		0.757	
L. Likelihood	-506.778		-117.186		-375.348		-472.951		-245.592	
γ	1.988***	0.186	1.186***	0.265	2.231***	0.259	--		--	
σ^2	1.385***	0.003	0.897***	0.007	1.525***	0.005	--		--	
$\sigma_{(u)}$	--		--		--		0.465	1.006	0.510	0.500
$\sigma_{(v)}$	--		--		--		1.260***	0.189	0.655***	0.177
RHO $_{(w,v)}$	--		--		--		-0.851***	0.071	-0.234	0.639
N	***371***		***109***		***262***		***109***		***262***	

*, P < 0.10; **, P< 0.05; ***, P< 0.01.

Table 4 Parameter estimates of the conventional and sample selection SPF models: Matched sample

Variables	Conventional SPF						Sample Selection SPF			
	P-M		B-M		C-M		B-M-S		C-M-S	
	Coeff.	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**
CONSTANT	9.923***	0.155	10.252***	0.183	9.761***	0.224	10.552***	0.338	9.236***	0.318
β_1	0.428***	0.049	0.492***	0.078	0.317***	0.056	0.332***	0.075	0.479***	0.100
β_2	0.078**	0.037	0.041	0.051	0.063	0.058	0.090	0.076	0.037	0.062
β_3	0.288***	0.089	0.220	0.137	0.457***	0.101	0.421***	0.148	0.226	0.165
β_{11}	0.101***	0.018	0.105***	0.025	0.068***	0.022	0.068**	0.029	0.106***	0.032
β_{22}	0.017	0.011	0.011	0.014	0.009	0.015	0.013	0.020	0.011	0.017
β_{33}	0.202	0.142	0.353*	0.194	-0.122	0.187	-0.101	0.239	0.408**	0.199
β_{12}	0.015*	0.009	0.041*	0.022	-0.003	0.010	-0.002	0.018	0.034	0.034
β_{13}	-0.073*	0.043	-.141**	0.060	0.048	0.054	0.048	0.073	-0.149**	0.065
β_{23}	-0.023	0.022	-0.051	0.034	0.009	0.027	0.010	0.044	-0.044	0.066
β_{ATL}	0.014	0.014	0.044**	0.021	0.009	0.020	0.006	0.025	0.048**	0.021
β_{BEN}	0.130	0.102	--		--		--		--	
RTS	0.794		0.753		0.837		0.843		0.742	
L. Likelihood	-247.445		-117.186		-118.056		-191.335		-190.998	
γ	1.587***	0.226	1.186***	0.265	3.404***	1.048	--		--	
σ^2	1.025***	0.003	0.897***	0.007	1.157***	0.008	--		--	
$\sigma_{(u)}$	--		--		--		0.817*	0.484	0.921***	0.155
$\sigma_{(v)}$	--		--		--		0.783*	0.424	0.712***	0.117
RHO $_{(w,v)}$	--		--		--		-0.926***	0.126	0.965***	0.139
N	***218***		***109***		***109***		***109***		***109***	

*, P < 0.10; **, P< 0.05; ***, P< 0.01.

Table 5 Parameter estimates of the probit selection equation

Parameter	Unmatched Sample		Matched Sample	
	Coeff.	**S.E.**	**Coeff.**	**S.E.**
CONSTANT	0.880	0.8515	2.208**	1.1081
α_1	-0.077**	0.0323	-0.100**	0.0430
α_{11}	0.001**	0.0003	0.001**	0.0004
α_2	0.148**	0.0651	0.047	0.0825
α_{22}	-0.014**	0.0064	-0.003	0.0085
α_3	0.071**	0.0304	0.011	0.0366
α_4	-0.003	0.0031	0.007	0.0077
α_5	-0.138	0.1829	-0.171	0.2230
L. Likelihood	*-216.08*		*-147.63*	
Chi-Square	*17.12**		*18.78***	
N	*371*		*218*	

*, P < 0.10; **, P< 0.05; ***, P< 0.01.

Table 6 Technical Efficiency (TE) levels and differentials across models

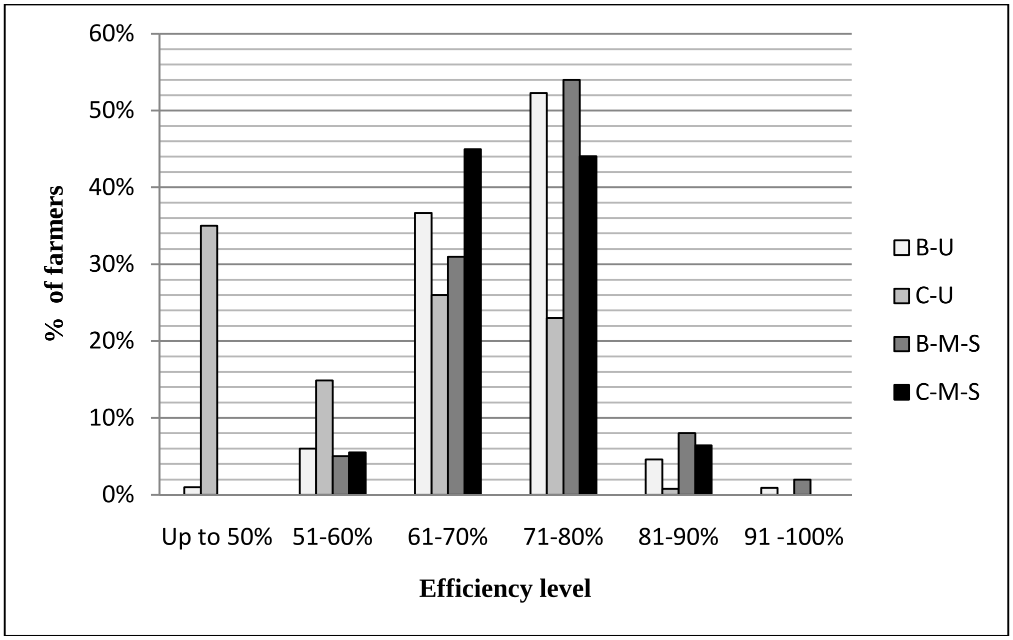
Index	CONVENTIONAL SPF							SAMPLE SELECTION SPF				
	P-U		B-U		C-U		Test of Means†	B-U-S		C-U-S		Test of Means
	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
TE	0.43	0.16	0.67	0.15	0.40	0.17	15.79***	0.70	0.05	0.59	0.08	13.27***
Differential ‡				*67.5%*						*19%*		
	P-M		B-M		C-M		Test of Means	B-M-S		C-M-S		Test of Means
	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
TE	0.53	0.16	0.67	0.15	0.48	0.21	7.69***	0.75	0.18	0.65	0.11	4.95***
Differential				*39%*						*15%*		

† A t-test is use to determine if TE means are significantly different between the beneficiaries and control groups.
‡The formula for percentage increase was used to calculate the TE differential between beneficiaries and control groups.
*, P < 0.10; **, P< 0.05; ***, P< 0.01.

Table 7 Predicted frontier output after bias correction

	Min	**Mean**	**Max**
B-M-S	20,751	51,603	97,122
C-M-S	17,933	45,401	89,153
Differential	16%	14%	9%

Figure 1. Distribution of efficiency score for extreme models



References

Angrist J, Pischke J (2009) Mostly harmless econometrics. Princeton University Press, New Jersey

Battese G, Coelli T (1995) A model for technical inefficiency effects in a stochastic frontier production for panel data. Empir Econ 20:325–332

Becker SO, Ichino A (2002) Estimation of average treatment effects based on propensity scores. Stata J 2:358-77

Bradford D, Kleit A, Krousel-Wood M, Re R (2001) Stochastic frontier estimation of cost models within the hospital. Rev Econ Stat 83:302–309

Bravo-Ureta, BE, Solís D, Moreira V, Maripani J, Thiam A, Rivas T (2007) Technical Efficiency in Farming: A Meta-Regression Analysis. J Prod Anal 27:57-72

Bravo-Ureta BE, Almeida A, Solís D, Inestroza A (2011) The Economic Impact of MARENA's Investments on Sustainable Agricultural Systems in Honduras. J Agric Econ: Forthcoming

Bravo-Ureta BE (2009) Evaluación Final de Resultados del Programa MARENA. Office of Evaluation and Oversight, Inter-American Development Bank.

Caliendo, M, Kopeinig S (2008) Some Practical Guidance for the Implementation of Propensity Score Matching. J Econ Surveys 22(1): 31-72.

Cameron A, Trivedi P (2005) Microeconometrics: Methods and applications. Cambridge University Press, Cambridge

Cavatassi R, Gonzalez M, Winters P, Andrade-Piedra J, Espinosa P, Thiele G (2009) Linking smallholders to the new agricultural economy: An evaluation of the *Plataformas p*rogram in Ecuador. ESA Working Paper No. 09-06, Food and Agricultural Organization.

Cerdán-Infantes P, Maffioli A, Ubfal D (2008) The impact of agricultural extension services: The case of grape production in Argentina. Office of Evaluation and Oversight, Inter-American Development Bank.

Chavas J, Petrie R, Roth M (2005) Farm household production inefficiency in the Gambia: resource constraints and market failures. Am J Agric Econ 87:160-179

Collins A, Harris R (2005) The impact of foreign ownership and efficiency on pollution abatement expenditures by chemical plants: Some UK evidence. Scott J Polit Econ 52:757–768

Dillon A (2011) Do Differences in the scale of irrigation projects generate different impacts on poverty and production? J Agric Econ: Forthcoming.

Dinar A, Karagiannis G, Tzouvelekas V (2007) Evaluating the impact of agricultural extension on farm's performance in Crete: A non-neutral Stochastic Frontier Approach. Agric Econ 36:135-146

Duflo E, Glennerster R, Kremer M (2008) Using randomization in development economics research: A toolkit. In: Schultz T, Strauss J (eds) Handbook of development economics. Elsevier Amsterdam, pp 3895-3962

ESA (2008) Segunda medición de los indicadores del programa MARENA: Informe final. ESA Consultores, Tegucigalpa, Honduras

Greene W (2007) Econometric analysis, 6th ed. Prentice Hall, New Jersey

Greene W (2008) The econometric approach to efficiency analysis. In: Fried H, Knox Lovell C, Schmidt S (eds) The measurement of productive efficiency and productivity growth. Oxford University Press, Oxford

Greene W (2010) A Stochastic Frontier Model with correction for sample selection. J Prod Anal 34:15-24

González M, López L (2007) Political violence and farm household efficiency in Colombia. Econ Dev Cult Change 55:367-392

Heckman J (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Imbens G, J Wooldridge (2008). Recent developments in the econometrics of program evaluation. Research Working Paper No. 14251, National Bureau of Economic Research

Jondrow J, Lovell K, Materov I, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econom 19:233–238

Kalirajan K (1991) The importance of efficiency use in the adoption of technology: a micro panel data analysis. J Prod Anal 2:113-126

Kaparakis E, Miller S, Noulas A (1994) Short run cost inefficiency of commercial banks: A flexible stochastic frontier approach. J Money Credit Bank 26:21–28

Khandker SR, Koolwal GB, Samad HA (2010) Handbook on impact evaluation: Quantitative methods and practices. The World Bank, Washington D.C.

Koop G, Steel M (2001) Bayesian analysis of stochastic frontier models. In: Baltagi B (ed) Companion to theoretical econometrics. Blackwell, Oxford

Kumbhakar S, Tsionas M, Sipilainen T (2009) Joint estimation of technology choice and technical efficiency: An application to organic and conventional dairy farming. J Prod Anal 31:151–162

Lai H, Polachek S, Wang H (2009) Estimation of a stochastic frontier model with a sample selection problem. Working Paper, Department of Economics, National Chung Cheng University
Leuven E, Sianesi B (2003) *PSMATCH2: STATA* module to perform full Mahalanobis and Propensity Score Matching, common support graphing, and covariate imbalance testing. Statistical Software Components Series S432001, Department of Economics, Boston College

Maddala G (1983) Limited dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Mayen C, Balagtas J, Alexander C (2010). Technology adoption and technical efficiency: Organic and conventional dairy farms in the United States. Am J Agr Econ 92:181-195

Monteiro N (2010) Using propensity matching estimators to evaluate the impact of privatization on wages. Appl Econ 42:1293-1313

Murphy K, Topel R (2002) Estimation and inference in two stem econometric models. J Bus Econ Stat 20:88–97

Pearson M (2009) Achieving the MDGs: At what cost. Technical Approach Paper, HLSP Institute

Radelet S (2004) Aid effectiveness and the Millennium Development Goals. Center for Global Development, Washington DC

Rahman S, Wiboonpongse A, Sriboonchitta S, Chaovanapoonphol Y (2009) Production efficiency of Jasmine rice producers in northern and north-eastern Thailand. J Agric Econ 60:419-435

Ravallion M (2008) Evaluating anti-poverty programs. In: Schultz T, Strauss J. (eds) Handbook of development economics. Elsevier, Amsterdam, pp 3787-3846

Rodriguez D, Rejesus R, Aragon M (2007) Impacts of an agricultural development program for poor coconut producers in the Philippines: An approach using panel data and Propensity Score Matching Techniques. J Agric Resour Econ 32:534-557

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

Rosenbaum P, Rubin D (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Amer Statistician 39:33-38.

Sianesi B (2001). Implementing Propensity Score Matching estimators with STATA. Paper presented at VII Meeting UK STATA Group, London

Sipiläinen T, Oude Lansink A (2005) Learning in switching to organic farming. NJF-Seminar 369, NJF Report Vol 1 No 1, Nordic Association of Agricultural Scientists

Solís D, Bravo-Ureta BE, Quiroga R (2009) Technical efficiency among peasant farmers participating in natural resource management programs in Central America. J Agric Econ 60:202-219

Solís D, Bravo-Ureta BE and Quiroga R (2007) Soil conservation and technical efficiency among hillside farmers in Central America: A switching regression model. Aust J Agric Resour Econ 51:491-510

Wintgens J (2009) Coffee: Growing, processing, sustainable production. Wiley-VCH, Weinheim

World Bank (2006) Impact evaluation: the experience of the Independent Evaluation Group of the World Bank. Independent Evaluation Group, World Bank