academicJournals

Vol.6(3), pp. 94-107, March, 2014
DOI: 10.5897/JAERD12.084
ISSN 2141-2170 ©2014 Academic Journals
http://www.academicjournals.org/JAERD

Journal of Agricultural Extension and Rural Development

Full Length Research Paper

Impact of farmer research group interventions on maize farmers in Central Rift Valley of Oromia: An empirical study

Wole Kinati[1]*, Adam Bekele[2] and K. P. M Chinnan[3]

[1]Adami Tulu Agricultural Research Center, P. O. Box 35, Batu, Ethiopia.
[2]Melkasa Agricultural Research Center, Ethiopia.
[3]Haramaya University, Ethiopia.

Accepted 13 January, 2014

The project on "strengthening Technology Development, Verification, Transfer and Adoption through Farmers Research Group (FRG)" implemented in the Central Rift-Valley of Oromia from 2004 to 2009 was used to promote and institutionalize participatory research in Ethiopian agricultural research system. A group of farmers were established as maize FRG working on maize improvement in two districts. Hence, this study was initiated with the objective to provide robust evidence for policy makers, donors, farmers, and implementing actors on whether the FRG approach can contribute to household productivity and income. A Cross sectional data were collected from a 180 randomly identified participant and nonparticipants. The empirical result of impact of our estimator indicated that the program increased participant households' productivity on average by 36%. However, further analysis revealed a positive and insignificant difference for the net income generated from the intervention. Adopting interventions that follow a value chain approach is recommended in order to make the program more comprehensive in bringing significant change not only in the production but also in the subsequent livelihood outcomes.

Key words: Farmer research group, central rift-valley, maize, productivity, income, propensity score matching.

INTRODUCTION

In order for agricultural research to properly address farmers' bio-physical and socio-economic constraints and be impact oriented by addressing the needs of its clients, it has to be participatory. The Ethiopian Agricultural Research System has been trying to promote participatory research to develop and promote technologies with farmers' active involvement. Encouraging results have been observed in the process, particularly by improving interaction among stakeholders. This has brought up a need to further improve and institutionalize participatory research in the research system for quick and tangible research impacts on the client. Owing to this, the project entitled "strengthening Technology Development, Verification, Transfer and Adoption through Farmers Research Group (FRG)" was implemented in the Central Rift-Valley of Oromiya National Regional State from 2004 to 2009. This valley largely encompasses the East Shewa Zone of Oromia and has about 40 to 60 km wide and more than 1000 km length bounded by highland plateaus. The altitude ranges from 500 to 2000 m.a.s.l. and has a semi-arid type of climate. It has an erratic, unreliable and low rainfall is

*Corresponding author. E-mail: wolekinati@gmail.com.

bimodal with the long rain from June to September. The farming system is characterized by mixed crop-livestock (Abule et al., 1998).

The project operated in three Zones, namely East Shewa, Arsi, and West Arsi in Oromia Regional State. The following districts were covered: Adama, Boset, Dodota, Adami Tulu Jidokombolcha, Bora, Dugda, Arsi Negelle, and Shala. The project fully entered into operation in 2005. The aim of the project was to promote participatory agricultural research method for enabling research outputs meet farmers' needs and priorities as well as capacitate farmers to innovate so as to raise the productivity of small holders through generation, transfer and adoption of improved technologies. The project was funded by the Japan International Cooperation Agency (JICA). It was jointly implemented by the Ethiopian Institute of Agricultural Research (IARI) and Oromiya Agricultural Research Institute (OARI) for a period of five years. The two implementing centers were Melkasa Agricultural Research Center and Adami Tulu Agricultural Research Center (MARC & ATARC).

One of the goals set in the project document was to increase in the production of major commodities of the farmers around the target FRG. One of the major commodities considered by the project was maize. In Ethiopia Maize (*Zea mays*) is mainly produced for local consumption. In additional leaves are used as feed for animals and the stake is used as fuel and for construction. Millions people depend on maize as a staple food. In view of its high demand for food grains and high yield per unit area, maize has been among the leading food grains selected to achieve food self sufficiency in Ethiopia (Benti et al., 1993 cited in Chimdo, 2001). Hence maize is one of the top priority crops to which substantial resources are being allocated by the National Extension package program. Despite its importance, the national average yield of maize is around 2 ton/ha. This is really half of the world yield average of 3.7 ton/ha (Chimdo, 2001).

Several reasons were suggested for the low productivity of maize in the study are among which the major one is shortage of improved maize seed varieties. Yet while these varieties are currently being promoted through demonstration trials with smallholders throughout the Rift Valley area, widespread adoption has been tempered by difficulties in delivering improved seed to smallholders. Despite an active agricultural extension system, a sizeable state-owned seed enterprise, and the recent liberalization of seed market regulations, the availability and adoption of improved seed in the Rift Valley area remains low (Muhammad et al., 2003).

Unlike the conventional research approach where farmers are considered as the end users of technology developed at research centers, the project involves farmers (Farmers Research Group) directly into the research process. The direct involvement of farmers into research makes the technology dissemination quicker and demand driven. Interested and hard working farmers who could conduct the experiment were identified as farmer researcher for on farm trial of technologies. Each of such groups of farmer had about 15 to 20 farmers who formed FRG. Pre-extension technologies and/or completed technologies were tested by farmers in a group with a guide of researchers.

The research topics were identified by the community and facilitated by a multidisciplinary research team so that different kinds of farmers' problems were addressed. Inputs needed for the technology trial were provided by farmers and the project. The FRG approach intended to accelerate the technology dissemination process and create confidence in farmers developing their capacity to develop, modify and adopt agricultural technologies.

METHODOLOGY

FRGs were established at different locations in the target districts. Accordingly, in two districts, Adami Tulu Jido Kombolcha (ATJK) and Boset, maize producing FRGs were established. Selection of households into the program involved local consultation (experts and administrators) and a non-random placement. In the first place, peasant associations were identified in the district based on certain criteria like their accessibility to road and availability of agricultural extension services and willingness of the farmers to participate and the opportunity and potential of the peasant associations for specific commodity of intervention. Households who have been involved in FRG since 2007/2008 were considered as participants. Each FRGs consisted of 15 to 20 farmers. Although, the whole process of FRG activities intended to develop farmers' capacity, scheduled farmer trainings occurred on regular bases. Working in groups, farmers would observe and discuss dynamics of the maize's ecosystem and the crop development. The objective of these learning processes is to develop farmer expertise in crop management that then enables them to make their own decisions.

The study was targeted at these two districts where maize FRGs was established by the project. Adami Tulu Jido Kombolcha, is locate in the southern part of Oromiya where as Boset is located in the eastern part. Adami Tulu Jido Kombolcha and Boset have 1403.3 and 1461.88 km^2 of land inhabited by about 141745 and 109578 people respectively of which more than 85% are living in the rural. All the farmers are subsistence, whose livelihoods depend mainly on mixed farming of crop and livestock. Acacia species and other species generally characterize the vegetation cover of the area.

Agro-ecologically, the areas are categorized under the semi arid, with altitudes ranges from 1500 to 2000 and below 1500 m.a.s.l. for Adami Tulu Jido Kombolcha and Boset respectively. The average annual rain fall ranges from 650 to 750 mm and the distribution is highly variable between and within years. The identified major type of soil is fine sandy loam with sand silt clay (Abule et al., 1998). Open woodland consists of Acacia species and other species generally characterizes the vegetation cover of the areas.

Sample size and sampling techniques

Based on the data from the FRG project document, there were about 143 farmers involved in maize FRG in these districts. Table 1 presents the number of farmers by sex from each implementing centers.

A random total sample size of 180 was identified for the study. Seventy two participant households with sampling proportion of 50% (72 farmers out of 143) were selected randomly using

Table 1. Sample size by peasant associations.

Districts	Peasant associations	No. of FRG participants	Farmers interviewed					
			FRG participants		Non-FRG participants		Total	
			N	%	N	%	N	%
ATJK	Anano shisho	67	34	47	-	-	34	18
	Desta abijata	-	-	-	51	47	51	28
	Sub total	*67*	*34*	*47*	*51*	*47*	*85*	*48*
Boset	Dongore furda	40	20	28	-	-	20	11
	Dongore tiyo	36	18	25	-	-	18	10
	Hurufa kurkufa	-	-	-	57	53	57	32
	Sub total	*76*	*38*	*53*	*57*	*53*	*95*	*53*
	Grand total	143	72	100	108	100	180	101*

*, Results do not add up to 100 because of rounding.

probability proportionate to sample size technique. A second random sample of 108 farmers was drawn from the population of nonparticipating maize growers living in the same district where the FRG project took place from a purposively selected *Kebeles*[1]. In doing so these *Kebeles* were purposively identified using agro-ecological criteria to provide representation of maize dominating cropping system. There were also two other reasons: there has to be a substantial difference in terms of distance so that information exchange between FRG participant (treatment group) and non-FRG participant (control group) is minimized and the selected *Kebeles* should be accessible. Then, a list of households in each *Kebeles* was drawn up and maize producers were identified. For this purpose, the survey team constructed lists of nonparticipating maize farmers for the given locality in consultation with Development Agents (DAs) of the selected peasant associations, supervisors and village elders. As the interest was in maize producers, only farmers producing maize included in the sampling frame and households were randomly sampled.

Sources of data and methods of data collection

Cross-sectional data was collected from the sample households by administering interview schedule. The interview schedule was pretested by administering it to selected respondents which we excluded from the sample frame during sampling. On the basis of the results obtained from the pre-test, necessary modification was made on the interview schedule. Both sampled FRG and non-FRG farmers in the selected enumeration area were visited and interviewed using the same scheduled interviews and data collection done from December, 2009 to January, 2010.

Analytical methodology

In the more general extension literature, extension impacts per se are very difficult to show, especially in terms of dealing with attribution issues and linking cause and effect quantitatively (Purcell and Anderson, 1997 cited in Davis et al., 2010). Many infrastructural variables and other factors affect agricultural performance in complex and contradictory ways, and benefits are difficult to quantify (Anderson, 2007). Impact studies basically face three interrelated challenges: (a) Establishing a viable counterfactual (the predicted outcome in the absence of the intervention, that is, what would have happened to the participants had they not participated in the FRG; (b) Attributing the impact to an intervention; and (c) Coping with long and unpredictable lag times (Alston and Pardey, 2001; Salter and Martin, 2001 quoted in Davis et al., 2010). Other issues that may confound studies include endogeneity in program placement and extension–farmer interactions, farmer-to-farmer information flow, selection bias, and policies that affect various measures. Very few studies use an experimental design, and some studies that have used control groups have run into design problems (Davis et al., 2010).

Two common sources of bias are program placement or targeting bias, in which the location or target population of the program is not random, and self-selection bias, in which households choose whether or not to participate, and thus may be different in their experiences, endowments, and abilities.

The most accepted method to address the previously mentioned biases is to use an experimental approach to construct an estimate of the counterfactual situation by randomly assigning households to treatment (participant) and control (nonparticipant) groups. Random assignment ensures that both groups are statistically similar (that is, drawn from the same distribution) in both observable and unobservable characteristics, thus avoiding program placement and self-selection biases. Such an approach is not feasible in demand-driven programs in which participants make their own decisions of whether to participate and about the kind of activities to do in the

learning process. Likewise, random assignment also conflicts with the nature of community-driven development programs like FRG.

To address the problems of showing impact, several quasi-experimental methods have been developed to net out the impacts of other factors. These include; double difference or difference-in-difference (DID), reflexive comparison and propensity score matching (PSM). A common approach is the use of PSM method. Thus, using a cross sectional household survey for this study, we isolate the causal effect of participating in FRG on the outcome variables by using PSM method.

Propensity score matching method

Several matching methods have been developed to estimate causal treatment effects. A commonly used matching method is propensity score matching (PSM). It applies for all situations where one has a treatment, a group of treated individuals and a group of untreated individuals (Caliendo and kopeinig, 2008). The impact of FRG intervention on household's given outcome is the difference in households' mean outcome with the program and without the

[1] The smallest administrative structure next to Woreda

program. However, households participating in the program cannot be simultaneously observed in two states. A household can either be in the program or outside the program. Thus, the fundamental problem of such an impact evaluation is a missing data problem. In other words, we are interested in answering the research question "what would have been the productivity and income outcomes of participating households if FRG was not in place?" Hence, this study applies a propensity score matching technique, which is a widely applied impact evaluation instrument in the absence of baseline survey data for impact evaluation.

The preference of PSM over the other conventional regression methods lies in its unique characteristics in which it compares outcome for observations who share similar observable characteristics[2] and only compares households lay in the common support region and excluded others from the analysis.

This study attempts to estimate the average impact of treatment on treated (ATT). According to Bryson et al. (2002), ATT refers to mean impact[3] of the program on individuals who actually participated. In this study "treatment" implies participation in the program (in FRG). In employing PSM method in assessing treatment effect, according to Caliendo and Kopeinig (2008), there are procedures to be followed. These are estimation of the propensity scores, choosing a matching algorism, checking overlap/common support condition and testing the matching quality/effect estimation.

Propensity score estimation procedure

The first step in PSM method is to estimate the propensity scores. As described by Rosenbaum and Rubin (1983), matching can be performed conditioning on P(X) alone rather than on X, where P(X) = Prob(D=1|X) is the probability of participating in the program conditional on X. If outcomes without the intervention are independent of participation given X, then they are also independent of participation given P(X). This reduces a multidimensional matching problem to a single dimensional problem (ibid.).

A logit model was used to estimate propensity scores using a composite of pre-intervention characteristics of the sampled households (Rosenbaum and Robin, 1983) and matching was then performed using propensity scores of each observation. In estimating the logit model, the dependent variable was participation, which takes the value of 1 if a household participated in the program and 0 otherwise. The mathematical formulation of logit model is as follows:

$$P_i = \frac{e^{Z_i}}{\quad} \qquad (1)$$

Where, P_i is the probability of participation.

$$Z_i = a_0 + \sum_{i=1}^{n} a_i X_i + U_i \qquad (2)$$

Where, i = 1, 2, 3, - --, n; a_0 = intercept; a_i = regression coefficients to be estimated; X_i = pre-intervention characteristics, and U_i = a disturbance term, and the probability that a household belongs to non participant is:

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \qquad (3)$$

According to matching theory (Rosenbaum and Robin, 1983; Jalan and Ravallion, 2003; Bryson et al., 2002), the logit model via which the propensity score is generated should include predictor variables that influence the selection procedure or participation in the program and the outcome of interest. Several factors guide selection of predictor variables. In the present study, explanatory variables of the logit model were identified using findings of previous related empirical studies, FRG targeting criteria, and own field observation. We included as many explanatory variables as possible to minimize the problem of unobservable characteristics in our evaluation of the impact of the program.

Matching estimators

After estimation of the propensity scores, seeking an appropriate matching estimator is the major task of a program evaluator. There are different matching estimators in theory. The most common once are NN, Caliper and Kernel matching[4].

[2] PSM technique has attracted attention of social program evaluators since the last fifteen years (see for e.g., Jalan and Ravallion, 2003; Dehejia and Wahba, 1999). The PSM technique enables us to extract from the sample of non-participating households a set of matching households that look like the participating households in all relevant pre-intervention characteristics. In other words, PSM matches each participant household with a non-participant household that has (almost) the same likelihood of participating into the program.

[3] "Impact" is meant for the change in production and income using productivity and income level as an outcome indicator. On the other hand, "control" stands for non-participant/non-treated households used for comparison.

[4] **Nearest Neighbour (NN) Matching**: according to Caliendo (2008), the most straightforward matching estimator is Nearest Neighbour. The individual from the comparison group is chosen as a matching partner for a treated individual that is closest in terms of the propensity score. NN matching can be done with or without replacement options. In the case of the NN matching with replacement, a comparison individual can be matched to more than one treatment individuals, which would result in increased quality of matches and decreased precision of estimates. On the other hand, in the case of NN matching without replacement, a comparison individual can be used only once. Matching without replacement increases bias but it could improve the precision of the estimates. In cases where the treatment and comparison units are very different, finding a satisfactory match by matching without replacement can be very problematic (Dehejia and Wahba, 2002). It means that by matching without replacement, when there are few comparison units similar to the treated units, we may be forced to match treated units to comparison units that are quite different in terms of the estimated propensity score.

Caliper Matching: NN matching faces the risk of bad matches if the closest neighbor is far away (Caliendo, 2008). To avoid this problem researchers use the second alternative matching algorism called caliper matching by imposing a tolerance level on the maximum propensity score distance(caliper). Caliper matching means that an individual from the comparison group is chosen as a matching partner for a treated individual that lies within a given caliper (propensity score range) and is closest in terms of propensity score (Kopeinig, 2005). If the dimension of the neighborhood is set to be very small, it is possible that some treated units are not matched because the neighborhood does not contain a control unit. On the other hand, the smaller the size of the neighborhood the better is the quality of the matches (Becker and Ichino, 2002). One possible drawback in caliper matching as Smith and Todd (2005) cited in Caliendo (2008) indicated is that it is difficult to know a *priori* what choice for the tolerance level is reasonable.

Kernel Matching: the matching algorisms discussed so far have in common that only a few observations from the comparison group are used to construct the counterfactual outcome of a treated individual. Kernel matching is nonparametric matching estimator that use weighted averages of (nearly) all individuals in the control group to construct the counter factual outcome. Accordingly, all treated units are matched with a weighted average of all controls with weights which are inversely proportional to the distance between the propensity scores of treated and controls (Becker and Ichino 2002; Venetoklis, 2004). Kernel weights the contribution of each comparison group member so that more importance is attached to those comparators providing a better match. The difference from caliper matching, however, is that those who are included are weighted according to their proximity with respect to the

Generally, the choice of a given matching estimator depends on the nature of the available dataset (Bryson et al., 2002). In other words, it should be clear that there is no `winner' for all situations and that the choice of a matching estimator crucially depends on the situation at hand. The choice of a specific method depends on the data in question, and in particular on the degree of overlap between the treatment and comparison groups in terms of the propensity score. When there is substantial overlap in the distribution of the propensity score between the comparison and treatment groups, most of the matching algorithms will yield similar results. In case there are only a few control observations, it makes no sense to match without replacement. On the other hand, if there are a lot of comparable untreated individuals it might be worth using more than one nearest neighbor to gain more precision in estimates (Caliendo and Kopeinig, 2005).

Overlap and common support condition

As ATT is only defined in the region of common support; Heckman et al. (1997) quoted in Caliendo and Kopeinig (2008) point out that a violation of the common support condition is a measure of evaluation bias as conventionally measured. Comparing the incomparable must be avoided, that is, only the subset of the comparison group that is comparable to the treatment group should be used in the analysis. Hence, an important step is to check the overlap and the region of common support between treatment and comparison group.

Imposing a common support condition ensures that any combination of characteristics observed in the treatment group can also be observed among the control group (Bryson et al., 2002). The common support region is the area which contains the minimum and maximum propensity scores of treatment and control group households, respectively. It requires deleting of all observations whose propensity scores is smaller than the minimum and larger than the maximum of treatment and control, respectively (Caliendo and Kopeinig, 2005).

Assessing the matching quality

According to Caliendo (2008), matching quality has to be checked if the matching procedure is able to balance the distribution of the relevant variables in both the control and treatment group, since we do not condition on all covariates but on the propensity score. To do this, several procedures used in the literature includes standard bias, t-test, joint significance and pseudo-R^2 and stratification test. The basic idea of all approaches is to compare the situation before and after matching and check if there any differences after conditioning on the propensity score.

The primary purpose of the PSM is that it serves as a balancing method for covariates between the two groups since differences in covariates are expected before matching and should be avoided after matching. Consequently, the idea behind balancing tests is to check whether the propensity score is adequately balanced. In other words, a balancing test seeks to examine if at each value of the propensity score, a given characteristic has the same distribution for the treatment and comparison groups. The propensity scores themselves serve only as devices to balance the observed distribution of covariates between the treated and comparison groups. The success of propensity score estimation is therefore assessed by the resultant balance rather than by the fit of the models used to create the estimated propensity scores (Lee, 2006).

Finally, using predicted probabilities of participation in the program (that is, propensity score) match pairs are constructed using alternative methods of matching estimators. Then the impact estimation is the difference between simple mean of outcome variable of interest for participant and non participant households. In our case, the mean stands for household productivity and income. The difference involvement in FRG between treatment and matched control households is then computed. The ATT is obtained by averaging these differences in FRG outcomes (Y_i) across the k matched pairs of households as follows:

$$ATT = \sum_{j=1}^{P}\left(Y_{ij1} - \sum_{i=1}^{NP} Y_{ij0}\right)/P \tag{4}$$

Where, ATT is productivity and income, Y_{ij1} is the post intervention productivity and income of household j, Y_{ij0} is the productivity and income of household of the i^{th} non-participant matched to the j^{th} participant, NP is the total number of non-participants and P is the total number of participants. A positive (negative) value of ATT suggests that households who have participated in FRG have higher (lower) of outcome variable Y_i non-participants.

Sensitivity analysis

It should be clear that matching estimators are not robust against 'hidden biases due to un observable characteristics, selection bias. Different researchers become increasingly aware that it is important to test the robustness of results to departures from the identifying assumption. Since it is not possible to estimate the magnitude of selection bias with non-experimental data, the problem can be addressed by sensitivity analysis.

Rosenbaum and Robin (1983) proposed using Rosenbaum bounding approach in order to check the sensitivity of the estimated ATT with respect to deviation from the Conditional Independence Assumption (CIA). The basic question to be answered here is whether inference about treatment effects may be altered by unobserved factors. In other words, one wants to determine how strongly an unmeasured variable must influence the selection process in order to undermine the implications of matching analysis. Rosenbaum bounds provide evidence on the degree to which any significance results hinge on this untestable assumption. If the results turn out to be sensitive, the evaluator might have to think about the validity of his identifying assumption and consider other estimation strategies.

Variable choice and its definitions

Choice and definition of explanatory variables

There are no general rules for which variables to include in the model (Anderson et al., 2009). However, Bryson et al. (2002) suggest that, economic theory and knowledge about previous research and also information about the institutional settings should guide the researcher to know which observables (explanatory variables) affect both participation and the outcomes of interest[5].

propensity score. The most common approach is to use the normal distribution (with a mean of zero) as a kernel, where the weight attached to a particular comparator is proportional to the frequency of the distribution for the difference in scores observed (Bryson et al., 2002). According to Caliendo (2008), a drawback of this method is that possibly bad matches are used as the estimator includes comparator observations for all treatment observation. Hence, the proper imposition of the common support condition is of major importance for kernel matching method. A practical objection to its use is that it will often not be obvious how to set the tolerance. However, according to Mendola (2007) kernel matching with 0.25 band width is most commonly used.

[5] In the estimation of the propensity score, we are not interested in the effects of covariates on the propensity score because the purpose of our work is to

Accordingly, different socioeconomic, demographic, institutional and location factors were identified as shown in Table 2.

Choice, indicators and measurements of the outcome variables

Impact on crop productivity: Crop productivity is defined as the value of production per unit area (Davis et al., 2010). This is one of the outcome variables for which this study intends to measure. It is expected that FRG interventions will improves the productivity of commodity of intervention. The effect of FRG interventions on the commodity of intervention is measured in yield per unit of area (quintal/ha) increase.

Impact on household net income: Household net income is also one of the outcome variables as a result of the household's participation in FRG which is measured in Birr. Household net income is calculated as the difference between the total revenue generated from sale of commodity of intervention (maize) and total cost incurred by households for the production of this particular commodity of intervention (Davis et al., 2010).

Before estimating the models, it was necessary to check if multicollinearity exists among the explanatory variables. The existence of strong multicollinearity seriously affects the parameter estimates of the regression models, it is necessary to check it's occurrence among the explanatory variables. Accordingly, Variance Inflation Factor (VIF) technique was employed to detect the problem of multicollinearity for the variables (Gujarati, 2004). It was calculated as:

$$VIF(x_j) = \frac{1}{1 - R_j^{\,2}} \qquad (5)$$

Where R_j^2 is the squared multiple correlation coefficient between and other explanatory variables. Each selected variable is regressed on all the other variables, the coefficient of determination (R_j^2) being constructed in each case. If a strong linear relationship exists among the explanatory variables then this would result in large VIF value. The larger value of VIF (X_j), the more troublesome, as a rule of thumb, if the VIF of a variable exceeds 10 (this will happen if R_i^2 exceeds 0.95), that variable is said to be highly collinear (Gujarati, 1995), and is used as a signal for the existence of a severe multicollinearity among explanatory variables. In the same way, for dummy variables contingency coefficient test were employed using the following formula:

$$= \sqrt{\frac{X^2}{n + X^2}} \qquad (6)$$

Where, *C* is coefficient of contingency, x^2 is chi-square test and *n* is total sample size. For dummy variables if the value of contingency coefficients is greater than 0.75 the variable is said to be collinear. Another problem in regression analysis is the problem of heteroscedasticity in the data and this was detected by using Breusch-Pagen test (hettest) in STATA.

To analyze the data, the estimation was run by employing propensity score matching algorism with STATA 10.0 Software using the STATA code written by Leuven and Sianesi (2003).

RESULTS AND DISCUSSION

The effects of FRG on maize crop productivity, market surplus, and agricultural income (income from maize by households) using the analytical methods explained earlier was examined here. In doing so the important steps followed to arrive at the impact of the program was also described here. It explains the estimation of propensity scores, matching methods, common support region, balancing test and treatment effect.

Propensity scores

The results of the logistic regression model which was used to estimate propensity scores for matching program households with non-program households was presented here. The dependent variable in this model is a binary variable indicating whether the household was a participant in the program. In the estimation data from the two groups; namely, program and nonprogram households were pooled such that the dependent variable takes a value 1 if the household was participant and 0 otherwise. Before proceeding to impact estimation, Variance Inflation Factor (VIF) was applied to test for the presence of strong multicollinearity problem among the continuous explanatory variables. Moreover, by using contingence coefficients (C) multicollinerty between discrete variables were checked.

There were no explanatory variables dropped from the estimation model since no serious problem of multicollinearity was detected from the VIF results. Similarly, heteroscedasticity was tested by using Breusch-Pagen test. This test resulted in the existence of heteroscedasticity problem as it is significant at 5% probability level (p = 0.0294) suggesting the need for standard error robust. Hence, robust standard error was conducted accordingly.

Table 3 shows the estimation results of the logit model. The estimated model appears to perform well for our intended matching exercise. The pseudo-R^2 value is 0.3075. According to Pradhan and Rawlings (2002), a low R^2 value shows that the allocation of the program has been de facto random. In other words, a low R^2 value means that program households do not have much distinct characteristics over all and as such finding a good match between program and non-program households becomes easier. The pseudo-R^2 indicates how well the regressors explain the participation probability. After matching there should be no systematic differences in the distribution of covariates between both groups and therefore, the pseudo- $R2$ should be fairly low (Caliendo

assess the impact of FRG interventions on outcome variables. However, the choice of covariates to be included in the first step (propensity score estimation) is an issue. Heckman et al. (1997) and Dehejia and Wahba (1999) cited in Caliendo (2008) argue that omitting important variables can increase the bias in the resulting estimation. Only variables that influence simultaneously the participation decision and outcome variable should be included. Accordingly, variables that determine households' decision to participate in FRG could also affect the outcome variable mentioned above. Here, pre-intervention characteristics, which bring variation in outcomes of interest among program participants and non-participants, were used. In other word, variables which are not affected by being participate in the program or not or those explanatory variables which are fixed throughout are assumed to be used as explanatory variables.

Table 2. Type, definitions and measurement of variables.

Variable	Types and definition	Measurements
Dependent variables		
Treatment	Dummy, participation in FRG maize	1 if yes,0 otherwise
Covariates		
AGEHH	Continuous, age of the household head	in years
EDULHH	Dummy, education level of household head	1 if literate, 0 otherwise
FAMSIZE	Continuous, total family size of the household	number of household
FAREXP	Continuous, farming experience of household head	in years
TLOWN	Continuous, total land owned	in hectare
TLU	Continuous, livestock holding size	tropical livestock unit
DISNMKT	Continuous, distance to the nearest market	in kilometers
DISEXTO	Continuous, distance to extension office	in kilometers
DPCRTO	Continuous, dependency ratio	number of dependents in the household

Source: Own definitions.

Table 3. Logit estimation results of household program participation.

Covariates	Coefficients	Robust Std. Err.	Z
AGEHH	-0.0567592	0.0397389	-1.43
EDULHH	-0.2737916	0.4495109	-0.61
FAREXP	0.0294119	0.0412445	0.71
FAMSIZE	-0.0445323	0.0684756	-0.65
DPCRTO	-0.1228543	0.3757797	-0.33
TLU	0.0467609	0.0166224	2.81***
TLDOWN	0.6372433	0.1940986	3.28***
DISEXTO	-2.300553	0.49666	-4.63***
DISNMKT	0.0250949	0.2235155	0.11
_cons	1.010433	1.247629	0.81
N	178		
Wald $chi^2(9)$	43.21		
Prob > chi^2	0.000		
Log pseudo likelihood	-83.176653		
Pseudo R2	0.3075		

Source: Own estimation result. ***, Significant at the 1% probability level.

and Kopeinig, 2005).

The logit estimation results, when looked into the estimated coefficients (Table 3), indicate that program participation is significantly influenced by three explanatory variables. Sizes of livestock ownership (in TLU), size of land ownership and distance from the nearest extension office are significant variables which affect the participation of the household to the program. Size of livestock ownership and land holding are found to have strong and positive relationship with household participation in the program. This means households with more size of livestock ownership and land holding are more likely to be included in the program. In the contrary, distance from the nearest extension office has strong and negative effect on the household participation suggesting that households leaving relatively far away from extension office have less likely to participate in the program.

Figure 1 portrays the distribution of the household with respect to the estimated propensity scores. In case of treatment households, most of them are found in partly the middle and partly in the right side of the distribution. On the other hand, most of the control households are partly found in the center and partly in the left side of the distribution.

Matching program and non-program households

Three main tasks were accomplished here before

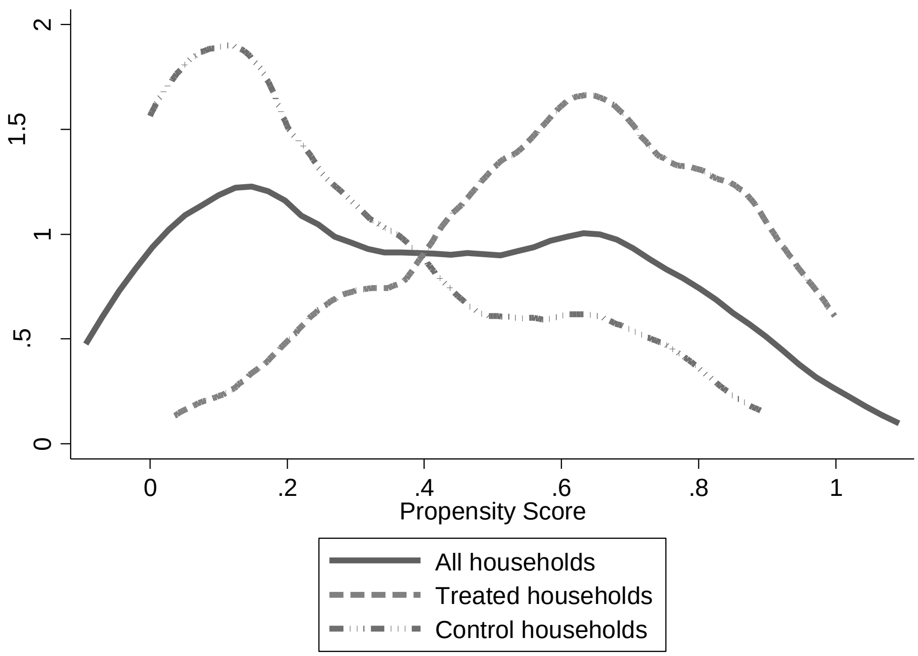

Figure 1. Kernel density of propensity scores

Table 4. Distribution of estimated propensity scores.

Group	Obs.	Mean	Std. Dev.	Minimum	Maximum
Total households	178	0.4045	0.2931	1.22e-06	0.9982
Treated households	72	0.6143	0.2286	0.0357	0.9982
Control households	106	0.2620	0.2426	1.22e-06	0.8899

Source: Own estimation result.

conducting the matching estimator. First, predicted values of program participation (propensity scores) was estimated for all households in the program and outside the program. Second, a common support condition was imposed on the propensity score distributions of household with and without the program. Then, thirdly, observations whose predicted propensity scores fall outside the range of the common support region was discarded. As shown in Table 4, the estimated propensity scores vary between 0.0357 and 0.9982 (mean = 0.6143) for program or treatment households and between 1.22E-06 and 0.8899 (mean = 0.262) for nonprogram (control) households. The common support region would then lie between 0.0357 and 0.8899. In other words, households whose estimated propensity score is less than 0.0357 and larger than 0.8899 are not considered for the matching exercise. As a result of this restriction, 31 households (10 program and 21 control households) were discarded and not used in computing the impact estimator.

As it can be observed from Figures 2 and 3, the distribution of estimated propensity scores, with and without the imposition of the common support condition, is around and less than 0.5 for program and non-program households, respectively.

Choice of matching algorism

Alternative matching estimators were tried in matching the treatment program and control households in the common support region. The final choice of a matching estimator was guided by different criteria such as equal means test referred to as the balancing test (Dehejia and Wahba, 2002), pseudo-R^2 and matched sample size. Specifically, a matching estimator which balances all explanatory variables (that is, results in insignificant mean differences between the two groups), bears a low R^2 value and also results in large matched sample size is preferable.

Table 5 presents the estimated results of tests of matching quality based on the above mentioned

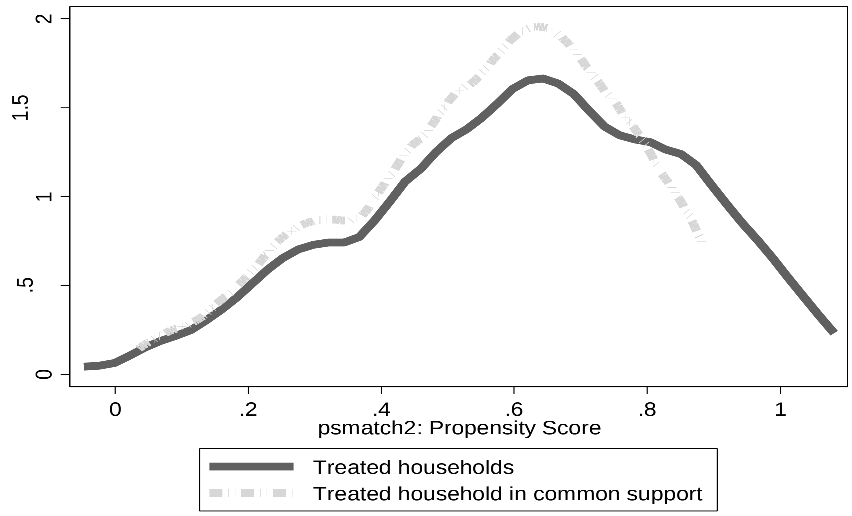

Figure 2. Kernel density of propensity scores of program households.

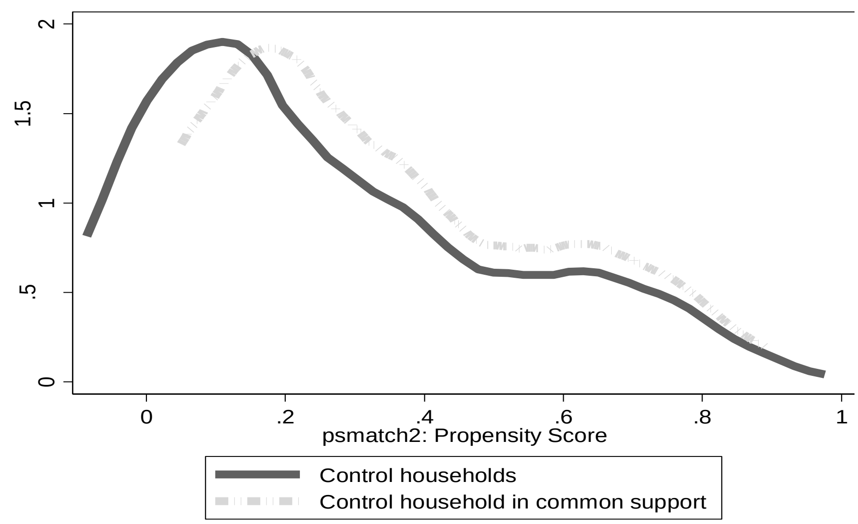

Figure 3. Kernel density of propensity scores of non-program households.

performance criteria. After looking into the results, it has been found that kernel matching with a band width of 0.1 is the best estimator for the data we have. As such, in what follows estimation results and discussion are the direct outcomes of the kernel matching algorithm based on a band width of 0.1.

Testing the balance of propensity score and covariates

Table 6 shows the balancing test of covariates, before and after the matching. As the table indicates, program and non-program households were significantly different in terms of certain pre-intervention characteristics. However, these differences were removed after the matching was conducted.

The low pseudo-R^2 and the insignificant likelihood ratio tests support the hypothesis that both groups have the same distribution in covariates X after matching (Table 7). These results clearly show that the matching procedure is able to balance the characteristics in the treated and the matched comparison groups. We, therefore, used these results to evaluate the effect of FRG intervention among groups of households having similar observed characteristics. This allowed us to compare observed outcomes for participants with those of a comparison groups sharing a common support. The details of other Chi-square tests for joint significance for the three different matching algorithms are presented

Table 5. Performance of matching estimators.

Matching estimator	Performance criteria		
	Balancing test*	Pseudo-R^2	Matched sample size
NN			
NN(1)	9	0.007	147
NN(2)	9	0.007	147
NN(3)	9	0.007	147
NN(4)	9	0.007	147
NN(5)	9	0.007	147
Caliper			
0.01	9	0.033	85
0.25	9	0.015	147
0.50	8	0.099	147
Kernel			
Band width of 0.1	9	0.004	147
Band width of 0.25	9	0.009	147
Band width of 0.5	9	0.047	147

Source: Own estimation result. *Number of explanatory variables with no statistically significant mean differences between the matched groups of program and non-program households.

Table 6. Propensity score and covariate balance.

Variable	Before matching (178)			After matching (147)		
	Treatment (N=72)	Control (N= 106)	T-value	Treatment (N=62)	Control (N=85)	T-value
AGEHH	40.972	39.858	0.65	39.790	39.701	0.05
EDULHH	0.653	0.632	0.27	0.694	0.670	0.28
FAREXP	25.194	22.377	1.77*	24.000	24.053	-0.03
FAMSIZE	8.264	6.991	2.46**	8.032	8.112	-0.12
DPCRTO	0.696	0.724	-0.24	0.736	0.775	-0.22
TLU	19.678	12.914	3.32***	16.680	16.221	0.21
TLDOWN	2.948	1.939	5.1***	2.545	2.566	-0.11
DISEXTO	0.416	1.095	-5.9***	0.428	0.437	-0.14
DISNMKT	0.872	1.553	-4.31***	0.880	0.838	0.28

Source: Own estimation result. ***, ** and*, Significant at 1, 5 and 10% probability levels, respectively.

Table 7. Chi-square test for the joint significance of variables.

Sample	Pseudo R^2	LR chi^2	$p>chi^2$
Unmatched	0.308	74.07	0.000
Matched	0.004	0.67	1.000

Source: Own estimation result.

under Appendix 1.

All of the above tests suggest that the matching algorithm we have chosen is relatively the best one with the data we have at hand. Therefore, we can proceed to estimate ATT for households' in order to answer the second objective of this study.

Impacts of FRG on various outcomes

Here, the study provides evidence as to whether or not the FRG has brought significant changes on household productivity and income from the commodity of intervention. The estimation result presented in Table 8

Table 8. ATT for productivity commodity of intervention.

Outcome variable of interest	Treated	Controls	Difference	S.E.	T-value
Household maize productivity(quintal/ha)	25.5	16.3	9.2	2.9	3.13***

***, Significant at 1% probability level.

Table 9. ATT for proportion of produce sold for commodity of intervention.

Outcome variable of interest	Treated	Controls	Difference	S.E.	T-value
Household market surplus(qtls)	31.55	19.97	11.58	6.85	1.69*

*, Significant at 10% probability level.

Table 10. ATT for household's gross income and net income.

Outcome variable of interest	Treated	Controls	Difference	S.E.	T-value
Households' gross income	20015.92	14103.64	5912.28	3358.95	1.76*
Net income(birr)	11883.33	8220.93	3662.40	2627.21	1.39

*, Significant at 10% probability level.

provides a supportive evidence of statistically significant effect of the program on household maize productivity, market surplus and gross income measured in quintals per hectare, portion of yield marketed in quintal and Birr respectively. However, the result showed that there is positive and insignificant difference between program participant and nonparticipant in terms of net income generated from the sale of increased maize produce.

After controlling for pre-intervention differences in demographic, location, institutional and asset endowment characteristics of the FRG and non-FRG households, it has been found that, on average, the program has increased maize productivity of the participating households by 9.2 quintals or by 36% (Table 8). The result is consistent with several other studies showing positive effects of similar interventions on crop productivity (Davis et al., 2010; Gockowski et al., 2006).

Our findings in Table 9 indicate that the proportion of maize sale is high for treated (31.1quintals) as compared to their counterparts (19.97quintals). In other words when the difference is tested, it is statistically significant at 10% probability level.

Similarly the result of our impact estimation proved that the project has succeeded in increasing the participant household's gross income by 5912.28 birr (Table 10). However, the empirical analysis for the net income from the sale of maize indicates that the difference between the two groups does not yield statistically significant effect (P>0.1). In other words when the total variable cost is deducted from this gross income, the result became positive but statistically insignificant. This could be attributed to the high cost of input by the program participants due to the inefficient input delivery system which involves high transaction costs and the nonexistent of concurrent market interventions for the produce in line with the commodity improvement intervention by the project that could help to achieve the ultimate objectives of the program-improved household income.

Table 11 shows the result of sensitivity of FRG intervention effects on different outcome variables in order to control for unobservable biases. The first row presents the critical level of e^{γ}, at which the causal inference of significant FRG intervention effect has to be questioned. As noted by Hujer et al. (2004), sensitivity analysis for insignificant effects is not meaningful and is therefore not considered here. Given that the estimated FRG intervention is positive for the significant outcomes, the lower bounds under the assumption that the true treatment effect has been underestimated were less interesting (Becker and Caliendo, 2007) and therefore not reported in this study. Rosenbaum bounds were calculated for FRG intervention effects that are positive and significantly different from zero. The first column of the table shows those outcome variables which bears statistical difference between treated and control households in our impact estimate above. The rest of the values which corresponds to each row of the significant outcome variables are p-critical values (or the upper bound of Wilcoxon significance level -Sig^{+}) at different critical value of e^{γ}.

Result show that the inference for the effect of FRG intervention is not changing though the participants and non participant households has been allowed to differ in their odds of being treated up to 200% ($e^{\gamma}=3$) in terms of unobserved covariates. That means for all outcome

Table 11. The result of sensitivity of FRG intervention effects on different outcome variables.

No.	Outcome variables	$e^{\gamma}=1$	$e^{\gamma}=1.25$	$e^{\gamma}=1.5$	$e^{\gamma}=1.75$	$e^{\gamma}=2$	$e^{\gamma}=2.25$	$e^{\gamma}=2.5$	$e^{\gamma}=2.75$	$e^{\gamma}=3$
1	Households' maize productivity	0	0	0	0	1.1e-16	4.4e-15	9.4e-14	1.2e-12	9.3e-12
2	Households' market surplus	0	0	0	0	2.6e-15	8.1e-14	1.3e-12	1.3e-11	8.4e-11
3	Households' gross income	0	0	0	0	1.1e-16	4.4e-15	9. 5e-14	1.2e-12	9.4e-14

Source: Own estimation. $e^{\gamma}=$ (Gamma)=log odds of differential due to unobserved factors where Wilcoxon significance level for each significant outcome variable is calculated.

variables estimated, at various levels of critical value of e^{γ}, the p- critical values are significant which further indicate that we have considered important covariates that affected both participation and outcome variables. We could not get the critical value e^{γ} where the estimated ATT is questioned even if we have set e^{γ} largely up to 3. Thus, it can be concluded that impact estimates (ATT) of this study are insensitive to unobserved selection bias and are a pure effect of FRG intervention.

CONCLUSIONS AND POLICY IMPLICATIONS

This study provides crucial insights into and important evidence on the impact of Farmer Research Group (FRG) implemented in the Central Rifty Valley of Oromia on the maize FRG farmers using cross sectional data collected for the same purpose. Using matching estimator (propensity score matching), the study evaluated the FRG program.

The result revealed that, as expected, participation in the program was determined by a combination of factors. Program participation is significantly influenced by three explanatory variables. Sizes of livestock ownership (in TLU), size of land ownership and distance from the nearest extension office are the significant variables which affect the participation of the household in the program. Households with more size of livestock ownership and land holding are more likely to be included in the program. By contrast, distance from the nearest extension office has strong and negative effect on the household participation suggesting that households leaving relatively far away from extension office are less likely to participate in the program[6].

After controlling for such characteristics, the empirical findings revealed that FRG had the largest impact on crop productivity. Significantly raising maize productivity of participating households in the study area. More particularly, the program increased participating households' productivity on average by 9.2 quintals per hectare. Which is in fact 36% more than what they would have produced in the absence of the program. The impact of the project on the proportion of produce sold to the market is also significant. Treated households sold significantly large proportion of their produce compared to their counterparts. However, when the gain is converted in to monetary value, after the total variable cost is deducted in order to see the net income, the estimated result revealed that the result became positive but insignificant. This could be attributed to the high cost of input due to inefficient input delivery system which involves high transaction costs and the nonexistent of concurrent market interventions for the program participants' produce in line with the commodity improvement intervention by the project that could help to achieve the ultimate objectives of the program-improved household income. In conclusion, the results of this study tell us that it is misleading looking only in to the productivity as indicators for program performance.

FRGs as participatory approaches are important research and development efforts to improve livelihoods of farmers if implemented properly. Based on the empirical findings reported in this study, the following policy recommendations are forwarded: As it can be observed from the empirical results, this study has found evidence that FRG in the study area has worked in significantly increasing household productivity. This sends an encouraging signal for program designers, implementers, and funding agencies. On the other hand, further improvement in the households' productivity and income from similar interventions could be enhanced for better

[6] Finding a reliable estimate of the program impact thus necessitates controlling for all such factors adequately. In doing so, propensity score matching has resulted in 62 program households to be matched with 85 non-program households. In other words, a matched comparison of different outcome variables of interest was performed on these households who shared similar pre-intervention characteristics except the program intervention. The resulting matches passed a variety of matching quality tests and were fit for answering the study's main objective.

livelihoods outcome by taking the following policy measures.

First, adopting interventions that follow a value chain approach is very important so that the program will be more comprehensive in bringing significant change not only in the production but also in the subsequent livelihood outcomes. Therefore, under this recommendation the following points were found crucially important but missed in the current program under study:

1. Effective and efficient input delivery mechanisms should be combined with productivity and income improvement programs. This can be possible through the use of the same approach (farmers group) so that access to input services can be enhanced. Furthermore, cost of input delivery can be minimized by linking farmers groups with input dealers.
2. On the other hand, lack of access to market has a potential in significantly reducing farmers income from their produce if market interventions are not part of the program as revealed by this study.

Second, strengthening actors involved along the value chain is recommended in order to reduce the transaction costs created in the input delivery and output marketing processes. Hence, policy makers can also increase household productivity and income for the betterment of rural livelihoods by furthering investment on those interventions giving considerable attention to the participation of target peoples in their programs.

ACKNOWLEDGEMENTS

The authors are very grateful to the Oromia Agricultural Research Institute for financing this study. They also gratefully acknowledge financial support of the JICA (Japan International Cooperation Agency) as main sponsor of the FRG project.

REFERENCES

Abule E, Gremew E, Aliye H (1998). Adami Tulu Agricultural Research Center, Bulletin No 1, Oromia Agricultural Development Bureau, Finfine, Ethiopia.

Bryson A, Dorsett R, Purdon S (2002). The use of propensity score matching in the evaluation of labor market policies. Working Paper No. 4. Department for Work and Pensions.

Caliendo M, Kopeinig S (2005). Some Practical Guidance for the Implementation of Propensity Score Matching, Discussion Paper No. 1588, University of Cologne.

Caliendo M, Kopeinig S (2008). Some practical guidance for the implementation of propensity score matching. IZA Discussion Paper No. 1588, University of Cologne.

Chimdo A (2001). Research center based maize technology transfer: Efforts and achievements. EARO, Addis Ababa, Ethiopia.

Davis K, Nkonya E, Kato E, Daniel A, Odendo M, Miiro R, Nkuba J (2010). Impact of Farmer Field Schools on Agricultural Productivity and Poverty in East Africa. IFPRI Discussion P. 00992.

Dehejia RH, Wahba S (2002). Propensity score matching methods for non experimental causal studies. Rev. Econ. Stat. 84(1):151-161.

Gockowski J, Asamoah C, David S, kamleu GB, Gyamfi I, Agordorku S, Kuti MA (2006). An evaluation of farmer field school training on the livelihoods of cocoa farmers in Atwima District, Ashanti Region, Ghana. Sustainable Tree Crops Program. Working Papers Series 1. Ibadan, Nigeria: International Institute for Tropical Agriculture.

Gujarati DN (1995). Basic Econometrics. 3rd Edition, McGraw Hill, Inc., New York.

Gujarati DN (2004). Essentials of Econometrics. 4th edition. Mc-Graw-Hill Companies Gilligan D, Hoddinott J. and Alemayehu Seyoum, 2008. The Impact of Ethiopia's Productive Safety Net Programme and its linkages. IFPRI Discussion P. 00839.

Heckman JJ, Ichimura H, Todd P (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme, Rev Econ. Stud. 64:605-654.

Jalan J, Ravallion M (2003). Estimating the benefit incidence of an antipoverty program by propensity score matching. J. Bus. Econ. Stat. 21(1):19-30.

Lee WS (2006). Propensity Score Matching and Variations on the Balancing Test: Melbourne Institute of Applied Economic and Social Research, the University of Melbourne.

Leuven E, Sianesi B (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html.Accessed on January, 2011.

Muhammad L, Njoroge K, Bett C, Mwangi W, Verkuijl H, De Groote H (2003). The seed industry for dryland crops in Eastern Kenya. Mexico, D.F.: CIMMYT and Kenya Agricultural Research Institute (KARI).

Pradhan M, Rawlings LB (2002). The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund, World Bank Econ. Rev. 16(2):275-295.

Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70(1): 41-55.

Anderson JR (2007). Agricultural advisory services: background paper to Science and Technology for pro-poor growth. World development report 2008, ed. D. Byerlee. World Bank, Washington D.C.

Andersson C, Alemu M, Stage S (2009). Impacts of the productive safety net program in Ethiopia on livestock and tree holdings of rural households. Discussion Paper Series. Environment for Development (EfD) Initiative. EfD DP 09-05.

Hujer R, Caliendo M, Thomson SL (2004). New evidence on the effects of job creation schemes in Germany. A matching approach with threefold heterogeneity. Res. Econ. 58(4):257–302.

Appendix

Joint significance test (likelihood ratio test).

Matching algorithms	Sample	Pseudo R^2	LRchi^2	P>chi^2
NN(1)	Unmatched	0.308	74.07	0.000
	Matched	0.007	1.17	1.000
NN(2)	Unmatched	0.308	74.07	0.000
	Matched	0.007	1.17	1.000
NN(3)	Unmatched	0.308	74.07	0.000
	Matched	0.007	1.17	1.000
NN(4)	Unmatched	0.308	74.07	0.000
	Matched	0.007	1.17	1.000
NN(5)	Unmatched	0.308	74.07	0.000
	Matched	0.007	1.17	1.000
Caliper(0.01)	Unmatched	0.308	74.07	0.000
	Matched	0.033	3.67	0.961
Caliper(0.25)	Unmatched	0.308	74.07	0.000
	Matched	0.015	2.61	0.989
Caliper(0.5)	Unmatched	0.308	74.07	0.000
	Matched	0.099	16.99	0.074
Kernel(0.1)	Unmatched	0.308	74.07	0.000
	Matched	0.004	0.67	1.000
Kernel(0.25)	Unmatched	0.308	74.07	0.000
	Matched	0.009	1.60	0.999
Kernel(0.5)	Unmatched	0.308	74.07	0.000
	Matched	0.047	8.09	0.620