Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA ECONOMIA E STATICTICA AGROALIMENTARE

Ciclo XXVI

Settore Concorsuale di afferenza: 13/A2- POLITICA ECONOMICA
Settore Scientifico disciplinare: SECS-P/02 POLITICA ECONOMICA

# THE IMPACT OF GRANTS IN THE AGRICULTURE SECTOR IN ALBANIA: A COUNTERFACTUAL APPROACH

Presentata da: Oltion Preka

Coordinatore Dottorato:
Prof. Roberto Fanfani

Relatore: Prof. Roberto Fanfani

Esame finale anno 2014

Table of Content

# List of Abbreviations

ATE – Average Treatment Effect ( on population )

ATET – Average Treatment Effect on Treated

ATU - Average Treatment Effect on Untreated

AZHBR – Agency for Agriculture and Rural Development

CAP - Common Agriculture Policy

CAP – Common Agriculture Policy

CIA – Conditional Independence Assumption

FAO – Food and Agriculture Organization

K - Kernel matching

MARDA – Ministry of Agriculture and Rural Development in Albania

NN – Nearest Neighbour matching

NSS - National Subsidy Scheme

PSM - Propensity Score Matching

SSAF - Sectoral Strategy of Agriculture and Food

WB – World Bank

# Introduction

In the last decades, evaluation activity in the social science has become increasingly important. A wide range of national, European and international organizations have carried out evaluation programs in order to pursue different objectives. The literature on impact evaluation distinguishes between two prospective of impact evaluation. The first one is related to accountability purposes and the second has to do with research purposes.
This study focuses on the second order of research activity whose principal aim is to explore the causal relationships between action and the effects supposed to be caused by it. Throughout this study we will refer to the term impact as to the one defined by the EU Commission in the Programming Period 2014-2020 which is the following: the change that can be credibly attributed to an intervention ( European Commission 2011)

The evaluation of impacts in the agriculture sector is not an easy task. Due to the particularities characterizing the sector, impact evaluation is considered a real challenge. This in confirmed by Pingal ( 2001) and (Horton) 1986, 1998) affirming that impact evaluation studies have made less of a difference than might be expected and much less than desired by those who conduct them.

As the following sections will outline, it is difficult to establish a clear causal relationship between a program intervention in agriculture and its main expected effects.

Another issue of concern when evaluating agriculture programs is the impact of interest. Most institutions are focused on measuring macro-economic impact of interventions in the agriculture sector. Macro-economic evaluation relies on measuring macro-economic indicators such as GDP.

Impact evaluation is considered crucial from several institutions. The European Union, in particular, in compliance with its objectives, is mainly interested in the convergence between european countries. This implies a major focus on impact evaluation of the macro-economic effects. Differently, the World Bank (WB), the Food and Agriculture Organization (FAO) and other international organizations are more interested in micro impact evaluation of a given activity. This derives from the nature of these organizations, whose institutional objectives aim at improving living conditions for the people living in rural areas by providing sustainment to single programs in a small scale from an institutional and financial perspective. In particular, these institutions WB, FAO, IFC - has long been studying, identifying and analyzing the causes of poverty in rural areas. The principal aim of such research activities has been to find the reasons and underlying causes of poor living conditions and identify the appropriate instruments and mechanisms to bring people out of the poverty line. Little attention has been paid to micro-economic effects of rural support on everyday life of farmers. This is confirmed by the WB that has suggested that despite the wide literature dealing with the impact of agriculture intervention, the scientific rigour remains questionable.

This study seeks to shed light on the effectiveness of Subsidy schemes in Albania. The primary objective is to estimate the impact of the intervention by measuring its effectiveness. It tries to provide an answer to the following question: does the subsidy scheme achieve its main objectives related to increased production and yield? How did it affect the farm size structure? We do this by establishing a causal relationship between intervention and the above outcomes, ruling out all other factors that could influence the result.

There is little evidence about the impact of small scale projects on living conditions in the rural areas and how this impact has been produced. According to the WB, only half the studies that refer to the 'evaluation program' phase present elements of evaluation method, and even less than half adopt rigorous methods such as matching techniques.

Within this context, the scientific importance of this study is twofold. On the one side, it tries to evaluate the micro-economic impact of an agriculture program, when it is acknowledged that data availability on individual farm level is limited (WBcita). On the other one, it represents the first attempt to provide scientific evidence shedding light on agriculture program effects by using rigorous evaluation methods.

According to the best of our knowledge, until now no one has applied rigorous methodological approaches to measure the effectiveness of the national subsidy scheme on production. Therefore, this study has the great ambition to be the first evaluation study to measure the impact of the subsidy schemes in Albania on production output following scientific standard evaluation methodologies. From a more theoretical perspective, it aims to provide further contribution to the growth of empirical evidence on micro-economic effect caused by the agriculture programs. The scarcity of impact evaluations is most pronounced for farm-level interventions seeking to enhance farm use through improved water access and interventions seeking to improve the quality or access to inputs such as seeds or fertilizer technology (IEG, 2011). Furthermore, the understanding of the mechanism of creation of causal impacts remains limited and not explored yet.

We apply a mixed method design in which the typical counterfactual approach is complemented by the theory-based approach. In doing so, after providing evidence of causal impact between program and production, we try to explain the quantitative results found out by the quantitative approach by shedding some light on the mechanism, what has functioned, and what has not, and under what circumstances.

This study is organized as follows. The first section provides and overviews of the agriculture sector in Albania. National subsidy scheme is analysed the second section. After a description of the national framework and of the evaluation of the program in years, we estimate some outcomes that are important for the measurement of the impact on productivity. The third section is of the core of our study. It describes first the main evaluation approaches present in the literature of social programs evaluation. Afterwards, matching approaches is described in general, and propensity score matching in particular. The empirical model used in this study is presented in section four, including also a detailed description of the data set. In addition, the application of the steps of propensity score matching is described and finally is presented the estimates of the program. Conclusions and policy recommendation are given in section five which is concluded by the presentation of several policy recommendations.

## 1. Overview of the Agricultural Sector in Albania

After the Second World War, Albania made part of the communist bloc of the east. Even though it is located in the Western Balkans, just in front of Apulia Region in the other side of the Adriatic Sea, it has been for almost five decades the most isolated country in Europe and one of the most in the World. During all this period of time Albania has a run stated economy where everything belonged to the state and the citizens had no propriety rights. The particular historical moment, market by the fall of the Berlin Wall, touched even Albania. The dictatorship fall and the country adopted the democratic political system and the free market economic system.

Since the beginnings of the 90', the economy of Albania has undergone profound transformations. The economy shifted from a completely stated controlled system under a planned economy, to a private one. This transformation was accompanied by an organic privatization process in all sectors of the economy. It was judged by European Institutions as one the fastest and most successful privatization process in the Eastern Countries. More than half of small to medium state enterprises was changed ownership from state to private hands within the first 5 year after system change. The whole privatization process is considered completed, except for several big state enterprises operating in sectors considered strategic to the national interest of Albania.

Within this scenario, the agriculture sector does not make any difference. It presents similar transformation patterns to those observed in the other sectors. During the communism system, the organization of the agriculture sector was based on two pillars, cooperatives, on one side, and so called Agriculture Enterprises, on the other. In both of them, land ownership was of the state.

More in general, people leaving in the rural areas had no property rights on land, as it was previously nationalised by force. Exception was made only for the small plots, minor than 0.05 hectare, around home.

One of the first economic reforms implemented after the change of system was the Land Reform, approved by the Albanian Parliament in 1991. Unlike other countries in the Eastern Europe, Land Reform did not give back the land to the legitimate proprietary, that is the owner before the communism period. Rather, the Reform distributed the whole agriculture land to those living in rural areas, regardless of previous ownership status. In particular, to all households living in the rural areas was given agriculture land for free, in proportion to the number household members who, at the time the law was approved, were major than 18 years old.

This reform, whose rational was the land goes to whom works it has determined and shaped the agriculture sector till now. Land fragmentation was one of the most important direct consequences which continue till now ( according to the data provided by Ministry of Agriculture and Rural Development of Albania ( MARDA), 350.916 households economies share 695.520 hectare of land. Although it has been registered a slight and slow process of transformation, the average farm size is still very small compared to other Balkan and eastern countries.

State intervention in Albania shows different patterns from those seen above. Despite the fact that the socio-economic importance of the sector is similar, if not even more relevant, compared to other countries, Albania has not a long experience of government intervention in agriculture sector. Albania government was not engaged in supporting programs mainly due to insufficient financial resources. Programs providing direct or indirect support to the farmers, similar to those offered by national governments in Eastern Europe or Europe in general, have been almost completely absent. The only exception is made by 2KR project financing introduced in 2002. Being unique in its category has not prevented the program from being charged for inefficiency due to corruption practices in the allocation of financial resources. For these reasons, indeed, the head of 2KR program has been persecuted by law.

# 2. The National Agricultural Subsidy Scheme

## 2.1. Description of the national subsidy scheme

The national subsidy scheme has been first introduced in Albania in 2007. The scheme represented elements of a general intervention in agriculture sector, in contrast to specific programs aiming at improvements of the living conditions of a specific socio-economic group of rural population. The target group of subsidies is not a specific group. Rather the scheme was directed to the category of farmers, potentially all of them, willing to create or expand the area under cultivation with olives and vineyards by planting at least plots of a minimal dimension equal to 0.3 hectare. The only restrictions were related to the cultures planted - even though the program has later provided financial sustainment to other instruments in favor of including many cultures - and the region. The region restriction is functional to climate conditions favoring the cultivation of those cultures that suit better to regional weather conditions.

Generally speaking, designing a clear picture of what a program aims to achieve, on one hand, and how the intervention is expected to lead to that outcome, on the other, is of crucial importance for assuring a successful evaluation program. Surely, defining the intervention logic of an intervention program is not an easy task, even more in the agriculture sector.

On the other hand, the evaluator must be careful when willing to carry out evaluation on a wide range of outcomes (i.e. poverty alleviation, increasing income, technology adoption ecc.) because this result in a very risky activity from the methodological point of view. The more potential benefits one tries to measure, the greater the risk grows of unreliable results.

In the first year of introduction the scheme supported only 3 measures. In the following years, not only the number of sustained measures has increased, but also the area and the nature of support has increased as well. In a six year period, the number of measures financed passed from X to in the 2007, to X in the 2013. In particular, the National Scheme can be distinguished in the following 20 intervention measures in 2012:

1. Planting of olive trees
2. Planting of nuts
3. Planting of fruit trees
4. Planting of subtropical fruits
5. Planting of vineyards
6. Drop irrigation
7. Direct payment for production of extra virgin olive oil
8. Production of bio agricultural products from cultivated plants.
9. Purchase of new plastic sheet for establishing new greenhouses
10. Purchase of new plastic sheet for tunnels for watermelon
11. Direct payment for livestock farms
12. Support farms that breed
13. Support for matriculated sheep
14. Support for bee hives
15. Support for breeding heifers
16. Subsidizing interest rate for 5 years
17. Collection of chestnuts and pomegranates
18. Intensive snail breeding
19. Supporting the cultivation of the medicinal plants
20. Interest free loans.

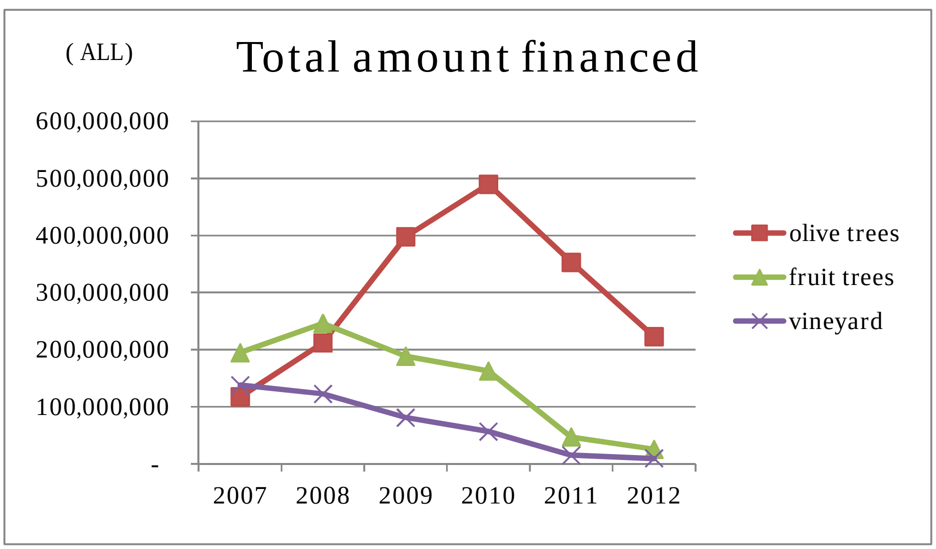
As it can be seen, the National Subsidy Scheme covers a broad range of measures, ranging from sustainment to new plantation of olive and vineyards, to support for purchasing plastic sheet, to direct payments for milk and oil producers and, finally, to financial support for bio farming practices and bee hives. Analyzing the scheme from the nature of support prospective, the sustainment instruments it provides can be classified/fall into in two broad categories, with another third one added only recently, in particular are oriented towards:

- investment support,
- direct payment, and
- credits.

## 2.2. Evaluation of outcomes

From 2007, the total expenditure for these three supporting schemes is about Eur 24 million, with olive scheme accounting for more than 60% of the budget. It is worth noting that the number of projects supported has kept growing, especially for olive scheme, despite decreasing budget. This pattern can be explained by lower investment contribution for each hectare planted.

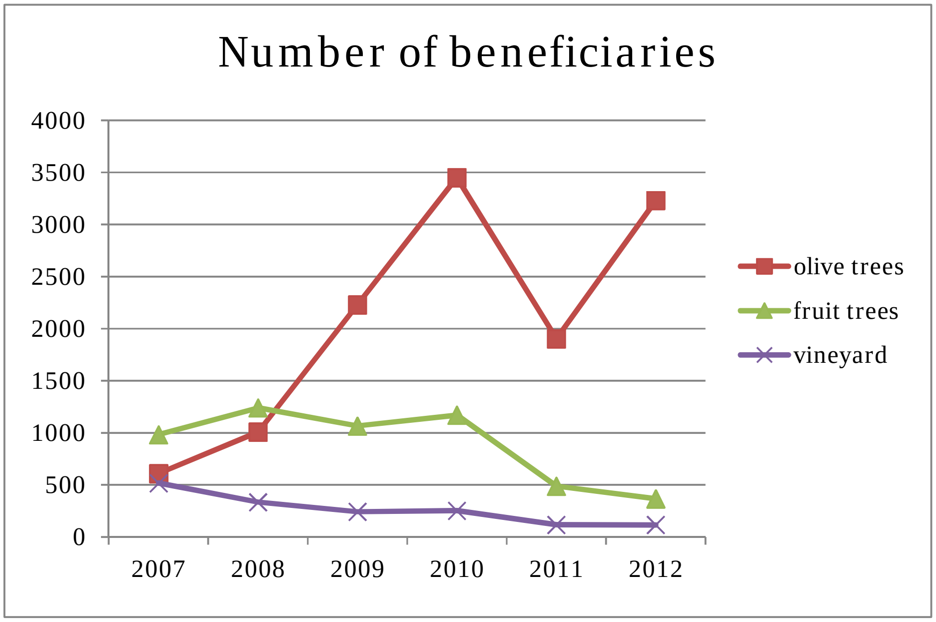
Graph nr 1 – Total financed ( in ALL)




Source: our computation on MARDA data

Graph 2 shows that the largest part of beneficiaries regards the olive plantation with a great distance from fruits and vineyards. The biggest number of beneficiaries for olive trees is in 2010 and their number has followed a growing trend since 2007-2012, compared to the beneficiaries of fruit trees and vineyard that has a descending trend from 2007-2012.

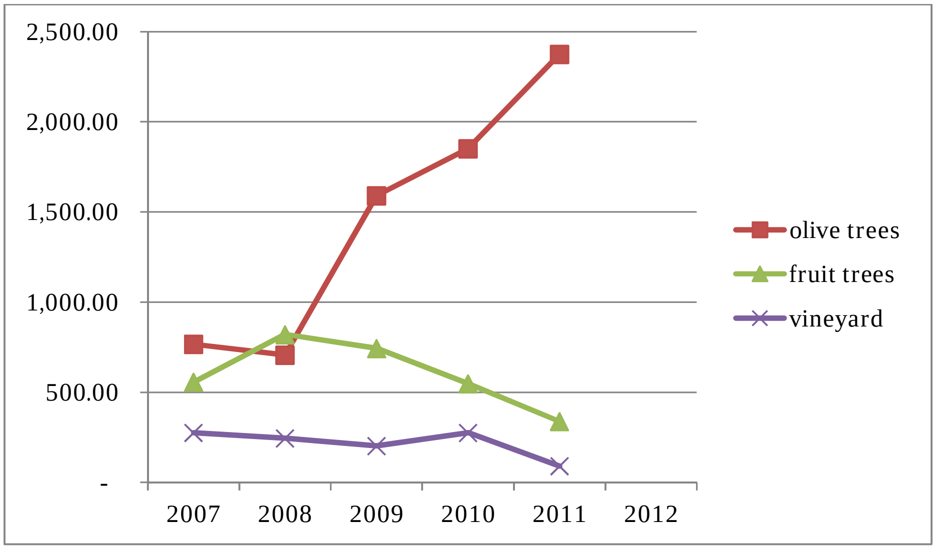
Graph nr 2 – Number of beneficiaries




Source: our computation on MARDA data

The graph nr 3 clearly shows the success of olive plantation scheme. The surface planted of olive trees is the biggest compared to the two other schemes, the trend has increased for olive grove scheme since 2007 to 2012 compared to two other schemes. The trend of fruit trees and vineyard has decreased since 2007 but fruit trees surface is two times higher compared to the surface of vineyard.

Graph nr 3 – New area planted




Source: our computation on MARDA data

Even though the number of measure has come increasing as well as the total budget associated with it, some measures has been preferred to others. If on analyzes the distribution of the budget among measures, together with the number of beneficiaries and, most importantly, the financial support for hectare, it emerges clearly the political decision to favor the plantation of olive groves in particular. Such orientation toward olive plantations is even more clear if one considers the trend of the financial contribution per hectare. While the grant for vineyard and other cultures has fallen sharply, the financial contribution for olives has been reduced at a smaller rate.  Even more if considered in terms of total costs covered

## 2.3. Evaluation of program's leverage effects

The sustainment program, besides the direct effect of cultivation new plantation with olive and vineyards has also triggered private investments for those who received financial support. This is known as the leverage effect. The leverage sample effect caused by the scheme is equal to a rate of 1, 03.  That is, for every

ALL provided by the government to the beneficiaries, these have invested in addition almost the same amount of money. Note that this result is valuable only for our sample, and cannot be generalized. In particular, beneficiaries included in our sample received a financial support from the program of 28’698’500 Lek, or nearly 235 thousand euro[1]. Considered all together, the beneficiaries has invested from own sources of money about 29’619’500 ALL ( 242’782 euro ) in addition to the subsidies received. The sources of personal investment were for more than half own money, intended as coming from previous work, savings and other unspecified sources different for remittances and credits. Nearly one third of the invested amount was financed by remittances of the household head or his son(s) working abroad. The share of investment financed by credit issued by conventional channels such as banks or minor organization dealing with credit sustainment in the rural areas was smaller, reaching something more than 12%. Finally same farmers have made financial unofficial agreements with their relative or friend, but only for a minimum part, only 3 % of the total amount. Regarding financial credit, note that the percentage of those who received is only 3% and 1%, respectively for conventional and non conventional credit channels. This last finding is in line with the general situation, characterized by credit squeeze faced by farmers. It should be noted also that farmers has received credit by means of an instrument of the scheme, which finances credit interests for investment in agriculture sector. In other words, farmers received credit with favorable interest rate of 0%.

Table 1 - Leverage effect of the scheme ( ALL)

| | | |
|---|---|---|
| Total amount spend on subsidies for olive | 28.698.500 | |
| total investments in addition to subsidy received | 29.619.500 | |
| of which: | | % of total |

[1] Exchange rate of 2008 was nearly 1 ALL = 122 EUR

| | | |
|---|---|---|
| own money | 15.873.000 | 53.5 % |
| remittances | 9.246.500 | 31.2% |
| credit | 3.650.000 | 12.3% |
| credit from cousins | 850.0000 | 3% |

Source: own elaboration on AZHBR data

We have also analyzed the leverage effects of subsidies, but this time divided by cultures, in order to see if there is any difference. Since the relative cost of plantation, on the one side, and the financial support, on the other, are different for the same area planted with olive and vineyards, we expect this pattern to be reflected also in the added value of investments. Indeed, table 2 shows that the leverage effect on investments for olive plantation has been below the mean, registering only 64% more investments financed by own resources. In other words, olive farms has spent 0,64 EUR of additional investment to every euro received by the government. This is consistent with our interviews as the famers seem to agree on the fact that the amount financed for each dynym covered completely the cost of plantation.

Regarding the share of each financing source, the so considered private money accounts for more than 64% on the total private investment, followed by remittances which has financed 23% of total private value invested in olives. Credit in the olive sector is almost inexistent providing financing in only in two cases, one of which is between relatives. Only 50 thousand of lek ( 40.000 EUR) was provided by Banks and other institutions financing and one famer has declared that he received 850 thousands of lek from relatives to make a relatively big investment in olive plantations.

Table 2 – Leverage effect of subsidies on olives ( in ALL)

| | | |
|---|---|---|
| Total amount spend on subsidies for olive | 11.216.000 | |
| total investments in addition to subsidy received | 7.213.000 | |
| of which: | | % of total invested |

| | | |
|---|---|---|
| own money | 4.618.000 | 64.3 % |
| remittances | 1.695.000 | 23.4% |
| credit | 50.000 | 0.6% |
| credit from cousins | 850.000 | 11.7% |

Source: own elaboration on AZHBR data

The measure sustaining the plantation of vineyards, though providing less money than the olive measure, both in absolute terms and in terms of covering the investment required, has yielded a leverage rate of 1.28. So vineyards have attracted as much as twice additional investment in Olives, compared to the public money received respectively. Subsidies provided in 2008 for the plantation of vineyards have triggered other private investments in the amount of 130% exceeding public money. Besides, own money and remittances, whose share on the private financing is in line with the average figures, table 3 highlights the sizeable amount of credit issued by financial institutions both, in absolute terms and percentage of the total, respectively 3.600.000 ALL and 16%. Despite the value, the number of cases is also higher than in the olive files, with 3 cases.

Table nr 3 – Leverage effect of subsidies on vineyards ( in ALL)

| | | |
|---|---|---|
| Total spend on Subsidy | 17’382’500 | |
| total invested other than subsidy | 22‘406’500 | |
| of which: | | % of total invested |
| own money | 11’255’000 | 50.3 % |
| remittances | 7’551’500 | 33.7% |
| credit | 3’600’000 | 16 % |
| credit from cousins | 0 | 0% |

Source: own elaboration on AZHBR data

From the analyses carried out above, it has merged that the program has yielded a respectable leverage effect on private investments. Subsidies in the field of olive and vineyards plantations have triggered private investments in the same

amount as the one spent. If this figure could be generalized to the entire population, it is of course a great result.
What is a bit surprising, at least from our point of view, is that the leverage effect was most pronounced for V than for olives. However, these could be explained by less financing amount for vineyards, so who was willing to plant had to spend more by him to cover the gap of total cost needed. But, in the other hand, program designers have adopted different subsidy levels of the final cost to stimulate one culture with respect to the other.

## 3. Evaluation Design and Methodology

Considering the gap and the evaluation need, as well as in compliance with UE recommendations on evolution issues, this study approach combines together quantitative and qualitative methods. The logic underlying such a decision is justified by the need to enter in the thoroughly into the information provided by quantitative result so that to have additional explanations of them It is worth saying that the qualitative part of this theses does not have high ambitions. Its main goal is to shed light into the reasons and mechanism through which the intervention has operated. Understanding something more on qualitative aspects allows us to formulate policy recommendations in order to improve policy intervention in the future not only in the agriculture sector but also even in other economic sectors which face similar problems and thus have the same intervention logic.
A number of approaches can be adopted to evaluate the outcomes and impacts of the government support schemes in the agricultural sector. In designing the appropriate evaluation strategy one needs to choose the methodology that maximise the result of the evaluation in order to produce an “unbiased” impact estimate, given the data available and the purpose of the evaluation.
In this chapter we will briefly review the challenges deriving from the evaluation of agricultural programmes and then briefly review the possible approaches to build a counterfactual in order to perform a sound evaluation of such

programmes. We will finally move to the application, that is, we will present the methodology employed in our impact evaluation of the subsidies received by the agriculture sector in Albania, namely the olive and vineyard subsectors.

### i. *Qualitative analyses*

This study adopts a mixed methodological approach combining quantitative elements with qualitative analyses. The core part of our study is represented by the estimation of the impact of subsides provided to cultivators of olive and vineyards of the production, productivity and farm size as well as on labor employed in agriculture carried out by means of propensity score matching methodology. Once quantitative results have been pointed out, they will be interpreted by qualitative analyses. Thus the role played by the qualitative analysis is mostly to enter thoroughly inside the results in order to provide a larger overviews of the results emerged from quantitative analyses. On this respect, quantitative and qualitative approaches are complementary between them.

In general, qualitative information integrates quantitative data by entering thoroughly into the issues being analyses, providing, in this way, valuable information otherwise not available.

As regards the qualitative analyses conducted in this study in particular, there are a number of issues to be discussed. First, the qualitative analysis draws mainly on two different sources, though of the same nature. The first one comes originates from face to face interviews conducted with head HHs subject to quantitative observation. As it will be better explained later on, nearly 279 face to face interviews have been conducted to gather the data on which the program's impact has been measured. Besides for quantitative nature information, these interviews have been extremely valuable for providing qualitative information not only about the questions included in the questionnaire, specifically on issues impossible to be captured by quantitative information. But also interview resulted extremely valuable in understanding the farmes' problems from more a general prospective. This has allowed creating a completely different perception of interrelated and interdisciplinary nature of problems.

The second source of qualitative information relies on interviews with public administration officials with the purpose of gaining information regarding intervention logic and institutional setting of the national subsidy scheme. In order to do this, public officials working on different levels of administration were consulted. Officials taken into interview, covered institutional position ranging from a combination of central to local level of central and local governing bodies. This has allowed to shedding light on different aspects of the intervention. The interview included:

- Mr. Ndoc Faslia, vice minister of MARDA
- Mr. Gjok Vuksani, Head of AZHBR
- Mr. Drini Imami, who has participated in the program design
- Head of the Agriculture Offices responsible for the Region of Scutari and Fier
- Specialist covering extension services for the Region of Scutari and Fier
- Specialist covering NSS program for the Region of Scutari and Fier
- Specialist making part of the controlling group of implementation of NSS
- Mr. Bardhok Biba, head of the Muninipality of Hajmel.
- Local specialist engaged in agriculture consultation of different Municipalities.

## 3.1. Experimental design

The literature dealing with the establishment of a causal effect relationship between two or more factors relies mainly on the on experimental design. In general, it consists in administering a certain treatment – to a group of units, persons, animals - called “Treated”(T), while another group of individuals, called control or comparison group “Non-Treated” (NT) will not receive such treatment.

The main point here is that units are randomly assigned to either the T group or NT-group. Randomization produces the best estimate of the treatment impact as all characteristics of the observations are equally distributed among the treated and non-treated individuals making the two groups not differ systematically. The overall impact of the treatment can then be measured by the sample average difference in the outcome of the two groups, as such difference could then be solely attributed to the effect of the treatment. This experiments are frequently used in medicine, pharmaceutical sector, and, more in general, in all those fields of study where controlling the attribution of the treatment is possible.

To explain the effect of the randomization more formally, the impact of a treatment $\delta_i$ for each individual $i$ is defined by equation [1] as:

$$\delta_i = Y_{1i} - Y_{0i} \, [1]$$

where $Y_{1i}$ indicates the outcome in case of treatment and $Y_{0i}$ in absence of treatment. Having defined that, the overall purpose of an evaluation exercise is that of measuring the average impact of the treatment across all the individuals, *Average Treatment Effect (ATE)* defined by [2]

$$ATE = E(\delta_i) = E(Y_{1i} - Y_{0i})[2]$$

Where $E(.)$ represents the expected value. The $ATE$ can be rewritten as

$$ATE = E(Y_{1i} - Y_{0i})[2]$$

Equation [2] represents the difference in means between the outcome in case of treatment and the outcome in absence of treatment.

In the experimental design, the random assignment to the treatment guarantees that the treatment status D is uncorrelated with any other observable or unobservable variables, making the potential outcomes $Y_0, Y_1$ statistically independent from the treatment status.

As it is impossible to simultaneously treat and non-treat the same units, it is necessary to have untreated individuals that can proxy the effect of the absence of treatment on the treated group (*Counterfactual*). The $ATE$ could then be estimated as the difference between the mean observed outcomes for the Treated and Non-Treated [3]

$$ATE = E(Y|D = 1) - E(Y|D = 0)[3]$$

Equation [3] is valid if there is no selection bias$SB$, i.e. if the average outcome that treated individuals would obtain in absence of treatment $E(Y_0|D = 1)$equals the observed outcome for the untreated individuals [4]

$$SB = E(Y_0|D = 1) - E(Y_0|D = 0)[4]$$

The selection bias is often not equal to 0 as in many cases those participating to a treatment self-selected or increased the likelihood to receive the treatment thanks to their own characteristics (e.g. if one intends to measure the impact of a training on the future income, may need to consider that people who enrolled in the training demonstrated higher motivation than those who didn't, therefore in the future may also present higher incomes due to the higher devotion to their job, pro-activeness, etc.).

### 3.2. Quasi-Experimental Approach

In the real world experimental design is rare since it is very difficult to properly design and implement an experiment. This results even more difficult in the social sciences since the assignment to treatment is most probably not random, but depends on a variety of reasons and may lead to a number of political and social concerns (think for example to the apriori exclusion of a element from the assignment of public funds).

In the quasi-experimental approach however it is maintained the necessity to measure the impact of the treatment by understanding what would have happened in its absence. Therefore, there is the need to construct a proper *counterfactual*, i.e. a group of individuals replicating the same characteristics of the treated individuals. Following the same reasoning explained before, the impact of a general treatment program could be measured in this case by the differences between the treated and non-treated. But what if the two groups differ between them even in the absence of the treatment due to different characteristics? The differences in the outcomes of each group are to be completely attributed to the treatment intervention, or are due to systematic differences. That is to say, individuals might perform differently even in the complete absence of intervention. Can it be established a clear causal effect between intervention and outcome differences?

In order to overcome this problem, some authors have suggested the use of the quasi-experimental approach. The underlying logic is the same: the need to build an appropriate counterfactual

In this thesis we will focus on the Propensity score matching, a method developed by Paul Rosenbaum and Donald Rubin in 1983 that we will present in the next paragraph.

**POTENTIAL OUTCOME APPROACH OR ROY – RUBIN MODEL** (Roy, 1951; Rubin, 1974)

The so called Roy model, introduced first by Roy in 1951 and then further extended by Rubin in 1974, is a relatively simple model. It relies on only three elements, individuals/units, treatment and potential outcome. In the context of this study, individual either participate or not participate into the program, instead of multi level. Thus treatment is to be considered binary variable and it is denoted by Di. The dummy treatment variable equals one the ith individual receives treatment and zero otherwise.

Potential outcomes are then defined as Yi(D)i, where i denotes each individual for i takes values from 1 to N ( i = 1, …..N) and N denotes the total population. The outcome of individuals who receive treatment will be Yi(1) and for those who do not receive will be Yi(0).

The treatment effect for the ith individual ( for an individual i) will be the difference of between outcome of who received the treatment and who not.

$$= Yi(1) – Yi(0) \quad (1)$$

*Parameter of interest*

The literature suggests two main parameters of interest. The first one is the Average treatment effect on population ( ATE) and the second is average treatment effect on treated (ATET). ATE is the difference of expected outcomes before and after participation. Formally, it is denoted as:

$$tATE = E\ (t) = E[Y(1) – Y(0)] \qquad (2)$$

This parameter answer the question: "What is the expected effect on the outcome if individuals in the population were randomly assigned to treatment? (Caliendo&Copeining, 2008). As Heckman notes, this parameter estimates the effect of all individuals ( whole population), those who participated and those who did not. On the other hand, Average treatment effect on treated (ATET) the second evaluation parameter, focuses explicitly on the effects on those for whom the program is actually intended. It is given by (Caliendo&Copeining) :

$$tATT = E(t|D=1) = E\ [Y(1)|D=1] - E\ [Y(0)|D=1] \qquad (3)$$

Heckam (1997b) suggest other parameters, where the average treatment effect on untreated ( ATU) is the most important.

## 3.3. Matching approach

Estimating treatment effect seems to be an important topic in the literature of empirical economics. There are several methods for estimating treatments effects, but five categories of estimators are widely used, matching, randomized social experiments, natural experiments, selection model and structural simulation model ( Blundell, Dias, 2000), Imbens (2004).

Matching approach popularity might be due to some attractive features. In particular, it is extremely easy to implement if compared to other estimators and most importantly matching does not require consistent nonparametric estimation of unknown functions ( Abadie & Imbens 2006). specification of the functional model. The latter represents a great advantage with respect to other models in all those situations in which there is not enough empirical evidence regarding the structural model to be employed.

The idea underlying matching method is without doubt very simple. For each treated individual it finds (one or more) individuals with the same characteristics of the first one. It does so by collecting sufficient observable factors that any two individuals with the same values of these factors show no systematic differences in the reaction to the treatment ( Blundell, 2002). Afterwards, individuals of one group are matched with individuals of the opposite group who share the same characteristics with respect to the observed covariates. At this point, the impact of the treatment can be measured by the average of differences between matched individuals with different treatment status.

Matching approach tackle the data missing problem differently from the regression method. Whereas the latter recovers the unknown potential outcome using estimated regression model, the former creates the counterfactual situation. Relying in similar individual not participating in the treatment (program), it recovers how the treated would behave, with respect to the outcome, had they not received the treatment. How are created matched couples between treated and non treated is another issue which will be explained later. From the above discussion, it can be deduced that the selection of the right variables that makes individual similar, expect for the treatment status is of crucial importance for estimating the correct effect of the treatment. In first analyses, whether or not the appropriate controlling/matching variables has been chosen is a matter of faith, but remember that it is the same for choice of the right comparison group in diff-in-diff method. On this regard, several experimental studies have tried to estimate the performance of matching estimators

Are we more interested on the overall impact of the program (ATE). This is because the supporting scheme object of analyses is not targeted to a specific group in certain socio-economic conditions or located to a specific region, but it is opened to everyone willing to cultivate more than 0.3 hectares of O and V in compliance/respecting only with the modest requirements. However, we also will estimate ATT in order to have a larger insight on the dimensions of the effects.

## 3.4. The Propensity Score Matching

### 3.4.1 Propensity score

Rosebaum and Rubin ( 1983b) has simplified radically the matching approach by introducing propensity score matching. It has extended the possible application to all those situations in which, due to large number of covariates, matching was not an easy tast. For instance, if one has to match on s covariates of vector X being all dichotomous, there would be 2s possible matches.

Rosebaum and Rubin ( 1983b) demonstrate that if potential outcomes are independent of treatment conditional on covariates X, they are also independent of treatment conditional on a balancing score b(X). As a consequence we will have that the propensity score :

$$P(D=1 \mid X) = P(X).$$

That is, the probability of an individual to participate in a treatment given his observed covariates X, is equal to one possible balancing score. PSM estimator computes /indicates simply the mean difference/difference in means in outcomes over the common support region , appropriately weighted by the propensity score distribution of participants ( Caliendo and Kopeinig ).

Propensity score relies on the following assumptions:

- CIA
- Common support

As any other model, Roy model relies on several assumptions. The first and most important one states that the outcome is conditionally independent on assignment to treatment. The idea behind this assumption is that treatment satisfies some form of erogeneity ( Caliendo Kopeining). In the literature can be found different versions of this assumption. Rosenbaum and Rubin ( 1983b) refer to it as unconfoundedness, Heckman and Robb ( 1985) as selection on observables, whereas Lechner ( 1999) call it conditional independence assumption (CIA), Though with same differences between versions, in this study this terminology will be used interchangeably. In the formal way:

let's denote Yt the outcome for not treated and Yt1 the outcome for non treated.

$$Y(0), Y(1) \text{ VS } D \mid X \qquad (4)$$

it indicates that conditional on covariates X, the outcome and the participation into treatment are independent between them

This expression (4) ) implies that the covariates chosen should be important to explain both the propensity score and the outcome. With specific reference to this study, this means that the variables included in the model should explain the decision to being part of the program as well as the productivity.

Estimation of the pscore is not sufficient to estimate the parameter of interest, ATE or ATT depending on the specific objective of each study ( Becker and Ichino). This is because we know that for continues variables the probability to observe to two units with the same probability is zero. As the propensity score p(X) is fundamentally a continues variables which takes value in the interval [ 0 – 1], the probability two observe to individuals showing the same probability will equal zero as well. Therefore, in order to get through this problem it is needed some other method, precisely a specific algorithm. Several propensity score matching algorithms have been proposed, each one characterized by strengths and pitfalls. This section provides a brief overview of the characteristics of the most important ones as proposed by the empirical evaluation literature, nearest neighbors matching (NN) with various versions of it.

Several versions of this matching algorithm might be implemented. The first is NN with or without replacement. In matching without replacement once a control unit is used to be matched with the correspondent treated unit is discarded.

The idea behind matching with replacement, instead, is to use the same untreated unit more than once if it is a best matching alternative for another treated unit. This is the case when the near a particular treated unit, there are no other better control to match with than the already used unit. Using the same untreated unit several times leads to a loss of estimation precision

NN with caliper - At the end of the process, the average of the differences in the outcome between treated units and untreated are computed, regardless of the way in which best matches are selected. Note that nearest neighbor finds a match to all treated units. This raises doubts about the quality of matching. In the case in which the propensity scores are distributed differently among groups, it might be that the nearest neighbor of a treated unit be anyway a bad match for it.

To solve this problem, Caliper and Radius Matching can be used. Radius matching imposes a caliper, that is the maximum distance of the propensity scores between treated and untreated units for whom the matching is allowed. If the untreated nearest neighbor of a treated units falls within the caliper region, the matching between two individuals will be performed. If it falls outside the established tolerance region, matching will not take place. In this way the researcher can control the quality of matching, that is the extent to which treated and untreated units should be similar in order to be matched. It is clear the quality of matches will be as high as the smaller is the size of the neighborhood.

Kernel matching – Kernel matching, instead, is based on a different approach. Differently from the NN it does not matches singularly treated with untreated unites, but rather all treated are matched with a weighted average of the controls. Note that weights are inversely proportional to the distance between propensity scores of two categories.

Many authors agree on the fact that the perfect algorithm does not exist, since their performance depend on the data available. As ( X) puts it, there is not a perfect estimator for all situations. The effectiveness of each one depends heavily on the data available. The right should be chosen depending on the data

availability and one suite well to the dataset available. There are algorithms that suits better for a specific set of data available, rather than for another one.
Even though asymptotically all matching estimators should yield the same results, the decision on which estimator to choose is important in small samples. Each matching estimator involves a trade – off between precision/effectiveness and variance.

### 3.4.2 Variance estimation

After computation of the treatments effects, testing for their statistical significance and their standard errors represents a challenge apart. The problem lays on the fact that the estimated variance of the treatment effect, besides the classical/normal variance of sampling variation, consist on three more components deriving from the variance due to estimation of the propensity score, the imputation of the common support and, when matching without replacement the order in which the treated units are matched (see Heckman 1998). This entails additional difficulties in computing the variance estimation.
Even though some progress has been made in the recent literature, variance estimation still remains a matter of discussion. Three main approaches for the estimation of standard errors have found wide diffusion in the empirical literature according to ( Kaliendo & Copening). Let us see them briefly.

*Bootstrapping*

The most popular way to deal with the difficulties of variance estimation of treatment effects is the one suggested by Lechner (2002). He suggests using bootstrapping when analytical estimates are biased or unavailable2. There are many application examples for bootstrapping, beginning from Heckman et. al. (1997a) who report bootstrap standard errors for LLM estimators, and continuing with Black and Smith (2004) and Sianesi (2004 ) respectively for nearest neighbor and KM, as well as for caliper matching.
Bootstrapping technique for standard errors estimations consists in reestimantion the results, including the first steps of the estimations (propensity score, common support ecc., for each bootstrap draw. If one repeats bootstrapping R time, R bootstrap samples and R estimated

[2] For a more dettailed discussion about boostrapping methods see Brownstone and Valletta (2001)

average effects will be produced. Distribution of these means is believed to approximate the sampling distribution of the population mean and thus the standard errors.

However, there is little formal evidence to justify bootstrapping method in estimating ate variance ( Imbens 2004). Abadie and Imbens (2006a) show that bootstrap fails in the case of nearest neighbor matching with replacement on a continues covariates.

*Variance estimation by Lechner*

Is the same Lechner who has proposed an alternative way to estimate the atet variance resulting after NN matching applying the following formula:

Var(ˆτ AT T ) = 1/N1 Var(Y (1) | D = 1) + _j∈{D=0}(w j )2/(N1)2 * Var ( Y(0)| D=0)   (X)

where N1 is the number of matched treated individuals and wj is the number of times individual j from the control group has been used. This applies when matching with replacement is performed. Note that this formula corresponds with the "traditional" variance formula if no unit is matched more than once.

Two key assumptions are required, the first that the variances of the outcome variables within treatment and control group satisfy homoschedasticity assumption and second that outcome variances be uncorrelated with the estimated propensity score. Finally, in the simulation made by Lechner, the variance calculated by formula x yields similar results with bootstrap variances.

*Variance estimation by Abadie and Imbens*

In contrast to the above approaches that does not make any distinction between population and sample variance, Abadie and Imbens ( 2006) start from the consideration that average treatment effects (ATE) is different from sample average treatment effect (SATE). Partly because matching estimators with a fixed number of matches are highly nonsmooth functionals of the distribution of the data, and therefore not amenable to standard asymptotic methods for smooth functional, the formal large sample properties of the SATE estimators have not been established (Imbens 2006). As a consequence standard bootstrap based on the assumption of consistent estimating of non parametric unknown function does not lead to valid confidence intervals for the simple matching estimators ( Abadie and Imbend 2005)

They propose a consistent estimator for the variance of ATE that relaxes the above assumption.
Unlike other method which use for σ2(x_w)

## 3.5. Application- The model

### 3.5.1. Data and descriptive statistics

This empirical analysis of the effectiveness of the subsidy scheme relies on cross-sectional dataset. One of the strengths of this stagy lay on the nature/kind of data employed. For this impact evaluation primary data has been collected personally by me administering a survey to 289 farmers belonging to the vineyard and olives sectors. Propensity score matching was applied to the final dataset of 225 observations since more than 60 questionnaires have been dropped due to important missing data.
Generally, data availability is of crucial importance for every scientific methodology. This is even truer when matching procedure is applied since it is a extremely sensitive to the dataset available for the following reasons. The first is that the variables available are critical to justifying the strong assumption underlying propensity score matching ( Maffioli 2010 ) On the other hand, data available determines in large part the appropriate matching algorithm to be applied for the estimation of the program effects.

### 3.5.1 Data collection

#### i. *Sampling methodology*

The data were gathered following a mixed method combining elements of stratification method with randomness one. The gathering process has two steps. The first one involved a stratification based on the region (Remember here that the study focuses on two regions, one in the south and one in the north). Within each region, in the second step, another stratification was adopted distinguishing respondents by the program participation status. After dividing beneficiaries from non beneficiaries, households were selected randomly from a shortlist of members of both categories located in different villages of each region. Within the same village, we tried to find more or less the same number of not-treated selected in a random way as well. As regards randomization, we have picked the first of the list, than the fourth, after that, the seventh and so on. In other words, based on the lists of beneficiaries and non beneficiaries, it has been chosen the first of the list, than the third 4, the seventh and so on till the end of the list members.

### *i. Regions Selection*

NSS is extended throughout Albania. However, for every region only several program components are sustained depending on each Region's agriculture characteristics. Financing for cultivation of olive groves is not provided, for instance, in mountain areas.

Ideally, a representative sample of all regions should have been selected but due to budget constraints it has been decided to focus selection on two regions, Shkodra and Fier. The region of Shkodra is located in the North-east of Albania and Fier, in the South-east part of Albania. The two regions have a long tradition in olive and grapes cultivation, either for direct consumption or for the transformation in olive oil and vine, and thus result two major producers. On the other hand, both regions have received consistent amount of financing within the program framework. Within each region all villages have been covered.

### *ii. Missing Responses*

Particular attention was paid to the non-respondents in order to ensure that the unwillingness or inability to provide an answer was not systematically due to specific characteristics of the respondents.
In other words, as the interviews were carried out in person it was possible to monitor the reasons why the respondents were unable to answer. In this way no connection between the missing answer and any recurrent characteristic of the farmers.
To have the questionnaires compiled properly and fully in all parts multiple-visits were done to the non-respondents, until the task was completed.

### *iii. Timing and periodicity*

In the evaluation of the agricultural programmes it is important to identify the best timing (month, year) to run the interviews. In our case one to one interviews were conducted during the moths of May and the begging of June 2013. The reduced timeframe collection allows for ruling out potential drawback specific to agriculture evaluation, rising from changes in time
Data collected via survey compiled a cross-section referring to the 5-year-after phase of the program implementation. Questions included in the questionnaire referred to the present situation as well as how the situation was 5 years ago, when the program was first introduced. We are aware that, such a choice, at least in theory, involves a great risk of reduced/ objectivity with respect to a question of the past, mostly due to alteration in time of interviews' judgment and mood. But this risk is minimized as questions referring to the past situations, regards mainly asset and other physical aspects, which are not subject to mood change.

### *iv. Survey design and administration*

Data were collected through a questionnaire delivered to 289 farmers.

Prior evaluation research has shown that it is also important for data to be drawn from the same sources for both beneficiaries and non beneficiaries individuals ( Maffioli 2002) This study handles the homogeneity of administering the same interview to the selected sample, comprised of both beneficiaries and non beneficiaries. The interviews were conducted in person by myself as to ensure that all questions were clearly understood and answered by the farmers. Whenever necessary a question was clarified and repeated more times until the the information requested was fully understood. Despite this process has been time-consuming - approximating half an hour per questionnaire - we have ensured through this method a high quality of the responses. On the same time however we have kept in mind that the time frame spent for each interview should be as long as to ensures farmers attention. This has been a challenge, as it was mostly depending on the ability of the interviewer not to lose the farmer's attention and this risk was even higher in the case of respondents from the agricultural sector. For this reason the strategy was to present and explain all questions within certain time-frame and clarify the contribution that this study would give to the planning of future funds.

### 3.5.2 Descriptive statistics

*i. Sample representativeness*

One of the main issues of concern of agriculture in Albania is the relative small farm size compared with neighboring countries. The average farm size varies depending on the culture and the region, with farms concentrated on field crops production and in certain regions being larger than others. According to the statistics provided by MARDA, the region of Fier, together with the region of Vlora, have the largest farms, averaging about 1.3 hectares. Note that since in our model we use the variable farm size with referred to the year 2008, here we are interested to see the distribution of this variable in that year. Overall, the statistics we are going to describe reveal that, if confronted with the data provided by MARDA, our sample is a quite good representation of the whole population. It represents the right proportions of land distribution by the several aspects.

In, particular, the farms part of our sample in 2008 had an average size of 1.63 hectares that is about 20% higher than the population one. But if one looks at the distribution and drops 5% of extreme values, the mean size turns in line with the national one. Moreover, the next table on distribution of farms by dimensions shows that almost 3 on 4 farms of the sample have a size less than 2 hectares which is consistent with the proportion of the whole population.

As regards the distribution of farm size by region, table X shows a slight difference between the regions of Fier and Shkodra at a 90% level of significance. In particular, the farmers located in Fier are on average 23% larger than farmers located in Shkoder. This is in line with the distribution of farm size across regions for Albania, in general, and with these two regions in particular. The average farm size of the sample is representative of the average farm size of the regions.

Table 3 - Average farm size by region

| | Shkodra | Fier | T-test |
|---|---|---|---|
| *Average farm size* | 13.9 | 16.9 | -1.841* |
| *significant at 90 % | | | |

A different pattern is observed in the distribution of farm size by sector. Sample cultivators of olive and vineyard share the same average size. The difference registered is not significant. It should be noted that in this case, the average dimension does not refer to the area under cultivation with olive or vineyard, but rather to the overall land owned by each cultivator.

Table 4 - Average farm size separately for olive and vineyards cultivators

| | Olive | Vineyard | T-test |
|---|---|---|---|
| *Average farm size* | 16.9 | 15.6 | 1.073 |
| *significant at 90 % | | | |

Furthermore, it has been controlled the distribution of the sample according to farm size. For this purpose the sample was split up in four different categories: very small, until 0.6 ha, small, from 0.6 to 1.2, small-medium, between 1.2 and 2, and finally the medium farms including those sized over 2 ha.The criteria for defining the groups dimension arethe same one adopted by MARDA in the presentation of yearbook of agriculture statistics.

Table 5 - Distribution of households by land owned

| Farm size ( in hectares ) | % |
|---|---|
| 0 – 0.59 | 13.78 |
| 0.6 – 1.19 | 22.22 |
| 1.2 – 1.99 | 38.67 |
| 2 + | 25.33 |

Source: our own elaborations on sample data

The prevalence of the farms part of our sample, almost 75%, has a size of less than 2 hectares and only one in four exceeds this threshold. Nearly 13% are very small, not exceeding 0.6 hectares per family, whereas more than 22% is between 0.6 and 1.2 hectares. Farmers owning between 1.2 and 2 hectares of land represent 39% of the sample. Such fragmentation of land should not be surprising considering the land reform of 1991 which distributed the whole agriculture land on the basis of persons leaving in rural areas. Most importantly, such a representation is consistent with the distribution of farm size by dimension in the whole country, according to data provided by MARDA.

*ii. Beneficiaries and not beneficiaries*

As already mentioned in this study we use a choice based sample. Table X displays the distribution of sample household farmers by participation status. There are 55 hh farmers who received subsidies for cultivation of olive groves and 46 for the cultivation of vineyards, for a total of 101 beneficiaries. On the other side, 124 non participants has been included in the sample, divided 60 and 64 respectively between olives and vineyards

Table 6 - Distribution by participations status

| | Non participating | Participating | Both |
|---|---|---|---|
| Olive | 60 | 55 | 115 |
| Vineyard | 64 | 46 | 110 |
| Nr of household | 124 | 101 | 225 |

Matching technique is based on similarities of distribution of covariates between individuals participating in a certain program and individuals not participating. Therefore, one of the questions which arise before applying matching is to what extent are different treated and comparison groups between them. After matching, these systematic differences between groups should completely disappear.

In this study a t-test for two samples with unequal variances has been performed for the covariates believed to be important for both treatment and outcome variables. It allows to check how different are the average characteristic of two groups with respect to all variables considered in the model, and, what's more important Note that t-test can be only used for continues variables, whereas for binary variables we used the percentage of individual presenting that character.

Table X outlines the respective values of t-test and the respective p-values for the two groups. Overall, t-test values reveal that there are not profound differences in means between who decided to participate into the program and who did not, even when there are differences in means. Beneficiaries and non beneficiaries differ with respect to some of characteristics, whereas for others covariates there are not statistically significant differences. The relative homogeneity between observables characteristics we control for is to be

considered a good feature for our sample. This facilitates the process of adjustment between groups.
In particular, the age of beneficiaries is not statistically different from the one of not beneficiaries. In the same way, there is only a slight difference in the average farm size, but not statistically significant. This suggests that land owned has not been determinant for participation into the program. As regards the area cultivated with olive and vineyards in correspondence with the year the subsidy program was introduced, difference in means are not significant. This result could be considered surprising since we were expecting HH who already are engaged in the cultivation of olive and vineyards were more stimulated to increase the area under cultivation compared to the HH who did not have previous experience on it. Since Land share is a combination of land owned variable and land cultivated with olive and vineyard it will not result significantly different as well. The percentage of self employment in the farm does not change between groups. So the hypotheses that who has not other sources of income except land is more likely to participate in programs sustaining the agriculture activity fall down, at least for our sample.
However there are some differences between groups. Beneficiaries and non beneficiaries differ with respect to the average family size, which is statistically different at 99% level. Participation in seminars, field demonstrations and other training programs seems to be correlated receiving subsidy support since the share of beneficiaries being active in other programs is significantly higher than to non beneficiaries. Investment in olive and vineyard plantation results a very good predictor of the participation status. The average of beneficiaries is almost as much as two times the average of non-beneficiaries and the difference is significant at 99% of confidence level. Therefore, bigger family size, higher investments and participation in training programs are positively correlated with the chances to be part of the subsidy programs.

Table 7 - Main characteristics of beneficiaries and non beneficiaries: summary statistics

| Variable | Non Beneficiaries | Beneficiaries | t-test | P – value |
|---|---|---|---|---|
| Age | 54.3 | 57.1 | -1.72 | 0.0745* |
| Nr of members household | 5.2 | 6.1 | -2.10 | 0.0364** |
| farm size in 2008 (ha) | 1.59 | 1.68 | -0.7739 | 0.4399 |
| Land Share | 0.72 | 0.79 | -1 | 0.2232 |
| Additional investment value ('000) | 155.37 | 296.19 | -2.2441 | 0.0263*** |
| Previous training programs | 0.30 | 0.63 | -5.1659 | 0*** |
| Farm self employment | 0.56 | 0.62 | -0.896 | 0.3708 |
| Area cultivated with Olive and Vineyards in 2008 | 0.37 | 0.25 | 1.98 | 0.0561* |

*90% confidence
**95% confidence
*** 99% confidence

We were interested also to check if there were differences between beneficiaries and non beneficiaries according to different farm size levels. In order to do this, farm size is distributed for both groups among four categories, very small, small, medium and upper medium. Table X shows the average values within the single category, as well as the respective t-test values. The mean farms size level does not differ between groups along different dimensions. Interestingly, mean values are exactly the same, with the only exception for farms smaller than 0.6 hectares. But even in this case, it is not significant.

Table 8 - Distribution of beneficiaries by farm size

| Variable | Non Beneficiaries | Beneficiaries | t-test | P – value |
|---|---|---|---|---|
| 0 – 0.59 | 0.29 | 0.44 | -1.9868 | 0.565 |
| 0.6 – 1.19 | 0.97 | 0.97 | -0.1466 | 0.8841 |
| 1.2 – 1.99 | 16.1 | 16.1 | -0.0325 | 0.9742 |
| 2 + | 1.6 | 1.5 | 0.6196 | 0.5372 |

Table 9 - Distribution of 4 categories of farm sizes for each Region ( in ha)

| Dimension | Shkodra | Fier | T-test | P–value |
|---|---|---|---|---|
| 0 – 0.6 | 0.4 | 0.35 | 1.0171 | 0.3209 |
| 06 – 1.2 | 0.9 | 0.9 | -0.7771 | 0.4474 |
| 1.2 – 2 | 1.5 | 1.6 | 0.4682 | 0.1551 |
| 2 - | 2.7 | 2.8 | -0.142 | 0.8887 |
| Overall | 13.5 | 16.6 | -1.8715 | 0.0661 |

Table 10 - Distribution of cultures by farm dimension (unequal variances)

| Farm size | Olive | Vineyard | t-test | P – value |
|---|---|---|---|---|
| 0 – 0.6 | 0.28 | 0.42 | -2.1045 | 0.0446** |
| 06 – 1.2 | 0.97 | 0.98 | -0.3225 | 0.7488 |
| 1.2 – 2 | 1.62 | 1.61 | 0.1449 | 0.8852 |
| 2 - | 2.9 | 2.6 | | |
| Overall | 16.5 | 15.5 | 0.8007 | 0.4242 |

*90% confidence
**95% confidence
*** 99% confidence

Table X displays/ report the distribution of average farm size by culture. To check for differences a t-test with unequal variances has been performed.

The same trend can be also observed when we divide/control by different dimensions of farm size. The only exception is made by the minor size group, up to 0.6 hectares, which shows significant difference between the two cultures at 1% level. Average of farm size for the other groups does not differ between the olives and vineyard. Though showing differences in average size, the same can be conformed even for the upper farm sizes (more than 2 hectare) as most probably the differences observed are due to the variation of the sample.

*Explanatory variables*

The questionnaire is designed to capture as much information as possible regarding not only tangible characteristics but also some intangibles aspects of household farmer. It was designed according to the standard practice in the literature and the input on experts in this sector.

Based on the above considerations, four categories of variables have been considered. The first one gathers information on personal characteristics of the head household (HH) and household (H) including household age, gender, education, main employment and the number of family members of the household, counting only those living home [3]. Educational and main employment is considered in the model as indicator variables, with three alternatives each. Education is divided in mandatory school, high school and university degree, whereas main employment considers employment in the public sectors, in the private one and finally self employment on farm.

The second group of variables considers information regarding farm characteristics. Initial asset endowment of farmers is considered crucial variables by the prevalent literature. Questions were relative to household land ownership, area under cultivation with olive or vineyards, distinguishing between own and ranted area; production; yields; average market price of production. Note that this information was referred to two different moments in time, for 2008 and five years later, 2012

Another group of questions was relative to the investment made in 2008. Precisely, household farmers were asked whether or not they participated into the subsidy scheme and how much financial support received by the program. The amount of investments, other than subsidy, and the financial source were also asked. It has been considered four sources of financement, own money, remittances from abroad, conventional credit issued by institutional organs, and credit by relatives or friends.

[3] Because of the emigration phenomenon many families have members living abroad whom by the statistics result living home.

The third category was designed to capture personal and motivational attitudes of respondents toward labour, in general and training and consultation activity. It included two questions, the first whether households have had any experience with aid, grant, subsidy or other form of instruments falling in the category of sustaining programs before NSS call. Previous credit experience was also included in this question. This is considered important in explaining program participation as who already has experienced dealing with administrative documents is more likely to apply. The second question asks whether they have participated in training and other consultation activities before 2008. Literature dealing with agriculture programs has found positive correlation between the first and successive participations in any kind of sustainment programs. That is to say farmers who already have had similar experiences are more likely to repeat it again.

It has been also included a group questions of a different kind compared with the other ones. They include judgments about NSS, more precisely on information received about the intervention, procedure, application process and so on.

On the other hand, it must be accounted for regional as well as sector differences. These dummy variables result of particular importance for the model since they permit to control for observable and unobservable factors that might differentiate between regions and sectors. Regional dummy not only isolate regional aspects related to whether conditions, climate, soil differences and more in general differences in asset endowment. But also it allows accounting for different institutional settings concerning aspect of designing and application of NSS. In the same way, the role played by the sectoral dummy is to take separate administrative settings of different measures, together with development patterns of olive groves and vineyards. For example, it might be that the implementation of olive scheme has been more successful than the olive scheme for several reasons ranging from farmers preferences, to financed amount, tradition reasons and so on. Introduction of these dummy variables is even more important if considered in the context of lack of specific covariate to account for institutional framework.

For a more detailed information about the covariates considered see ( appendix A )

Analyzing the intervention logic in previous section, it emerged that the main objective of the NSS is to increase the production. Also increasing farm size and yield, as well as improving farming practices, are some of specific objectives. Building on these considerations, effectiveness of the scheme is measured according to three dimensions, each corresponding to a specific objective. As already mentioned, the program entails several other second order objectives, whose measurement result impossible due to data unavailability. Though, this study tackles them from a qualitative appraisal.

Three outcome variables have been adopted in order to measure whether the intervention has achieved its desired results. Given the limited data availability based on a small sample, the sign of the impact, without a clear indication of the magnitude, is to be considered an appreciable finding. Since the intervention is production oriented, production is the most important variable for the objectives of our study. Production, defined simply as overall production, represents another outcome indicator. This variable was chosen to estimate how the program affects production of fruits. Yield is Another outcome indicator is Yield variable is defined as the total production quantity per dynym ( not hectare) and aims at capturing to what extent the program affects crop yield. Since the dataset available includes two different cultures which most probably are characterized by different yields patterns, it is necessary first to parameterize/standardized/to bring these two different cultures to homogenous units. We do this by reporting yields in the average unit for the respective culture. Technically, this is realized by dividing the value of yields originally observed for a HH farmer cultivating olive, by the average yield of olive trees, as results from our sample; the same for hh farmers cultivating vineyards. At the end of such a transformation, yields are measured in units of sample average yield. Differently to the above indicators whose nature is "productive", it has been introduced also the area under cultivation which allows checking for eventual structural effects of the intervention on olive and vineyard plantation size.

### 3.5.2. Estimation of the logistic regression

The estimated coefficients by the logit model cannot be interpreted directly for two main reasons. First, logistic regression was estimated on a choice based sample and thus does not represent the real proportions of beneficiaries and non beneficiaries of the population. Second, these are not the marginal effects of the explaining variables on the dependent variable. Those would have to be calculated separately. However to check if there is any differences in the estimated propensity to participate in the program between who really makes part and who does not, below are displayed the mean estimation of logit first for the sample and then, separately, for beneficiaries and non beneficiaries.
Considering that we are not sure about the best specification of the covariates which determines both outcome and the decision to participate in the program, we provide two different specifications of all observations. The first one (two specification model refers to the whole sample/all observations) is more parsimonious than the second specification.

Estimation of logit for the whole sample:

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| p-score | 225 | .4488889 | .3040083 | .0381506 | 1 |

Estimation of logit for beneficiaries:

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| p-score | 124 | .280419 | .2365027 | .0381506 | .9847584 |

Estimation of logit for non-beneficiaries:

---------------------------------------------------------------------

| | | | | | |
|---|---|---|---|---|---|
| ps \| | 101 | .6557233 | .244587 | .1212776 | 1 |

What is of importance here is the difference in the estimation of the probability to participate into the program on the basis of the initial characteristics for both groups. This is called also the conditional probability to participate in the program. As it can be observed from the above tables, it seems that logit estimates are good predictors of participation. The average conditional probability of non beneficiaries to receive subsidy is 28%, whereas the same estimated probability for beneficiaries is much higher, with an average of 65 %. This means that in 65% of cases, the logistic regression has predicted correctly farmers who practically were sustained by the program. The difference in means of the conditional probabilities is also statistically significant at 99, 99% level, as indicated by the two sample t-test. Note that the propensity score is a probability which takes values in the interval [ 0 - 1]. This makes possible to interpret the estimations as percentage of the total.

Two-sample t test with unequal variances:

| Group \| | Obs | Mean | Std. Err. | Std. Dev. | [99.9% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 \| | 124 | .280419 | .0212386 | .2365027 | .2088155 | .3520224 |
| 1 \| | 101 | .6557233 | .0243373 | .244587 | .5732078 | .7382387 |

t = -11.6188 Pr(|T| > |t|) = 0.0000

The logit model as a whole is highly significant. The likelihood ratio test is significant at a level of 99,99% for both model specifications which indicates the predicting capacity of the model compared to the one without any predictors. Pseudo R-square, which refereed to logit model has a different interpretation compared to the one with regression models, is bigger in the second model. Though, vago, it predicts the relation of independent variables with dependent variable, indicating that the less parsimonious specification of logit has a greater capacity of predicting the relations.

As concerns the interpretation of the logit coefficient, we should be very cautious. The first thing to borne in mind is that we use a choice based sample, most probably with a different representation of the proportions of treated and untreated of the population. Therefore it is suggested to use the odds ratio which account for the above mentioned fact.

Participation in the program is negatively correlated with Age of households, at a significance level of 10%, whereas age squared significant has a slight positive effect on participation. This reveals that the correlation between age and participation status is not linear and that young head householder are more likely to receive support. Other household characteristics, such as the number of family member in particular increase slightly the probability of participation. The level of education does not play any role in obtaining subsidies. All levels of education are not statistically significant. The same, for major employment. It is highly non significant for each alternative of the indicator variable, at levels minor than 10%. Thus it has been decided to exclude it completely form the model. Finally, remember that the dummy variable associated with gender (1 if male and 0 if female) results a perfect predictor of participation, thus it has been excluded by the final specification by logit itself.

Investment in olive and vineyard plantation financed by sources other than the program and the square value of investments are highly significant, with a p-value nearly zero, but these variable have only a minimal influence in the participation status. An excellent predictor of program participation is also the previous experience in consultation activities. Logit estimation reveals at 99.99% confidence that who is active through the participation in consultation supports, is 5 times more likely to receive subsidies compared with who did not. This dummy variable seems to capture motivational and experience characteristics of individuals and, most importantly is confirmed by the theory on aid program participation.

As regards farms characteristics, farm size does not predict the probability to receive aid. More complicated is the interpretation of the other covariate, LFree, which take into account the area already under cultivation with olive and vineyards. Though not significant, their odds ration influence on the participation status is controversial. The lower the Lfree variable, the lower the probability to participate, whereas the odds ratio of Lfree squared is extremely positively high.

## 4.1 Impact Evaluation

As the evaluation aims to estimate the impact of the national support scheme on the performance of the vineyard and olive sectors of the two Albanian regions of Shkoder and Fier. As explained previously, we decided to focus on the three *outcome variables*.

- *yields,*
- (ii)*production increase,*
- (iii) *plantation size.*

Quantitative results of this study seem to be somehow surprising suggesting there is no scientific evidence of impact on expected results of the program. From the methodological point of view, it has emerged that although average treatment effects seems to be high, it is not statistically significative due to large standard errors, measured by the formula provided by Abadie and Imbens ( 2006). However, the data reveal that there is a significant difference ( at 10% level) between groups of beneficiaries and non beneficiaries before matching. After matching such differences disappears, or better are not significant. To put it differently, accounting for selection bias for previous treatment characteristics

Such result suggests that who received financial suppor would have performed better than who did not received support, even in the absence of the program, because of better characteristics.

This result is in line with one the major concerns of the empirical evaluation methods which suggest that participation in the development programs is subject to selection bias.

However it seems that the program has affected positively the area under cultivation with olive and vineyards. All matching estimators show a positive difference in the area cultivated between beneficiaries and non beneficiaries, even though the magnitude of such effect varies among estimators. (Different estimator show different magnitude of change). NNM and NNM with caliper of 0.04 performed with only one neighbor show that the program has caused an increase of the area under cultivation in the order of more than 0.24 hectares with level of significance of 95%. The same result and this the same significance was confirmed by K estimator. Though with a lower magnitude, NNM with (5 neighbours) shows positive correlation between program participation and bigger area. Beneficiaries and non beneficiaries cultivated area differ on average 1.9 hectare. Notice however that this result is significant only at 90%. Indeed, NNM with more one neighbours increase the precision of the estimation at the cost of higher variance and considering the small sample size it is to be expected. But even a level of confidence of 90% should be considered appreciable. It is worth mentioning again that the standard errors were computed by means of Abadie and Imbens formula which are more reliable than bootstrapping standard errors. To confirm the results we have performed NN with boostrap standard errors and the result was even more emphasized. In particular la last estimator shows an average treatment effect of the subsidy program on the area with olive and vineyard of 2.5 significant at 99%.

The average effect on treated might be also of concern. Surprisingly, all estimators report that the influence of the subsidies scheme on beneficiaries, in term of area under cultivation, is less than the effect on population and is no significant at the conventional confidence level.

## 4.2 Drawbacks

Mainly because of the small sample size and the complexity of evaluation of program in agriculture, one must be cautious before interpreting the conclusions and results of this study.

*Unobserved heterogeneity*

As it has been put in evidence several times the model adopted in this study estimates the average treatment effects under the assumption of selection on observables. That is to say, the outcome has some form of exogeneity with respect to the treatment status, once we have controlled by observed covariates. In other word, we are assuming that variables which determine the treatment status as well as the value of the outcome have been identified and isolated. On the other hand, it should be clear that, as suggests, perfect exogeneity is only possible in experimental approaches characterized by an identical distribution of all covariates between groups. This is ensured by the random selection of unit to be treat within the group of untreated. Rosenbaum (2002) suggest that Hidden bias instead would arise in the situation in which there will be one or more unobserved variables correlated simultaneously with treatment status and outcome variable whose performance we are going to measure. There is increasingly awareness in the scientific community about the fact that selection on observables is a strong assumption which hardly holds. Its consistency is threatened by hidden bias between units which affect assignment into treatment. As a result matching estimators will not be robust against this risk. It is clear that the optimal situation to check for selection bias is to have experimental data, but since this is not possible then we address this problem by sensitivity analyses.

*Long-term effects*

This study does not take into consideration long – term effects and it only focuses on short-time effects. The methodology approach adopted neither has the intention, nor is suitable to detecting sustainability effects in the middle to long term.

### *Number of observations*

The limited number of observations constitutes another drawback for this study. Matching is defined as a “data hungry” method ( Mafioli ). In particular the consistency of the matching estimators is dubious as well as the variance will be higher on small samples, putting in doubt the significance of the results. After the validation process of data, from 289 observation gathered only 225 was used for matching. 64 was discarded because incomplete or contradictory data. For example, the area under cultivation was several times than farm size, without justifying this figure by rented land or other. In the same way, farmers stating that large areas with olive or vineyards has yielded 0 production without providing any reason for it, such as diseases or others, was discarded.

### *Differences on maturities cycles*

One of the main problems in measuring yields relates to different crop cycles between olives and grapes. Some fruits require more years from the moment of plantation to being production, with respect to other fruits. This is also true for the full production cycle. Olive and vineyards compared together are very susceptible to this problem. First, olive tree start producing fruits on average four months after plantation and enter in full production in the 8th -9th year of live. Whereas, vineyard reaches full production by the 3th year of plantation and does not increase in the years to come. Such production pattern of the culture makes it difficult to compares yields between trees in different growth state of production. However, researchers, including us, are ancient to know as soon as possible about program results, even with some inaccuracy of the outcome estimations. What is should be important is to take account of such inaccuracy when interpreting the findings of the study.

## 4. Conclusions and Policy Recommendations

Most of the evaluation projects have not adopted any scientific methodology. Often they rely on the difference before and after program implementation. Regarding this issue, the first thing to be borne in mind is that claiming that a certain situation has improved with respect to some indicators is not sufficient to provide scientific evidence of a program impact. In the social sciences, there are many factors affecting a certain outcome. Some of these factors are under the control of the program or at least might be influenced by it; most of them are out of the evaluator's control. In such a situation, the furthest an evaluator can do is to take into account the latter factors that influence the performance of the outcome on which the researcher is interested.

For example, if the performance of an economic indicator outcome - subject to e program/treatment - improves there are two possibilities: the least probable one is that it is the direct result of the program or treatment being evaluated; the other alternative is that it might be that other factors, such as general improved economic situation, have led to this result. This means that the group of individuals who participated in the program would have performed in the same way even in the total absence of the treatment. Therefore, the establishment of causal relationship between the program or treatment and the outcome indicator is not automatic. Rather it has to be proved and this involves very complicated and challenging tasks. The question to which the evaluator should give an answer is: what would have performed the treated group had they not received any program support? There are fundamentally two ways to determine whether or not a certain treatment in general has influenced the outcome. The methodology for evaluation, valid for most of the scientific disciplines can be divided in two broad categories: experimental methods and non-experimental methods. Experimental methods are based on experimental control of the treatment, generally in laboratory conditions. They divide the population in two groups. One group is subject to treatment, whose effects are going to be tested. The other one, called control group, is not treated. The assignment of each unit to the treated or untreated group is made by random. Randomness of assignment is the main distinguishing feature of experimental methods. It is of crucial importance because it ensures that members of both groups present similar characteristics except for treatment. Therefore, the effect of the treatment is computed simply by the average difference of treated and untreated units with respect to the outcome of interest. Experimental trials are widely employed in medicine or pharmacy and, more in general, in those sectors where the assignment to treatment is made randomly.

When it comes to the social field the situation is quite different. It is straightforward that carrying out experimental designs results difficult, if not completely impossible. The main reason lays on the fact that the assignment to treatment is on voluntary bases. Thus it cannot be sustained that is random. Rather, certain individuals under certain personal circumstances will decide to participate in the program. Others, with different characteristics, will decide not to be part of the program. This poses at serious doubt the randomness of assignment to treatment. As a result, experimental designs, with some rare exceptions, are not appropriate for evaluation of social programs. Some alternative method must be used. One of them is the quasi-experimental method. This method of evaluation involves the creation of a counterfactual approach to create a non-treatment scenario as much similar as possible to the treated one.

The importance of matching as a rigorous scientific technical of evaluation has been emphasized even by UE.

## List of References


Aakvik, A. (2001) Bounding a matching estimator: the case of a Norwegian training program. Oxford Bulletin of Economics and Statistics 63(1): 115–143.

Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. Journal of Econometrics 113: 231–263.

Abadie, A. (2005) Semiparametric difference-in-differences estimators. Reviewof Economic Studies 72(1): 1–19.

Abadie, A. and Imbens, G. (2006a) Large sample properties of matching estimators for average treatment effects. Econometrica 74(1): 235–267.

Abadie, A. and Imbens, G. (2006b) On the failure of the bootstrap for matching estimators.Working Paper, Harvard University.

Abadie, A., Drukker, D., Leber Herr, J. and Imbens, G. (2004) Implementing matching estimators for average treatment effects in STATA. Stata Journal 4(3): 290–311.

Angrist, J. and Hahn, J. (2004) When to control for covariates? Panel-asymptotic results for estimates of treatment effects. Review of Economics and Statistics 86(1): 58–72.

Augurzky, B. and Kluve, J. (2007) Assessing the performance of matching algorithms when selection into treatment is strong. Journal of Applied Econometrics 22(3): 533–557.

Augurzky, B. and Schmidt, C. (2001) The propensity score: a means to an end. Discussion Paper No. 271, IZA.

Becker, S.O. (1964) Human Capital. New York: Columbia University Press.

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

Becker, S.O. and Caliendo, M. (2007) Sensitivity analysis for average treatment effect. Stata Journal 7(1): 71–83.

Bergemann, A., Fitzenberger, B. and Speckesser, S. (2001) Evaluating the employment effects of public sector sponsored training in East Germany:

conditional differencein-differences and Ashenfelters' dip. Discussion Paper, University of Mannheim.

Black, D. and Smith, J. (2004) How robust is the evidence on the effects of the collegequality? Evidence from matching. Journal of Econometrics 121(1): 99–124.

Blundell, R. and Costa Dias, M. (2000) Evaluation methods for non-experimental data.Fiscal Studies 21(4): 427–468.

Blundell, R. and Costa Dias, M. (2002) Alternative approaches to evaluation in empiricalmicroeconomics. Portuguese Economic Journal 1: 91–115.

Blundell, R., Dearden, L. and Sianesi, B. (2005) Evaluating the impact of education onearnings in the UK: models, methods and results from the NCDS. Journal of the RoyalStatistical Society, Series A 168(3): 473–512.

Brand, J.E. and Halaby, C.N. (2006) Regression and matching estimates of the effectsof elite college attendance on educational and career achievement. Social ScienceResearch, 35(3): 749–770.

Breiman, L., Friedman, J., Olsen, R. and Stone, C. (1984) Classification and RegressionTrees. Belmont, CA: Wadsworth International Group.

Brodaty, T., Crepon, B. and Fougere, D. (2001) Using matching estimators to evaluatealternative youth employment programs: evidence from France, 1986–1988. In M.

Lechner and F. Pfeiffer (eds), Econometric Evaluation of Labour Market Policies(pp. 85–123). Heidelberg: Physica.

Brownstone, D. and Valletta, R. (2001) The bootstrap and multiple imputations: harnessingincreased computing power for improved statistical tests. Journal of EconomicPerspectives 15(4): 129–141.

Bryson, A. (2002) The union membership wage premium: an analysis using propensityscore matching. Discussion Paper No. 530, Centre for Economic Performance, London.

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

Caliendo, M. and Hujer, R. (2006) The microeconometric estimation of treatment effects– an overview. Allgemeines Statistisches Archiv 90(1): 197–212.

Caliendo, M., Hujer, R. and Thomsen, S. (2007) The employment effects of jobcreation schemes in Germany – a microeconometric evaluation. IZA Discussion PaperNo. 1512. Advances in Econometrics 21, forthcoming.

Cochran, W. (1968) The effectiveness of adjustment by subclassification in removing biasin observational studies. Biometrics 24: 295–314.

Crump, R., Hotz, V., Imbens, G. and Mitnik, O. (2005) Moving the goalposts: addressinglimited overlap in estimation of average treatment effects by changing the estimand.Working Paper, University of California at Berkeley.

Davies, R. and Kim, S. (2003) Matching and the estimated impact of interlisting. DiscussionPaper in Finance No. 2001-11, ISMA Centre, Reading.

Dehejia, R. (2005) Practical propensity score matching: a reply to Smith and Todd. Journalof Econometrics 125: 355–364.

Dehejia, R.H. and Wahba, S. (1999) Causal effects in nonexperimental studies: reevaluatingthe evaluation of training programs. Journal of the American Statistical Association94(448): 1053–1062.

Dehejia, R.H. and Wahba, S. (2002) Propensity score matching methods for nonexperimentalcausal studies. Review of Economics and Statistics 84(1): 151–161.

DiNardo, J. and Tobias, J. (2001) Nonparametric density and regression estimation. Journal of Economic Perspectives 15(4): 11–28.

DiPrete, T. and Gangl, M. (2004) Assessing bias in the estimation of causal effects:Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Sociological Methodology 34: 271–310.

Fredriksson, P. and Johansson, P. (2004) Dynamic treatment assignment – the consequences for evaluations using observational data. Discussion Paper No. 1062, IZA.

Galdo, J. (2004) Evaluating the performance of non-experimental estimators: evidence froma randomized UI program. Working Paper, Centre for Policy Research, Toronto.

Gerfin, M. and Lechner, M. (2002) A microeconometric evaluation of the active labourmarket policy in Switzerland. The Economic Journal 112(482): 854–893.

Greene, W.H. (2003) Econometric Analysis. New York: New York University.

Gu, X.S. and Rosenbaum, P.R. (1993) Comparison of multivariate matching methods:structures, distances and algorithms. Journal of Computational and Graphical Statistics2: 405–420.

Hahn, J. (1998) On the role of the propensity score in efficient semiparametric estimationof average treatment effects. Econometrica 66(2): 315–331.

Ham, J., Li, X. and Reagan, P. (2004) Propensity score matching, a distance-based measureof migration, and the wage growth of young men. Working Paper, Department ofEconomics, Ohio State University.

Heckman, J. (1997) Instrumental variables – a study of the implicit behavioral assumptionsused in making program evaluations. Journal of Human Resources 32(3): 441–462.

Heckman, J. and Robb, R. (1985) Alternative models for evaluating the impact ofinterventions. In J. Heckman and B. Singer (eds), Longitudinal Analysis of LaborMarket Data (pp. 156–245). Cambridge: Cambridge University Press.

Heckman, J. and Smith, J. (1999) The pre-program earnings dip and the determinants ofparticipation in a social program: implications for simple program evaluation strategies.Economic Journal 109(457): 313–348.

Heckman, J. and Todd, P. (2004) A note on adapting propensity score matching andselection models to choice based samples. Working Paper, first draft 1995, this draftNov. 2004, University of Chicago.

Heckman, J., Ichimura, H. and Todd, P. (1997a) Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Review of EconomicStudies 64(4): 605–654.

Heckman, J., Smith, J. and Clements, N. (1997b) Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programmeimpacts. Review of Economic Studies 64(4): 487–535.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998a) Characterizing selection bias using experimental data. Econometrica 66(5): 1017–1098.

Heckman, J., Ichimura, H. and Todd, P. (1998b) Matching as an econometric evaluation estimator. Review of Economic Studies 65(2): 261–294.

Heckman, J., LaLonde, R. and Smith, J. (1999) The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (eds), Handbook of LaborEconomics, (Vol. III, pp. 1865–2097). Amsterdam: Elsevier.

Hirano, K. and Imbens, G. (2002) Estimation of causal effects using propensity score weighting: an application to data on right heart catherization. Health Services andOutcomes Research Methodology 2(3–4): 259–278.

Hirano, K., Imbens, G. and Ridder, G. (2003) Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71(4): 1161–1189.

Hitt, L. and Frei, F. (2002) Do better customers utilize electronic distribution channels? The case of PC banking. Management Science 48(6): 732–748.

Holland, P. (1986) Statistics and causal inference. Journal of the American StatisticalAssociation 81(396): 945–960.

Ichino, A., Mealli, F. and Nannicini, T. (2006) From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity.Discussion Paper No. 2149, IZA, Bonn.

Imbens, G. (2000) The role of the propensity score in estimating dose–response functions. Biometrika 87(3): 706–710.

Imbens, G. (2003) Sensitivity to exogeneity assumptions in program evaluation. AmericanEconomic Review 93(2): 126–132.

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

Lechner, M. (1998) Mikr¨okonometrische Evaluationsstudien: Anmerkungen zu Theorie und Praxis. In F. Pfeiffer and W. Pohlmeier (eds), Qualifikation, Weiterbildung undArbeitsmarkterfolg. ZEW-Wirtschaftsanalysen Band 31. Baden-Baden: Nomos-Verlag.

Lechner, M. (1999) Earnings and employment effects of continuous off-the-job training in East Germany after unification. Journal of Business Economic Statistics 17(1): 74–90.

Lechner, M. (2000) An evaluation of public sector sponsored continuous vocational training programs in East Germany. Journal of Human Resources 35(2): 347–375.

Lechner, M. (2001a) Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M. Lechner and F. Pfeiffer (eds),

Econometric Evaluation of Labour Market Policies (pp. 1–18). Heidelberg: Physica.

Lechner, M. (2001b) A note on the common support problem in applied evaluation studies. Discussion Paper No. 2001-01, University of St Gallen, SIAW.

Lechner, M. (2002) Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. Journal of the Royal Statistical Society, A 165: 59–82.

Lechner, M. (2004) Sequential matching estimation of dynamic causal models. Discussion Paper No. 1042, IZA.

Lechner, M. and Miquel, R. (2005) Identification of the effects of dyanmic treatments by sequential conditional independence assumptions. Working Paper, SIAW.

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

Navarro-Lozano, S. (2002) Matching, selection and the propensity score: evidence from training in Mexico. Working Paper, University of Chicago.

Pagan, A. and Ullah, A. (1999) Nonparametric Econometrics. Cambridge: Cambridge University Press.

Perkins, S.M., Tu, W., Underhill, M.G., Zhou, X. and Murray, M.D. (2000) The useof propensity scores in pharmacoepidemiologic research. Pharmacoepidemiology andDrug Safety 9(2): 93–101.

Rosenbaum, P.R. (2002) Observational Studies. New York: Springer.

Rosenbaum, P. and Rubin, D. (1983a) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal StatisticalSociety, Series B 45: 212–218.

Rosenbaum, P. and Rubin, D. (1983b) The central role of the propensity score in observational studies for causal effects. Biometrika 70(1): 41–50.

Rosenbaum, P. and Rubin, D. (1984) Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association 79:516–524.

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

Roy, A. (1951) Some thoughts on the distribution of earnings. Oxford Economic Papers 3(2): 135–145.

Rubin, D. (1973) The use of matched sampling and regression adjustment to remove biasin observational studies. Biometrics 29(1): 185–203.

Rubin, D. (1974) Estimating causal effects to treatments in randomised and nonrandomised studies. Journal of Educational Psychology 66: 688–701.

Rubin, D. (1979) Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association 74(366):318–328.

Rubin, D. (1980) Comment on Basu, D. – Randomization analysis of experimental data: the Fisher randomization test. Journal of the American Statistical Association 75(371):591–593.

Rubin, D.B. and Thomas, N. (1996) Matching using estimated propensity scores: relating theory to practice. Biometrics 52(1): 249–264.

Sianesi, B. (2001) An evaluation of the active labour market programmes in Sweden. Working Paper No. 2001:5, IFAU – Office of Labour Market Policy Evaluation.

Sianesi, B. (2004) An evaluation of the Swedish system of active labour market programmes in the 1990s. Review of Economics and Statistics 86(1): 133–155.

Silverman, B. (1986) Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

Smith, H. (1997) Matching with multiple controls to estimate treatment effects inobservational studies. Sociological Methodology 27: 325–353.

Smith, J. (2000) A critical survey of empirical methods for evaluating active labor market policies. Schweizerische Zeitschrift fuer Volkswirtschaft und Statistik 136(3): 1–22.

Smith, J. and Todd, P. (2005) Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125(1–2): 305–353.

Wooldridge, J.M. (2004) Econometric Analysis of Cross Section and Panel Data. Boston,MA: Massachusetts Institute of Technology.

Zhao, Z. (2000) Data issues of using matching methods to estimate treatment effects: an illustration with NSW data set. Working Paper, China Centre for Economic Research.

Zhao, Z. (2004) Using matching to estimate treatment effects: data requirements, matching metrics, and Monte Carlo evidence. Review of Economics and Statistics 86(1): 91–107.

## Appendix A – List of variables

| Variable | Label | Description |
|---|---|---|
| S_2008 | | Dummy variable indicating 1 if the individual has received government subsidy and 0 otherwise |
| Age | A1 | Age of the household head when asked in 2013 |
| Age2 | Age2 | Squared age of the household head |
| FamilyMemb | A3 | Number of family member living in the household family |
| Education | A7 | Indicator variable education of the household head divided in 3 categories, 1 – for mandatory school, 2 – high school, 3 – university |
| Farm08 | B3 | Farm size in 2008 ( measured in dynym[4]) |
| Farm12 | B2 | Farm size in 2013 |
| | B6 | Cultived area with Olive or Vineyard in 2008 ( dy) |
| | B61 | Ranted area to be cultivated with Olive and Vineyard in 2008 |
| | B51 | Cultived area with Olive or Vineyard in 2012 ( dy) |
| | B8 | Production in 2008 |
| | B7 | Production in 2012 |
| LShare08 | | Proportion of farm area under cultivation with Olive or Vineyard in 2008 ( dy) |
| LShare08Square | | Proportion of farm area under cultivation with Olive or Vineyard in 2012 ( dy) |
| Qark | Qark | Dummy variable indicating 1 for Shkodra Region |
| Sectordummy | SecD | Dummy variable indicating 1 for vineyard |
| Farm-self employment | H2 | Dummy variable indicating 1 if the household is self employed in farm |
| InvTot | Invtot | The amount of investments from sources other from subsidy |
| InvTotS | InvTotS | The square of the amount of investments from sources other from subsidy |

[4] 1 hectare = 10 dynym

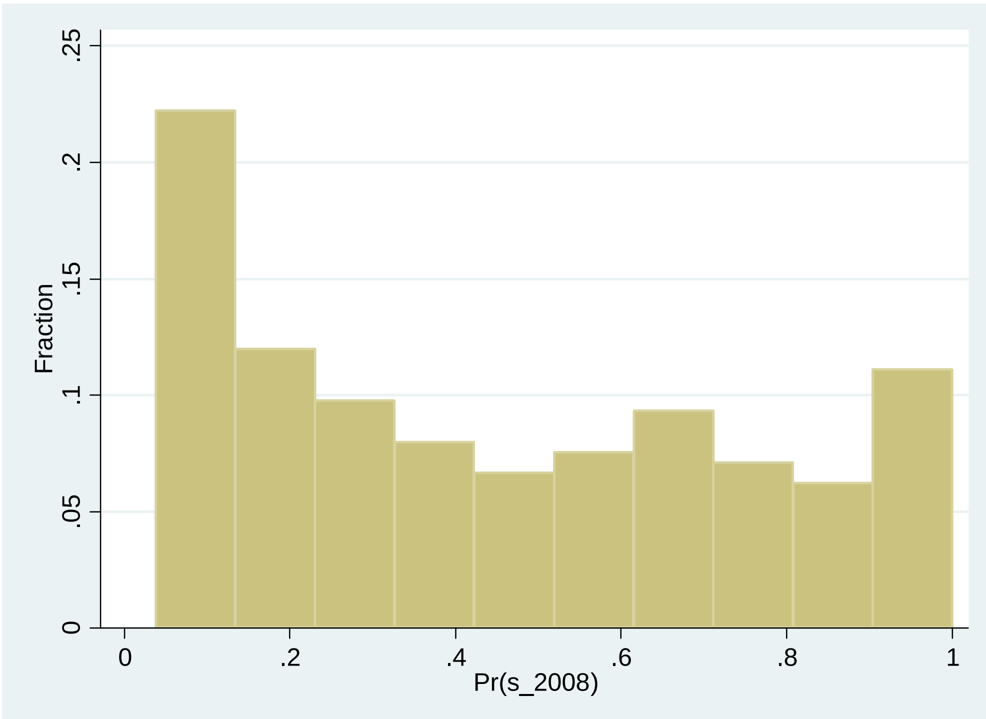
**Appendix B - Logit Results for the Estimation of the Propensity Score**

| Variable | Coeff. | Std. Error | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| *Demographic characteristics* | | | | | | |
| Age Of Hh. H | 0.81 | 0.12 | -1.82 | 0.07 | -0.45 | 0.02 |
| Age Of HH. H Squared | 1.00 | 0.00 | 1.96 | 0.05 | 0.00 | 0.00 |
| Family Members Of Household | 1.09 | 0.06 | 1.48 | 0.14 | -0.03 | 0.21 |
| *Education Level Of Hh. Head* | | | | | | |
| Mandatory School | 0.61 | 0.44 | -1.13 | 0.26 | -1.37 | 0.37 |
| High School | 1.00 | 0.55 | 0.01 | 0.99 | -1.08 | 1.09 |
| Bachelor Degree Or Superior | 0.62 | 0.44 | -1.69 | 0.73 | | |
| *Farm characteristics* | | | | | | |
| Farm Size | 0.10 | 0.02 | -0.59 | 0.56 | -0.06 | 0.03 |
| Percentage Land Free Of Olive And Vineyard (Lfree) | 0.12 | 2.49 | -0.87 | 0.39 | -7.04 | 2.72 |
| Lfree Squared | 23.56 | 2.05 | 1.54 | 0.12 | -0.86 | 7.18 |
| | 1.01 | 0.00 | 4.71 | 0.00 | 0.00 | 0.01 |
| Invtot | | | | | | |
| Invtots | 1.00 | -3.70 | 0.00 | 0.00 | 0.00 | |
| Exp prev prog ( if yes) | 5.06 | 0.36 | 4.49 | 0.00 | 0.91 | 2.33 |
| Region (If Fier) | 0.96 | 0.64 | -0.06 | 0.96 | -1.29 | 1.22 |
| Sector (If Vineyards ) | 0.55 | 0.40 | -1.50 | 0.14 | -1.40 | 0.19 |
| Hh. Head Farm Self-Employed | 1.10 | 0.41 | 0.23 | 0.81 | -0.70 | 0.89 |

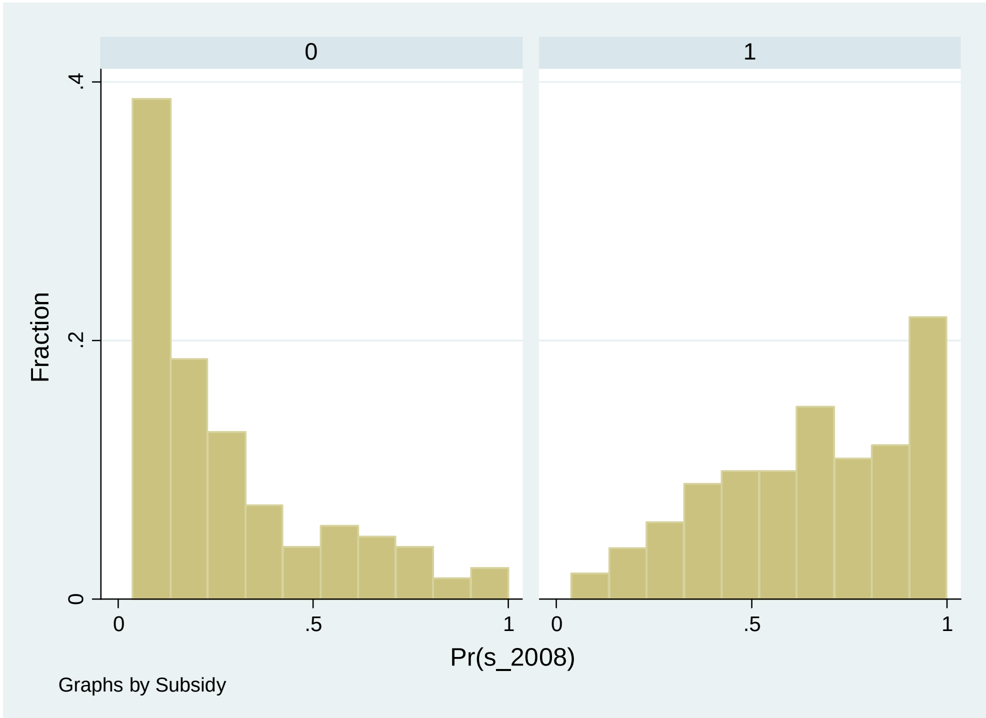
Number of obs. = 225
Log likelihood = -106.658
Pseudo-R2= 0.3109
LR $\chi$^2 Statistic [16 df]/(p-value) = 96.25/ (0.0000)

Note: 0 failures and 1 success completely determined.

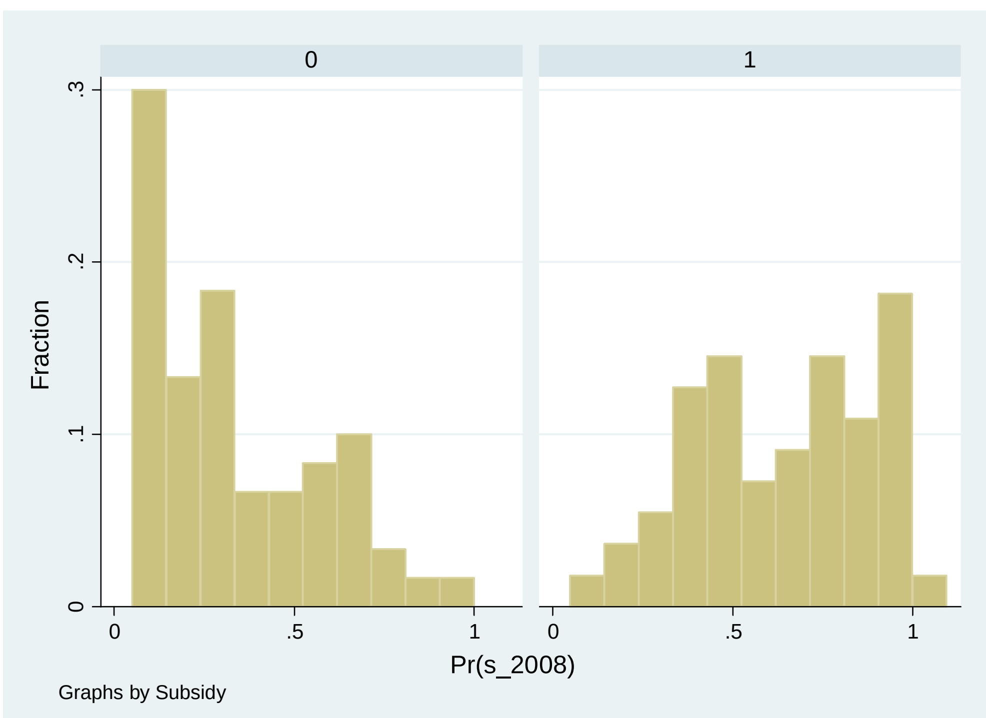
Graph nr 4 – Sample distribution of propensity score matching




Graph nr 5 – Distribution of propensity score matching by participation status

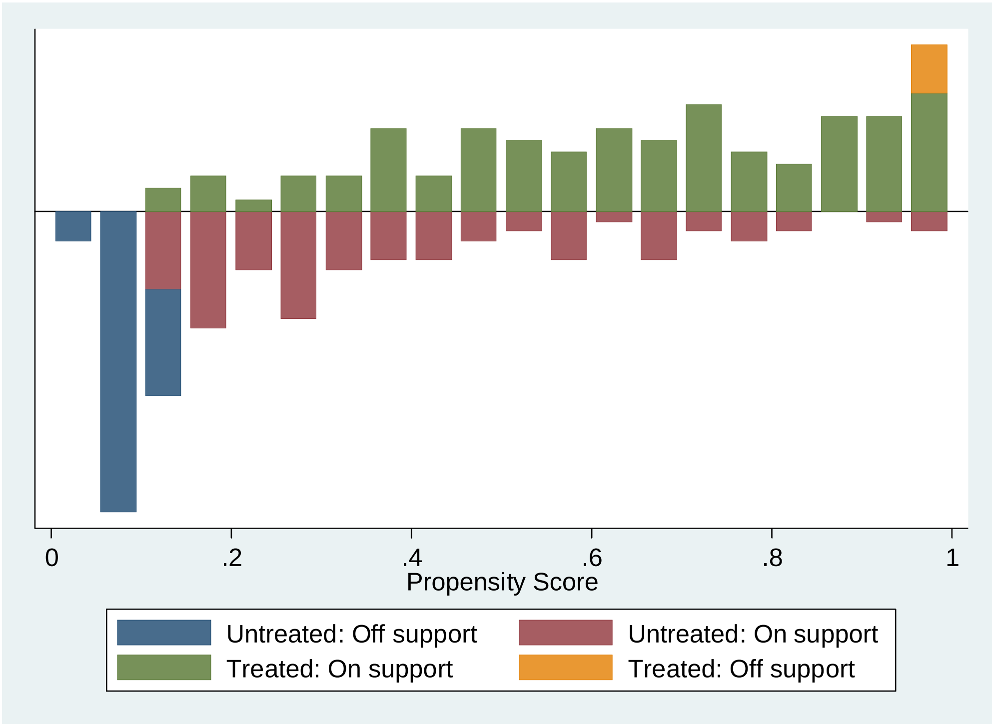



Graph nr 5 – Distribution of propensity score matching by sector



| psmatch2: Treatment assignment | psmatch2: Common support Off suppo | On suppor | Total |
|---|---|---|---|
| Untreated | 45 | 79 | 124 |
| Treated | 4 | 97 | 101 |

Figure 1 – Common support condition



Tabele 10

Matching results for Farm Size

| Variable | Difference Single NN replace | S.E. | T-stat |
| --- | --- | --- | --- |
| Farm size | 3.00 | .78 | 3.85 |
| ATT | 2.22 | 1.76 | 1.26 |
| ATU | 2.87 | 1.08 | 2.66 |
| ATE | 2.51 | 1.26 | 1.98 |

Note: Sample S.E.

(5) N replace

| Farm size | | | |
| --- | --- | --- | --- |
| (ATT) | 1.85 | 1.31 | 1.41 |
| (ATE) | 2.17 | 1.09 | 1.99 |

NN repl cal ( 0.04)

| Farm size | | | |
| --- | --- | --- | --- |
| (ATT) | 2.14 | 1.51 | 1.41 |
| (ATE) | 2.47 | 1.18 | 2.09 |

(5 )N repl cal ( 0.04)

| Farm size | | | |
| --- | --- | --- | --- |
| (ATT) | 1.24 | 1.4 | 0.87 |
| (ATE) | 1.08 | 1.15 | 1.61 |

Kernel (normal)

| Farm size | | | |
| --- | --- | --- | --- |
| (ATT) | 2.13 | 1.59 | 1.34 |
| (ATE) | 2.46 | 1.18 | 2.08 |

Matching results for Farm Size

| Variable | Difference Single NN replace | S.E. | T-stat |
| --- | --- | --- | --- |

| | | | |
|---|---|---|---|
| Production | 3.00 | .78 | 3.85 |
| ATT | 28.5 | 19.75 | 1.44 |
| ATE | 37.8 | 18.22 | 2.08 |

Table X - Ps test before matching ( only different covariates)

| | Mean | | | t-test | |
|---|---|---|---|---|---|
| Variable | Treated | Control | %bias | t | p>\|t\| |
| 3.a7 | .20792 | .14516 | 16.4 | 1.24 | 0.218 |
| 4.a7 | .12871 | .14516 | -4.8 | -0.35 | 0.723 |
| b3 | 16.877 | 15.919 | 10.4 | 0.77 | 0.441 |
| qark | .80198 | .81452 | -3.2 | -0.24 | 0.813 |
| sector_dummy | .45545 | .51613 | - 12.1 | -0.90 | 0.367 |
| LShare08 | .79284 | .72275 | 16.8 | 1.29 | 0.200 |
| farm_selfemployed | .62376 | .56452 | 12.0 | 0.90 | 0.371 |
| 1.qark#1.sector_dummy | .34653 | .43548 | -18.2 | -1.36 | 0.176 |

| Pseudo R2 | LR chi2 | p>chi2 | MeanB | MedB |
|---|---|---|---|---|
| 0.308 | 95.37 | 0.000 | 25.0 | 20.9 |

Figure 2 – Standardized bias before matching

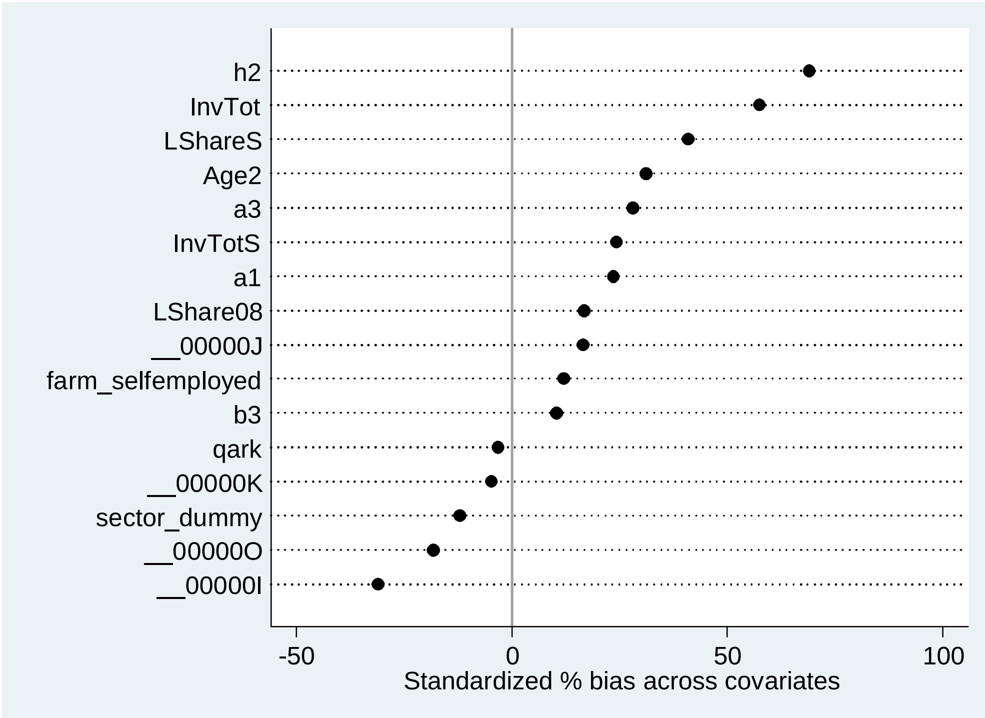



Figure 3 – Standardized bias after matching