COVER LETTER - TITLE PAGE

CAN THE PUBLIC TRAINING SCHOOLS PROGRAMS REDUCE THE TIME NEEDED TO FIND A JOB? A SURVEY FROM SPAIN.

Authors:

José Manuel Cansino Muñoz-Repiso
Universidad de Sevilla
Departamento de Teoría Económica y Economía Política
Facultad de CC. Económicas y Empresariales
Avda. Ramón y Cajal, nº 1
41018 Sevilla.
e-mail: jmcansino@us.es
Tf: + 34 954 55 75 28. Fax : + 34 954 55 76 29

Antonio Sánchez Braza
Universidad de Sevilla
Departamento de Teoría Económica y Economía Política
Facultad de CC. Económicas y Empresariales
Avda. Ramón y Cajal, nº 1
41018 Sevilla.
e-mail: asb@us.es
Tf: + 34 954 55 75 29. Fax : + 34 954 55 76 29

Author responsible for correspondence: Antonio Sánchez Braza: asb@us.es

CAN THE PUBLIC TRAINING SCHOOLS PROGRAMS REDUCE THE TIME NEEDED TO FIND A JOB? A SURVEY FROM SPAIN.

Abstract:

This paper estimes the average effect of a binary treatment on a scalar outcome (time needed to find a job). Such a treatment is the public training schools program implemented in the South of Spain (Seville).

Two methods are developed. The first uses an estimator which weights observations by the inverse of the propensity score. This estimator let us conclude that, for participants, the time needed to find a job is reduced in 471 days.

The second one is the differences estimator, that let us conclude that the time needed to find a job is reduced in 448 days.

Key words: Training programs evaluation, public policies, propensity score, differences estimator.

JEL Code: H30, C10

CAN THE PUBLIC TRAINING SCHOOLS PROGRAMS REDUCE THE TIME NEEDED TO FIND A JOB? A SURVEY FROM SPAIN.

1.- INTRODUCTION[1].

From Heckman, Clements and Smith (1997), it is generally accepted that social programs impact differently on individuals when they differ in characteristics like the previous educational level, professional experience, age, earnings, family background, etc. The individual's addecuate assignement to the set of programs disposable, begins a crucial issue in political decisions.

After controlling by covariates and by knowing the average effect of a program on an appropiatte outcome, the public decisor can decide which program would provoke the best impact on individuals by taking into account the program's average effect on subpopulations previously estimed. Social welfare can improve if public decisors follow an assignment rule which let them to determine which individuals must receive which treatment. Manski (2001) theorized this assuming the case of a finite set of rival treatments. Cansino and Roman (2007) explored this for the Spanish Accounting Court.

The aim of this paper is to estime the average effect of a binary treatment on a scalar outcome. Such a treatment is the training schools program implemented in the South of Spain (Seville) between 1997 and 1999. Speciffically, the paper estimes the average effect of this training program on the time needed to find a job. We select the province of Seville as this is the zone with the most widely developed number of training policies until now. The evaluation is carried out by estimating the program's average effect over the individuals' ability of the sample to find a job, individuals being unemployed between 16 and 25 years old.

Following Hirano, Imbens and Ridder (2003), the paper estimes the average effect by using an estimator which weights observations by the inverse of nonparametric estimates of the propensity score. The differences estimator is also implemented to compare results.

Difficulty in accesing microdata has obstructed in Spain this type of evaluation, largely developed in France, Germany, United Kingdom and USA. This paper contributes to the literature in the sense that no evaluation based of non experimental methods has been applied to this training program in Spain before, apart from Cansino and Sanchez (2008a, 2008b).

[1] Authors thank to Christopher Taber and Francisco Ortega their comments to a previous version of the paper.

The model and program characteristics are described in section 2. Section 3 is destined to explain the database used. In section 4, we define the estimator which weights observations by the inverse of the propensity -Hirano, Imbens and Ridder (2003)- and the results are obtained. The differences estimator is define in section 5 and the results are also obtained. Section 6 concludes.

2.- MODEL CHARACTERISTICS AND DATABASE.

2.1.- THE AVERAGE TREATMENT EFFECT.

The development of public policy evaluation has benefited from the use of causal inference[2]. One of the results is the Potencial Outcome Model –POM-, which allows us to compare participants and non participants in public programs[3]. A prolific development of the POM with regard to training programs evaluation comes thanks to Roy (1951) and Rubin[4] (1974, 1978). This paper support the Roy-Rubin Causal Model (RRM).

In the implementation of POM and RRM, the individual values of the main variables can be extracted from randomized experiments or from observational data. Both types of experiment will notably determine the evaluation and will promote different methodological developments.

In social sciences, randomized experiments face important problems related to cost, moral limitations, attrition and problems derived from the Hawthorne effect -Burtless (1995) and Cameron and Trivedi (2005)-. This can been solved by using observational data. In these cases, Rosenbaum (1999, p. 266) says that the researcher should design a treatment group and a control group from the individuals who have or have not been treated. The objective is to reproduce a scenario which is as similar as possible to a randomized experiment[5].

However, models which include counterfactual events (like an individual participating and not participating at the same time in a training program) are uneffective in individual causal effects

---

[2] Refering to the theoretical approach of casuality and its use in randomized experiments, see Cox (1992). Other authors such as Dawid (1979, 2000), Holland (1986), Heckman (1990) and Pearl (2000) also discuss the meaning of casuality in such an enviroment. Finally, in the specific case of training programs, we have refered to the seminal papers of Rubin (1974) and Heckman and Hotz (1989).

[3] Cameron and Trivedi (2005, pp. 31 and onwards) expose the POM advantages compared to alternative models.

[4] The first references that Rubin considered were Neyman (1923, 1935) and Fisher (1928, 1935).

[5] The seminal papers in this environment were implemented in Medicine. The papers of Cameron and Pauling (1976), Billewicz (1965) and Cochran (1968), must be highlighted. An interesting comment about these seminals papers is contained in Rosenbaum (1995, 1996). Two of the best known papers in quasi-experimental methods are Kiefer (1979) and Bassi (1984).

estimation. Holland (1986) named this problem as the fundamental identification problem. The fundamental identification problem makes us look for second best solutions in which researchers leave the estimation of the individual causal effects, opting for an average effect estimation, which usually refers to the following specification.

To solve the identification problem, we mantein throughtout the paper the unconfounddedness assumption (Rubin, 1978; Rosenbaum and Rubin, 1983), which is also known as the selection on observables assumption (Barnow, Cain and Goldberger, 1980), which asserts that conditional on the observed covariates, the treatment indicator is independent of the potential outcomes.

The Average Treatment Effect of the program[6] (ATE) is addressed in a partial equilibrium environment and, by using the potential outcome notation popularized by Rubin (1974), it is obtained as the average expected value from the difference between the potential values of Y_1 (the case of an individual treated) and Y_0 (the case of a non treated individual)[7]. Implicit in this notation is the stability assumption or SUTVA (Rubin, 1978) that individuals are not affected by receipt of treatment by others, and there is only one version of the treatment. As a consecuence, no general equilibrium effects are considered (Cameron and Trivedi, 2005, p. 872).

The Average Treatment Effect on the Treated (ATET) is given as the average expected value from the difference between the potential values of Y_1 and Y_0 but only with respect to individuals who have received treatment.

Given the fact that the validity of average effects can be damaged if treated and non treated individuals have different characteristics apart from their participation or not in the program, it will be possible to control the differences if they can be observed and the treated and non treated only differ

---

[6] Although in this paper only the most well-known average effects are used, Imbens (2004, p. 4) has summarized all the possible types of average effects of treatment in literature. In this sense, he refers first to the PATE (Population Average Treatment Effect) as the average effect that treatment causes on population, and to the PATT (Population Average Treatment effect for the Treated) as the average effect when only treated are considered. Secondly, the SATE (Selected Average Treatment Effect) would show the average treatment effect when the evaluation is carried out only taking into account only a sample of the population, and the SATT (Selected Average Treatment effect for the Treated) when the sample is extracted only from the treated population. Finally, the CATE (Conditional Average Treatment Effect) would estimate the average effect of the treatment conditioned to the covariates' distribution and the CATT (Conditional Average Treatment effect for the Treated) would estimate the same but only considering the treated population.

[7] The *ATE* is addresed in a partial equilibrium enviorment diferent, for example, to the mega mentioned in Cansino, Cardenete and Roman (2007).

in these characteristics. This is the base of the selection on observables method in which the characteristics are noted as the covariate or vector *X*.

2.2.- THE CARACTERISTICS OF THE PROGRAM EVALUATED

The training school program was designed as a nationwide experimental one implemented by the Spanish Department of Labor (more specifically, by the National Institute of Employment). Considering the first results of the program, the Spanish Department of Labor decided to convert it into a permanent program regulated by the Department Labor's order of march 29st - 1988. Finally, the Department Labor's order of august 3st - 1994 added this program into the set of the national policies of employment until now.

The training school organizes its activities into two steps; the first one gives a theoretical education to the unemployed and the second one offers a professional stage.

In order to judge the interest of this public training program, three parameters have been considered. The first is the number of participants. After the experimental period, the average number of participants had a range of between 45000 and 50000 young unemployed for every year. Compared to the whole of Spain, Seville is the zone with the largest number of implemented projects since 1985. This justifies the geographical focus of the paper.

Secondly, the size of the public funds absorbed as *Table 1* shows.

Thirdly, the EU authorities supported this training program allowing the use of the European Social Fund to finance it.

Table 1 shows the total budget for joint employment actions between the Spanish Department of Labor and the European Social Fund for the period 2000-2006. This allows us to form an idea of the importance that the European Social Fund has as co-financer of the training programs, given that its participation in these joint budget actions is 60'32%.

<table>
<tr><td colspan="2">Joint budget actions between the Spanish Department of Labor and the European Social Fund (2000-2006)</td></tr>
<tr><td colspan="2">TOTAL
9.540.537.406 Euros</td></tr>
<tr><td>***ESF***
5.768.208.916 Euros
60'46 %</td><td>***SDL***
3.772.328.490 Euros
39'54 %</td></tr>
</table>

Table 1
Source: INEM (National Institute of Employment)

2.3.- DEFINITION OF *D* AND *Y* VARIABLES.

We define D as the binary variable which indicates the participation of the individuals in the sample, taking values 1 or 0 depending on if the individual considered participates or not in the program.

$D_i = 1$ will indicate that individual i has participated in one of the programs and $D_i = 0$ will indicate that individual i has not participated in any program.

The scalar Y, is the response variable from which the program's average effects will be evaluated. We define Y as the ability of the individual i to find a job, and shows how much time he has to spend searching for a job[8].

The choice of the response variable is justified because the individuals of the sample, both the participants and the non participants (control group), are initially unemployed and included in the oficial census of people who are searching for jobs. Most of them have not been working before or have a short labor experience because of their age and lack of experience.

For that reason, it is relevant for the program evaluation to consider a response variable which allows us to measure the abilities of these people to find jobs[9].

In this sense, Y_i will measure the "ability of individual i to find a job", and we define Y_i as

$$Y_i = 1 - \frac{number\ of\ consecutive\ days\ until\ the\ individual\ "i"\ find\ a\ job}{total\ duration\ of\ observation} \tag{1}$$

[8] For a further investigation, we could define two outcomes in an alternative way. The first one let us to know the treatment average effect on the individual probability to find a job. The second would be the time needed to find a job conditioned to the unemployments subset (treated ond controls) who have find a job.

[9] The objective of this program is to act as an iniciative of young unemployed minors (less than 25 years old) in finding a job.

The period of observation we have considered consists of three years[10] (1095 days). We started to measure this time from the moment the participants finished the training program (generally at the end of 1999) and january 1st 2000 for the individuals of the control group[11].

The value of Y varies between 0 and 1. If Y is equal to 0, it means that individual i has not found a job during the period considered. This is the worse scenario for the program's effectiveness. If Y is close to 1, the individual i has found a job in a short period of time and if Y is equal to 1, it implies that individual i has found a job the first day after finishing the training program.

Table 2 summarizes the main descriptive data of Y, while *Figure 1* includes the frecuency distribution and the acumulated frecuency distribution of Y.

Descriptive statistics related to response variable Y "ability to find a job"			
Mean	0,578147	Kurtosis	-1,392020
Median	0,773516	Coeffcient of Asymmetry	-0,533554
Mode	0	Minimum	0
Standard deviation	0,380657	Maximum	1
Variance	0,144900	Range	1

Table 2
Source: Own elaboration

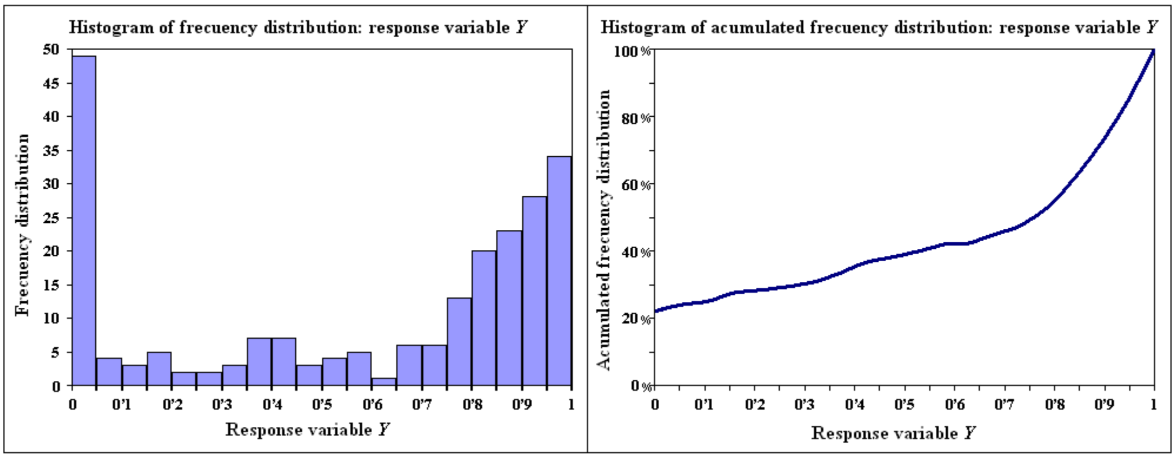

Figure 1
Source: Own elaboration

[10] A relatively broad period of time has been considered, three years (1095 days) due to specific problems of this collective in finding jobs.
[11] The date fixed to start the test for the control group coincides with the starting date for many of the individuals in the participants group.

2.4.- THE MODEL: METHODOLOGY BASED ON IDENTIFICATION AND SELECTION ON OBSERVABLES.

Because of the fact that the validity of average effects can be damaged if participants and controls show characteristics different from their participation or non participation in the training program, these charateristcs must be controlled because of their effect on the values of the response variable.

If the observed characteristics are the only individual (participants and controls) characteristics that differ, we can therefore control these differences. This is the base of the selection on observable model[12].

Selection on observables allows us to isolate the effect of a covariate (or a vector of covariates)[13] maintaining the independence between the treatment indicator variable *D* and the response variable Y. This condition can be expressed as

$$(Y_1 , Y_0) \perp D \mid X \tag{2}$$

Selection on observables supports the independence assumption typical in randomized experiments, contributing to the comparison between participants and controls.

Following Heckman and Hotz (1989, p. 865), selection on observables is recommended when the independence between *D* and *Y* is because of the covariate *X* (or vector of covariates), which has influence on the individual selection process, so by controlling *X* we give a solution to possible biased selection, making the dependency between *D* and *Y* disappear.

In the selection on observable context, when the independence assumption is guaranteed, we considered, according to Dehejia and Wahba (1999, p. 1057), that

$$\mathrm{E}\,[Y_1 - Y_0 \mid X] \;=\; \mathrm{E}\,[\,Y \mid X,\ D = 1\,] - \mathrm{E}\,[\,Y \mid X,\ D = 0\,] \tag{3}$$

Equation 3 lets us express the *ATE* as

$$\begin{aligned} ATE \;&=\; E\left[Y_1 - Y_0\right] \;=\; \int \left(E\left[Y_1 - Y_0 | X\right]\right) d\,P(X) \;= \\ &=\; \int \left(E\left[Y|X,\ D=1\right] - E\left[Y|X,\ D=0\right]\right) d\,P(X) \end{aligned} \tag{4}$$

---

[12] When controlled and treated differ in unobserved characteristics like psicologycal ones, average effect can be estimed by the differences in differences estimator. For the context of this paper, see Cansino and Sanchez (2008).

[13] As an introduction to the framework of the observational methods we recommended the examples that are used by Rosenbaum (1995, pp. 2 and onwards) in his exposition about these methods; this also can by said of Cochran (1968) and Cameron and Pauling (1976). We also recommend the papers of Billewicz (1965) and Moertel et al. (1985), both of them refering to the two previous examples.

In this way[14], it is possible to determine *ATE* from the difference between the average observed value of the response variable of the participants and the controls, calculating the difference for every possible value of *X*.

In a similar way, it is possible to calculate the average effect of the training program only for participants (*ATET*) as:

$$\begin{gathered} ATET = E\left[Y_1 - Y_0 | D=1\right] = \\ \int \left(E\left[Y|X,\ D=1\right] - E\left[Y|X,\ D=0\right]\right) dP(X|D=1) \end{gathered} \tag{5}$$

Therefore, *ATET* will be equal to the difference between the average observed values in the response variable of the participants and the average values of the controls for every different value of *X* when $D=1$.

2.5.- THE COVARIATE VECTOR DEFINITION $X^3 = (X_1, X_2, X_3)$.

With *D* defined, *X* will be a covariate[15] with respect to *D* if, for each of the individuals observed, its values remain the same for each value of *D*. That is to say $X_{0i} = X_{1i}$, being X_{0i} the *X* value before the event *D* ($D = 0$) and X_{1i} the *X* value after happening *D* ($D = 1$).

X covariate is also named contaminant because of the fact that *X* can contamine *Y*, adding its own effects[16] to those provoked by *D*.

The fact that *X* is predetermined with respect to *D* does not imply that this independence is bidirectional, because it is possible that, as a characteristic of considered population, dependence in an opposite direction can appear, making the value of *D* be affected for *X*.

From the sample information included in the database, we consider three predetermined variables which form the vector of covariates $X^3 = (X_1, X_2, X_3)$. The database only allows us to include in the model a complete information about this three covariates. We define the covariates in the following way:

- X_1: sex. This shows if the individual considered is male or female, taking $X_1 = 1$ in the case of a male and $X_1 = 2$ in the case of a female.
- X_2: age. This shows the individual's age at the beginning of the observational period. In the case of participants, X_2 shows the individual's age when the training program is over.

[14] To improve knowledge of selection on observables, we recomend Barnow, Cain and Goldberger (1980).

[15] We talk about one covariate but everything we state can be extrapolated for the case that *X* is a vector of n covariates, as $X^n = (X_1, X_2, \ldots, X_n)$.

[16] To read more, the comments of Rubin (1978) about covariates are very interesting.

For controls, X_2 shows the individual's age as of January 1 st, 2000.

Considering that age range for participants in one of the considered training programs is between 16 and 24 years old, and also considering that the program may extend for 1 or 2 years, X_2 covariate will have values of between 17 and 26 years old.

- $\boldsymbol{X_3}$: zone, showing the city where individuals took the training program or, in the case of controls, where individuals lived. For this last variable, we have divided the area of Seville (Spain) into four zones, named as zone 1 (Sevilla city), zone 2 (east and northeast of Seville), zone 3 (south and southwest) and zone 4 (west and northwest). The criterion of mapping is an operational one. So the $\boldsymbol{X_3}$ covariate will take the following values: $\boldsymbol{X_3} = 1$ for zone 1, $\boldsymbol{X_3} = 2$ for zone 2, $\boldsymbol{X_3} = 3$ for zone 3 and $\boldsymbol{X_3} = 4$ for zone 4.

Figure 2 includes the frecuency distribution of the covariates $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$.

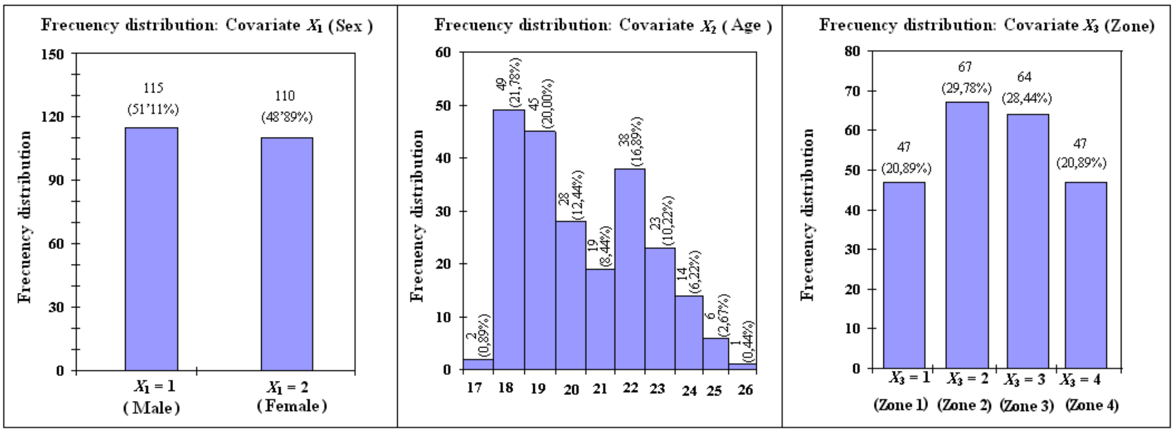

Figure 2
Source: Own elaboration

The inclusion in the regressions of each predetermined variable, depending on each variable's characteristics, is done with the following procedure:

- $\boldsymbol{X_1}$ is a qualitative variable and will be made up by dummy variables. To prevent perfect multicolineality, we introduce in the model as many dummies as categories less one.

  So in order to consider $\boldsymbol{X_1}$, we will include the dummy variable X_{11} which can value 0 and 1.

$$X_{11} = \left| \begin{array}{ll} 1 & \textit{in the case of a male} \\ 0 & \textit{in the case of a female} \end{array} \right.$$

- $\boldsymbol{X_2}$ is a quantitative variable taking values of between 17 and 26 years old. In this case, no dummies are necessary.

- Considering $\boldsymbol{X_3}$ as a qualitative variable, we need to define three dummies noted as $\boldsymbol{X_{31}}$, $\boldsymbol{X_{32}}$ and $\boldsymbol{X_{33}}$, taking values 0 or 1.

$$X_{31} = \left| \begin{array}{ll} 1 & \textit{if the individual belongs to zone } 1 \\ 0 & \textit{if the individual belongs to any of the other three zones} \end{array} \right.$$

$$X_{32} = \left| \begin{array}{ll} 1 & \textit{if the individual belongs to zone } 2 \\ 0 & \textit{if the individual belongs to any of the other three zones} \end{array} \right.$$

$$X_{33} = \left| \begin{array}{ll} 1 & \textit{if the individual belongs to zone } 3 \\ 0 & \textit{if the individual belongs to any of the other three zones} \end{array} \right.$$

3.- The "BASEVAFOR 96-03" database.

The "BASEVAFOR 96-03" database[17] has been constructed from individuals who have participated in the training programs carried out in the south of Spain (Seville) between 1996 and 2003[18]. We have selected those individuals who have finished the training program in 1999, the last year in which information was available when we started the evaluation[19]. Depending on the length of the program (1 or 2 years), programs finishing in 1999 started in either 1997 or 1998.

Only individuals who finished the training program have been considered, rejecting those who left the program before the end. The total of individuals was 1528, and from that figure we have selected a sample of 150. The selected individuals make up the participant group in our investigation. Similarly we have selected 75 individuals[20] to be the control group. The controls have similar characteristics to the participants[21].

The "BASEVAFOR 96-03" incudes two types of data related not only to participants but also with the controls. Firstly, "BASEVAFOR 96-03"gives us information about the periods of employment and

[17] The individual data came from the oficial employment agency (INEM).

[18] This time limit is due to the fact that it includes information related to the training programs that finished in 1999, meaning the program started in either 1997 or 1998. In some cases additional information of the individuals has been refered to from 1996. On other hand, as we refer to programs finished during 1999, the information relative to the response variable extends until 2002, to which some information of the individuals of 2003 is added.

[19] The requirements of the individuals were to be between the ages of 16 and 25 years, be unemployed and have signed up to the unemployment office, and have the minimum requirements to begin a training contract.

[20] We maintain the ratio 2/1 which was used in the evaluation of the JTPA study. The National Job Training Partnership Act (JTPA) Study was commissioned by the U.S. Department of Labor in 1986 to measure the benefits and cost of selected employment and training programs for economically disadvantaged adults and out-of-school youths. See Heckman, Ichimura and Todd (1997, 1998) and Heckman, Ichimura, Smith and Todd (1998). The dimension of the database is comparable to similar evaluations included in Deheija and Wahba (1999, p. 1056-Table 2-). The control group was constructed by the regional labor authorities for our evaluation.

For data, we assume some of them are missing at random (Rubin, 1976; Little and Rubin, 1987).

[21] They are individuals who, during the period of time considered, show the same characteristics of participants and have the requirements to participate in the training program. Also, it is possible that some of them could have applied for a program and could not participate due to place limitation.

unemployment of individuals, including data related to the number of times the individual has applied for a job. We will use this type of information to construct Y. Secondly, "BASEVAFOR 96-03" contains information related to the covariates considered; sex, age and resident zone.

4.- ESTIMATION OF THE AVERAGE EFFECT: EMPIRICAL RESULTS.

4.1. The propensity score.

To avoid the need to match individuals on the values of all covarities, Rosenbaum and Rubin (1983, 1984) develop an approach based on the propensity score[22], the probability of one individual to participate in a program (probability of $D = 1$), conditioned to the values of vector X. By making this probability $\varepsilon(X)$, we can express this as:

$$\varepsilon(X) = P(D=1 \mid X) \tag{6}$$

which is assumed to be bounded away from zero and one.

This shows that this probability is a function of X, which is usually unknown, and therefore it should be estimated by using the database.

Rosenbaum and Rubin (1983) also define the assumption of propensity score independence as

$$(Y_1, Y_0) \perp D \mid \varepsilon(X) \tag{7}$$

ε (X) being the probability of participating in a program conditioned on X. In this way, the independence assumption typical in randomized experiments is guaranteed. This assumption lets us argue that all of the observations with the same propensity score will have the same distribution as the vector X, which means that we can compare the data observed for either participants or controls with the same propensity score.

Following Hahn (1998, p. 316), the calculation of the conditioned probability of participation in a program, given certain observable characteristics, plays a crucial role in controlling bias in order to obtain an estimator of the program's effects.

By using propensity score, we proceed as if it were the case of an unidimensional variable, improving evaluation efficiency by avoiding the management of a large number of covariates included in vector X.

The way to estimate the effects of a training program using selection on observables and by applying propensity score, is divided into "two-stage".

[22] Really, there is not exist consensus on the number of covariates which recommended the use of propensity score instead of the covariates vector (Imbens, 2004). Any case, this is recommended when the overlap assumption can not be guaranted for all the covariates.

4.2. The calculation of propensity score on vector covariates $X^3 = (X_1, X_2, X_3)$ (sex, age and zone). From (6), we can now express the probability of an individual's participation in a program conditioned on the value of vector X, as

$$\varepsilon(X) = P(D=1 \mid X) = F(\beta X) \tag{8}$$

β is the parameter's vector associated with the covariates. The value of this probability will remain conditioned to the value of the distribution function at point βX_j ; X_j being every possible value that can adopt the vector of covariates X, with $j = 1, \ldots, k$.

Depending on the specific function of F, different selection models of binary response could be specified. From the possible non linear options, we have selected three: the Probit Model, the Logit Model and the Extreme Value Model Type I. There is not a generally accepted selection criterion in choosing one of these three models for the estimation of the "propensity score", so the way in which the choice is made is due only to practical reasons. We will estimate the three models and after analysing the obtained results, we will choose the one with the best results according to the criteria specified later on. Carrying out regressions on the vector of covariates $X^3 = (X_1, X_2, X_3)$, the results obtained are contained in *Tables 3*, *4 and 5*.

The calculation of the "propensity score" by using the Probit Model					
Dependent Variable: D (Prob. $D = 1$)					
Method: ML - Binary Probit					
Variable	**Coefficient**	**Coefficient Value**	**Std. Error**	**t -Statistic***	**Prob.**
Fixed effect	μ	-2'366701	0'848762	-2'788414	0'0053
$\boldsymbol{X_{11}}$	β_{11}	0'580077	0'185036	3'134942	0'0017
$\boldsymbol{X_2}$	β_2	0'125535	0'040936	3'066652	0'0022
$\boldsymbol{X_{31}}$	β_{31}	-0'254362	0'285963	-0'889491	0'3737
$\boldsymbol{X_{32}}$	β_{32}	0'002826	0'254795	0'011091	0'9912
$\boldsymbol{X_{33}}$	β_{33}	0'058707	0'257340	0'228131	0'8195
Mean dependent var		0'666667	S.D. dependent var		0'472456
S.E. of regression		0'460830	Akaike info criterion		1'249478
Sum squared resid		46'50776	Schwarz criterion		1'340574
Log likelihood		-134'5663	Hannan-Quinn criter.		1'286245
Restr. log likelihood		-143'2157	Avg. log likelihood		-0'598072
LR statistic (5 df)		17'29884	McFadden R-squared		0'060394
Probability (LR stat)		0'003967			
* t - Statistics adjusted by White's method					

Table 3
Source: Own elaboration

The calculation of the "propensity score" by using the Logic Model					
Dependent Variable: *D* (Prob. *D* = 1)					
Method: ML - Binary Logic					
Variable	**Coefficient**	**Coefficient Value**	**Std. Error**	**t -Statistic***	**Prob.**
Fixed effect	μ	-3'801614	1'384561	-2'745719	0'0060
X_{11}	β_{11}	0'959225	0'311150	3'082833	0'0021
X_2	β_2	0'201706	0'067164	3'003174	0'0027
X_{31}	β_{31}	-0'428509	0'477954	-0'896549	0'3700
X_{32}	β_{32}	0'007757	0'422600	0'018354	0'9854
X_{33}	β_{33}	0'090384	0'427989	0'211182	0'8327
Mean dependent var		0'666667	S.D. dependent var		0'472456
S.E. of regression		0'460893	Akaike info criterion		1'250555
Sum squared resid		46'52056	Schwarz criterion		1'341651
Log likelihood		-134'6875	Hannan-Quinn criter.		1'287322
Restr. log likelihood		-143'2157	Avg. log likelihood		-0'598611
LR statistic (5 df)		17'05642	McFadden R-squared		0'059548
Probability (LR stat)		0'004394			
* t - Statistics adjusted by White's method					

Table 4

Source: Own elaboration

The calculation of the "propensity score" by using the Extreme Value Model Type I					
Dependent Variable: *D* (Prob. *D* = 1)					
Method: ML - Binary Extreme Value Type I (Gompit function)					
Variable	**Coefficient**	**Coefficient Value**	**Std. Error**	**t -Statistic***	**Prob.**
Fixed effect	μ	-2'563204	1'079534	-2'374361	0'0176
X_{11}	β_{11}	0'773186	0'250257	3'089566	0'0020
X_2	β_2	0'156557	0'054409	2'877383	0'0040
X_{31}	β_{31}	-0'354458	0'376021	-0'942655	0'3459
X_{32}	β_{32}	-0'009549	0'331510	-0'028805	0'9770
X_{33}	β_{33}	0'048197	0'340991	0'141344	0'8876
Mean dependent var		0'666667	S.D. dependent var		0'472456
S.E. of regression		0'461152	Akaike info criterion		1'252412
Sum squared resid		46'57280	Schwarz criterion		1'343508
Log likelihood		-134'8964	Hannan-Quinn criter.		1'289179
Restr. log likelihood		-143'2157	Avg. log likelihood		-0'599539
LR statistic (5 df)		16'63867	McFadden R-squared		0'058090
Probability (LR stat)		0'005239			
* t - Statistics adjusted by White's method					

Table 5

Source: Own elaboration

In every case, in order to avoid possible heteroskedasticity problems, the t-statistic values are adjusted by White's method.

According to the corresponding value estimations of the parameters and t-statistics contained in *Tables 3*, *4 and 5*, the variables X_{11} and X_2 appear to be significant when calculating the probability of participation in the three models considered, while the dummy variables X_{31}, X_{32} and X_{33}, defined to include the zone covariate, appear insignificant. In any case, we have decided to mantain them because they help to improve the significance of all the estimated parameters and to improve goodness of fits.

From the three methods, the most efficient will be the one that shows less values of information criterion of Akaike, Schwarz and Hannan-Quinn and a higher value of the log likelihood function for each of the three models. This information appears in *Table 6*.

Comparison of the obtained results from the three binary response models applied			
	Probit Model	Logit Model	Extreme Value Model
Log likelihood function	-134’5663	-134’6875	-134’8964
Criterion Akaike	1’249478	1’250555	1’252412
Criterion Schwarz	1’340574	1’341651	1’343508
Criterion Hannan-Quinn	1’286245	1’287322	1’289179

Table 6

Source: Own elaboration

The model that shows the lowest values of the criterion of Akaike, Schwarz and Hannan-Quinn and a the highest value of the verosimilitude function is the Probit Model, and for this reason, it has been selected.

Following the procedure of the Probit Model, the equation which reflects the participation probability of an individual of the sample in the evaluated program conditioned on the of vector values is as follows:

$$P = -2'366701 + 0'580077\, X_{11} + 0'125535\, X_2 - 0'254362\, X_{31} + \\ + 0'002826\, X_{32} + 0'058707\, X_{33}$$

Table 7 contains the main data of descriptive statistics related to the probability of participation. The probability has been estimated for every individual of the sample by using the Probit Model. *Table 8* contains the estimation of the participation probability for every possible individual depending on the different values that vector $X^3 = (X_1, X_2, X_3)$ can have.

Descriptive statistics related to "Propensity score" obtained by using the Probit Model			
Mean	0'667158	Kurtosis	-0'795769
Median	0'681896	Coeffcient of Asymmetry	-0'145460
Mode	0'682903	Minimum	0'358889
Standard deviation	0'129550	Maximum	0'920798
Variance	0'016783	Range	0'561910

Table 7
Source: Own elaboration

"Propensity score" values depending on possible values that vector $X^3 = (X_1, X_2, X_3)$ can have, according to the Probit Model

	Male				Female			
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 1	Zone 2	Zone 3	Zone 4
17 years old	0'537092	0'636943	0'657695	0'635882	0'313141	0'409132	0'430973	0'408034
18 years old	0'586537	0'682903	0'702538	0'681896	0'358889	0'458488	0'480714	0'457367
19 years old	0'634645	0'726203	0'74449	0'725261	0'406757	0'508494	0'530757	0'507366
20 years old	0'680721	0'766358	0'783123	0'765491	0'456062	0'558366	0'580319	0'55725
21 years old	0'724162	0'803015	0'818146	0'80223	0'506053	0'607328	0'628636	0'606241
22 years old	0'76448	0'835956	0'849398	0'835257	0'55595	0'654647	0'675006	0'653605
23 years old	0'801313	0'865097	0'876852	0'864483	0'604975	0'699663	0'718811	0'69868
24 years old	0'834439	0'890472	0'900591	0'889942	0'65239	0'74182	0'759548	0'740906
25 years old	0'863765	0'912224	0'920798	0'911773	0'697532	0'780683	0'796839	0'779847
26 years old	0'889321	0'930579	0'93773	0'930201	0'739839	0'81595	0'830443	0'815197

Table 8
Source: Own elaboration

Finally, to every individual (participants and controls) the estimated value of his "propensity score" conditioned on vector X^3 has been assigned. After doing this we proceed to calculate *ATE* ($\hat{\alpha}_{ATE}$) and *ATET* ($\hat{\alpha}_{ATET}$).

4.3. Weighting observations by the inverse of the propensity score.

By weighting observations by the inverse of a nonparametric estime of the propensity score, we have an efficient estimator of the average effect. ATE and ATET's estimators are expressed as follows[23]:

$$\hat{\alpha}_{ATE \to WEIGHTING\ PS} = \frac{1}{n} \sum_{i=1}^{n} Y_i \cdot \frac{D_i - \hat{\varepsilon}(X_i)}{\hat{\varepsilon}(X_i) \cdot \left(1 - \hat{\varepsilon}(X_i)\right)} \tag{9}$$

[23] Hirano, Imbens and Ridder (2003) developed *ATE* ($\hat{\tau}$) and *ATET* ($\hat{\tau}_{treated}$) estimators obtaining,

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i \cdot T_i}{\hat{p}(X_i)} - \frac{Y_i \cdot (1-T_i)}{1-\hat{p}(X_i)}\right) \qquad 0 = \sum_{i=1}^{N}\hat{p}(X_i)\cdot\left(\frac{Y_i \cdot T_i}{\hat{p}(X_i)} - \frac{Y_i \cdot (1-T_i)}{1-\hat{p}(X_i)} - \tau_{treated}\right)$$

where Y_i is the outcome, T_i the binary variable which indicates if individual it's treated or control and the X covariate's vector which let us to define $\hat{p}(X_i)$ as the probability to participate in the program, conditioned on X.

$$\hat{\alpha}_{ATET \to WEIGHTING\ PS} = \frac{1}{n_1} \sum_{i=1}^{n} Y_i \cdot \frac{D_i - \hat{\varepsilon}(X_i)}{1 - \hat{\varepsilon}(X_i)} \quad (10)$$

where $\hat{\varepsilon}(X_i)$ is the estimed value of the propensity score for the i-individual on vector X.

The obtained results are $\hat{\alpha}_{ATE}$ = 0,421280 and $\hat{\alpha}_{ATET}$ =0,430409.

The estimed value of the ATE is positive. In average, the sample's individuals' ability to find a job increases by 0,421280. In the case of the ATET, the estimed value is also positive, meaning that there is a favourable causal effect from the program. This result indicates that participant's ability to find a job has increased, on average, by 0,430409.

5.- ESTIMATION OF THE AVERAGE EFFECT BY REGRESSION.

It is possible to determine the average effect of a training program on participants (*ATET*) by regression by using Least Squares, given that independence assumption is also guaranteed. According to Stock and Watson (2003, pp. 375 and onwards), we can obtain an estimator of the average effects of the evaluated training program on partipants ($\hat{\alpha}_{ATET}$) by using a linear model.

It is also possible to introduce covariates in the model as additional regresors. By doing that, we can measure the effects that contaminant variables have on *Y*. The inclusion of covariates in the model is shown by the following expression:

$$Y = \mu + \alpha \cdot D + \beta \cdot X + \varepsilon \quad (11)$$

Y is the dependent variable which shows the potential results of the individual and *D* is a binary explanatory variable. *X* is the covariates' vector and β is the vector of the associated parameters. The parameter μ collects the fixed effects in the model and the parameter ε collects the random error of the model, with an average value equal to 0, $E[\varepsilon \mid D, X] = 0$. The parameter α will determine the average effect of the program on participants. This parameter will be the "differences estimator with additional regressors". As independence assumption is guaranteed, if the necessary assumptions[24] for multiple regression by Least Squares are guaranteed too, this estimator will be unbiased and consistent.

[24] These are the four assumptions: a) the conditioned distribution of the random error, given the explanatory variates X_{1i}, X_{2i}, …, X_{ki}, is equal to 0 on average (in this case, the explanatory variables are D_i, that indicates program participation, and the covariates). b) all the observations *i = 1, 2, …, n* are distributed both independently and identically random. C) X_{1i}, X_{2i}, …, X_{ki} and ε_i have four moments. D) Perfect multicolineality does not exist.

The inclusion of the additional regressors in the model lets us improve the estimator's efficiency, reducing the random error variance. On the other hand, the addition allows us to test the randomness in the individual assigning procedure between the participants group and the control group, in the case that the assigning procedure is related to the additional covariates. That is to say, by including the covariates in the model we can control the probability of individuals being assigned between the participants group and the control group by adding the characteristics in which participants and controls differ.

The inclusion of predetermined variables in the model will consist of inserting the variables included in the vector $X^3 = (\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3})$ as additional regressors.

Table 9 shows the correlation matrix between the explanatory variables included in the model, allowing us to analyze possible multicolineality in the model.

Matrix correlation between the variable D and the variables included in the vector $X^3 = (\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3})$						
	$\boldsymbol{D}$	$\boldsymbol{X_{11}}$	$\boldsymbol{X_2}$	$\boldsymbol{X_{31}}$	$\boldsymbol{X_{32}}$	$\boldsymbol{X_{33}}$
$\boldsymbol{D}$	1'000000	0'194895	0'168779	-0'007731	0'006873	0'027864
$\boldsymbol{X_{11}}$	0'194895	1'000000	0'005540	0'218229	0'014691	-0'092845
$\boldsymbol{X_2}$	0'168779	0'005540	1'000000	0'143666	-0'105606	0'030735
$\boldsymbol{X_{31}}$	-0'007731	0'218229	0'143666	1'000000	-0'334617	-0'323978
$\boldsymbol{X_{32}}$	0'006873	0'014691	-0'105606	-0'334617	1'000000	-0'410569
$\boldsymbol{X_{33}}$	0'027864	-0'092845	0'030735	-0'323978	-0'410569	1'000000

Table 9
Source: Own elaboration

Linear correlation between variables is not obvious because all the coefficients are very low and far from ±1. With regard to the relationship between $\boldsymbol{X_{31}}$, $\boldsymbol{X_{32}}$ and $\boldsymbol{X_{33}}$, these variables show slightly high correlation because of the fact that they have been constructed to include the zone variable in the model. However, the values are never over ±0'5. On the other hand, the value of the determinant of the correlation matrix is 0'4464, far from 0. In addition, the condition number of the correlation matrix is C = 2'3240, far from the limits that determine multicolineality. Everything we have exposed lets us indicate that multicolineality problems are not relevant in the model.

With the previous specifications made, we implement the regression by Least Squares. Results are shown in *Table 10*.

The calculation of the "differences estimator with additional regressors"					
Dependent Variable: *Y* (Ability to find a job)					
Method: MCO					
Variable	**Coefficient**	**Coefficient Value**	**Std. Error**	**t -Statistic***	**Prob.**
Fixed effect	μ	-0'206095	0'216058	-0'953891	0'3412
$\boldsymbol{D}$	α	0'409173	0'050124	8'163276	0'0000
$\boldsymbol{X_{11}}$	β_{11}	0'128764	0'045467	2'832055	0'0051
$\boldsymbol{X_2}$	β_2	0'023647	0'010274	2'301544	0'0223
$\boldsymbol{X_{31}}$	β_{31}	-0'040937	0'065065	-0'629174	0'5299
$\boldsymbol{X_{32}}$	β_{32}	-0'028182	0'059367	-0'474715	0'6355
$\boldsymbol{X_{33}}$	β_{33}	-0'075899	0'058980	-1'286872	0'1995
R-squared		0'360720	Mean dependent var		0'578147
Adjusted R-squared		0'343125	S.D. dependent var		0'380657
S.E. of regression		0'308514	Akaike info criterion		0'516520
Sum squared resid		20'74949	Schwarz criterion		0'622799
Log likelihood		-51'10855	F-statistic		20'50139
Durbin-Watson stat		2'023048	Prob(F-statistic)		0'000000
* t - Statistics adjusted by White's method					

Table 10

Source: Own elaboration

After estimating the parameters of the regressors, the equation's model is:

$$Y = -0'206095 + 0'409173\,\boldsymbol{D} + 0'128764\,\boldsymbol{X_{11}} + 0'023647\,\boldsymbol{X_2}$$
$$-0'040937\,\boldsymbol{X_{31}} - 0'028182\,\boldsymbol{X_{32}} - 0'075899\,\boldsymbol{X_{33}}$$

The α parameter (0'409173) is the "differences estimator with additional regressors" from the program's effects on the participants ($\hat{\alpha}_{ATET}$).

The individual significance of the explanatory variables included in the model is demonstrated by the values obtained for the t-statistc and its associated probability. From these values D, $\boldsymbol{X_{11}}$ (sex) and $\boldsymbol{X_2}$ (age) appear to be significant variables. The $\boldsymbol{X_{31}}$, $\boldsymbol{X_{32}}$ and $\boldsymbol{X_{33}}$, defined to include the zone covariate, appear insignificant. In any case, we have decided to mantain them because they help to improve the significance of all the estimated parameters and to improve goodness of fit. In relation to the fixed effects, the adjustment shows non significance of the constant in the regression.

As in the previous estimation, the t-statistic values has been adjusted by White's method in order to correct possible problems of heteroskedasticity.

With respect to goodness of fit, the R-squared statistic equals 0'360720 and shows that the explanatory power of the considered variables is equal to 36'0720 percent, significantly improving the accuracy of the adjustment over the estimation without additional regressors. On the other hand, the value of R^2 adjusted is equal to 0'343125. The standard regression error is very low, 0'308514, and the estimated residual variance is 0'09518. The low values of the information criteria of Akaike (0'516520) and Scharwz (0'622799) also support the accurracy of the model. In addition, the value of the Durbin-Watson statistic is close to 2, 2'023048, indicating that autocorrelation problems of the residuals are not relevant.

The joint significance of all the model estimated parameters can also be tested from the value of the probability of the F-Snedecor contrast. In this case the probability is equal to 0'00000, meaning the acceptance of the joint significance of all the parameters of the model. This implies that we can consider all the parameters significantly different from 0 whilst, at the same time, having a high probability.

With the estimation carried out, the meaning of all the parameters of the model is important. From the results obtained:

- The α coeficient, associated with the explanatory variable D, shows that when an individual has participated in the program ($D = 1$), the response variable increases by 0'409173. This is the effect on the response variable of the participants, and means that the ability to find a job increases by 0'409173 over the non participants' value.
- The β_{11} coefficient, associated with the explanatory variable $\boldsymbol{X_{11}}$, shows that in the case of a male participant ($\boldsymbol{X_{11}} = 1$), the response variable increases by 0'128794. In the case of a woman participant ($\boldsymbol{X_{11}} = 0$), this effect is not added. This means that males have a better situation than females in terms of Y, which is higher by 0'128794 than the registered value in the case of females.
- The β_2 coefficient, associated with the explanatory variable $\boldsymbol{X_2}$, collects the effect of age on Y. Due to variable $\boldsymbol{X_2}$ being a quantitive variable, this effect will be related to the possible values of this variable, adding 0'023647 to the value of the response variable Y for every unitary change registered by $\boldsymbol{X_2}$. Therefore, individuals belonging to the sample (individuals whose age is between 17 and 26) show a higher ability to find a job as the value of $\boldsymbol{X_2}$ increases.
- The β_{31}, β_{32} and β_{33} coefficients, associated with the explanatory variables $\boldsymbol{X_{31}}$, $\boldsymbol{X_{32}}$ and $\boldsymbol{X_{33}}$, show that:

When the individual belongs to zone 1 (X_{31} = *1*), the effect on variable response *Y* is a reduction that equals 0'040937. When he belongs to zone 2 (X_{32} = *1*), the effect on variable response *Y* is a reduction that equals 0'028182. The reduction is equal to 0'075899 when the individual belongs to zone 3 (X_{33} = *1*). Finally, if the individual belongs to zone 4 (X_{31} = *0,* X_{32} = *0* and X_{33} = *0*), no effect is added due to the fact that we take this category as the base.

Therefore, this shows that individuals belonging to zone 4 have the best results with respect to the response variable Y, followed by individuals belonging to zone 2 and then individuals belonging to zone 1.

- The μ coefficient collects the effect on *Y* in the event that all of the explanatory variables become null. It shows the effect on *Y* in the event that the individual has not participated in the program (*D = 0*), is female (X_{11} = *0*) and belongs to zone 4 (X_{31} = *0,* X_{32} = *0* and X_{33} = *0*). This value equals 0'206095. We have to add to this value the effect provoked by the covariate X_2 (age).

To summarize, *Tables 11 and 12* contain the model estimated values for the response variable *Y* for every possible value of the explanatory variables. Table *11* contains the values for the case of participants while *Table 12* contains the values for the case of non participants.

Estimated values for the response variable *Y* for the participant individuals according to "the differences estimator with addtional regressors" model

	PARTICIPANT INDIVIDUALS							
	Male				Female			
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 1	Zone 2	Zone 3	Zone 4
17 years old	0'692904	0'705659	0'657942	0'733841	0'56414	0'576895	0'529178	0'605077
18 years old	0'716551	0'729306	0'681589	0'757488	0'587787	0'600542	0'552825	0'628724
19 years old	0'740198	0'752953	0'705236	0'781135	0'611434	0'624189	0'576472	0'652371
20 years old	0'763845	0'7766	0'728883	0'804782	0'635081	0'647836	0'600119	0'676018
21 years old	0'787492	0'800247	0'75253	0'828429	0'658728	0'671483	0'623766	0'699665
22 years old	0'811139	0'823894	0'776177	0'852076	0'682375	0'69513	0'647413	0'723312
23 years old	0'834786	0'847541	0'799824	0'875723	0'706022	0'718777	0'67106	0'746959
24 years old	0'858433	0'871188	0'823471	0'89937	0'729669	0'742424	0'694707	0'770606
25 years old	0'88208	0'894835	0'847118	0'923017	0'753316	0'766071	0'718354	0'794253
26 years old	0'905727	0'918482	0'870765	0'946664	0'776963	0'789718	0'742001	0'8179

Table 11
Source: Own Elaboration

Estimated values for the response variable Y for the control individuals according to "the differences estimator with addtional regressors" model

	CONTROL INDIVIDUALS							
	Male				Female			
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 1	Zone 2	Zone 3	Zone 4
17 years old	0’283731	0’296486	0’248769	0’324668	0’154967	0’167722	0’120005	0’195904
18 years old	0’307378	0’320133	0’272416	0’348315	0’178614	0’191369	0’143652	0’219551
19 years old	0’331025	0’34378	0’296063	0’371962	0’202261	0’215016	0’167299	0’243198
20 years old	0’354672	0’367427	0’31971	0’395609	0’225908	0’238663	0’190946	0’266845
21 years old	0’378319	0’391074	0’343357	0’419256	0’249555	0’26231	0’214593	0’290492
22 years old	0’401966	0’414721	0’367004	0’442903	0’273202	0’285957	0’23824	0’314139
23 years old	0’425613	0’438368	0’390651	0’46655	0’296849	0’309604	0’261887	0’337786
24 years old	0’44926	0’462015	0’414298	0’490197	0’320496	0’333251	0’285534	0’361433
25 years old	0’472907	0’485662	0’437945	0’513844	0’344143	0’356898	0’309181	0’38508
26 years old	0’496554	0’509309	0’461592	0’537491	0’36779	0’380545	0’332828	0’408727

Table 12
Source: Own Elaboration

5.- CONCLUSIONS.

The training schooll program’s average effect estimed by weighting observations by the inverse of a nonparametric estime of the propensity score let us to conclude that, for treated, the time needed to find a job is reduced in 471 days. As the program was designed to improved the employ between youngers unemployed; this results supports the effectiveness of this public policy.

The training schooll program’s average effect estimed by the differences estimator let us to conclude that, for treated, the time needed to find a job is reduced in 448 days. This result also supports the effectiveness of this public policy.

Another conclusion can be obtained from using the covariates information contained in the BASEVAFOR 96-03. More specifically, in the case of males, on average, the period needed for a treated for find a job, is reduced in 141 days. By considering age, the same period is reduced in 26 days per year from 16 to 25 years old.

According to the obtained results, the effectiveness of the training program from the “differences estimator with additional regressor” is positive too.

Both evaluations show evidence of that this program can reduce the time needed to find a job.

Further investigations might improve conclusions if public authorities let researchers to extend the database information with data related with others individual characteristics.

REFERENCES.

Barnow, B., G. Cain and A. Goldberger (1980): "Selection on Observables", en Stromsdorfer, E. W. and Farkas, G. (eds.), *Evaluation Studies Review Annual*, vol. 5: 43-59, Beverly Hills, California, Ed.: Sage Publications.

Bassi, L. J. (1984): "Estimating the Effect of Training Programs with Non-Random Selection", *Review of Economics and Statistics*, 66 (1): 36-43.

Billewicz, W. Z. (1965): "The Efficiency of Matched Samples: an Empirical Investigation", *Biometrics,* 21 (3): 623-644.

Burtless, G. (1995): "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, 9 (2): 63-84.

Cameron, E. and L. Pauling (1976): "Supplemental Ascorbate in the Supportive Treatment of Cancer: Prolongation of Survival Times in Terminal Human Cancer", *Proceedings of the National Academy of Sciences of the United States of America,* 73 (10): 3685-3689.

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometric. Methods and Applications*, New York, Ed.: Cambridge University Press.

Cansino, J.M., M. A. Cardenete and R. Roman: (2007): "Regional evaluation of a tax on retailer sales of some fuels through social accounting matrix", *Applied Economics Letters*, 14 (12): 877-880.

Cansino, J.M. y A. Sanchez (2008a): "Evaluación del programa de Escuelas Taller y Casas de Oficios a partir de su efecto sobre el tiempo de búsqueda del primer empleo. El caso de Sevilla", *Estudios de Economía Aplicada*, *forthcoming*.

Cansino, J.M. y A. Sanchez (2008b): "Average effect of training programs on the time needed to find a job. The case of the training programs schools in the south of Spain (Seville, 1997-1999)". *Funcas Workpapers*. Available in www.funcas.es.

Cochran, W. G. (1968): "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies", *Biometrics,* 24 (2): 295-313.

Cox, D. R., (1992): "Causality: Some Statistical Aspects," *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 155 (2): 291-301.

Dawid, A. P. (2000): "Causal Inference Without Counterfactuals", *Journal of the American Statistical Association*, 95 (2): 407-448.

Dawid, A. P. (1979): "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society, Series B (Statistics Methodological)*, 41 (1): 1-31.

Dehejia, R. H. and S. Wahba (1999): "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94 (448): 1053-1062.

Fisher, R. A. (1935): *The design of experiments*, Edinburgh, Ed.: Oliver and Boyd.

Fisher, R. A. (1928): *Statistical Methods for Research Workers*, 2ª edición, London, Ed.: Oliver and Boyd.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Ecometrica*, 66 (2): 315-331.

Heckman, J. J (1990): "Varities of Selection Bias", *American Economic Review*, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, 80 (2): 313-338.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47 (1): 153-162.

Heckman, J., N. Clements and J. Smith (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts" *Review of Economic Studies*, 64 (4): 487-535.

Heckman, J. J. and V. J. Hotz (1989): "Choosoing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, 84 (408): 862-874.

Heckman, J. J., H. Ichimura, J. Smith and P. E. Todd (1998): "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66 (5): 1017-1098.

Heckman, J. J., H. Ichimura and P. E. Todd (1998): "Matching As an Econometric Evaluation Estimator", *Review of Economics Studies*, 65 (2): 261-294.

Heckman, J. J., H. Ichimura and P. E. Todd (1997): "Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economics Studies*, 64 (4): 605-654.

Hirano, K., G. Imbens and G. Ridder, (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica*, 71(4): 1161-1189.

Holland, P. W. (1986): "Statistics and Causal Inference" (with discussion), *Journal of the American Statistical Association*, 81 (396): 945-970.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review", *Review of Economics and Statistics*, 86 (1): 4-29.

Kiefer, N. (1979): *Economic Benefits from Four Manpower Training Programs,* Garland Series of Outstanding Dissertations in Economics, New York,. Ed.: Garland Press.

Little, R. and D. Rubin (1987): *Statistical Analysis with Missing Data*, Nueva York, Ed.: Wiley.

Manski, C. F. (2001): "Designing programs for heterogeneous populations: The value of covariate information, *American Economic Review*, 91 (2): 103-106.

Moertel, C., T. Fleming, E. Creagan, J. Rubin, M. O'Connell and M. Ames (1985): "High-dose Vitamin C Versus Placebo in the Treatment of Patients with Advanced Cancer Who Have Had no Prior Chemotherapy: a Randomized Double-blind Comparison", *New England Journal of Medicine,* 312: 137-141.

Neyman, J. (1935): "Statistical Problems in Agricultural Experimentation", *Supplement to the Journal of the Royal Statistical Society,* 2: 107-180.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles.", re-edited en *Statistical Science* (con discusión), 1990, 5(4): 465-472.

Pearl, J. (2000): *Causality: Models, Reasoning and Inference*, Cambridge, Ed.: Cambridge University Press.

Rosenbaum, P. R. (1999): "Choice As an Alternative to Control in Observational Studies" (with discussion), *Statistical Science*, 14 (3): 259-304.

Rosenbaum, P. R. (1996): "Observational Studies and Nonrandomized Experiments", en S. Ghosh and Rao, C. R. (eds.), *Handbook of Statistics*, vol. 13, capítulo 6: 1277-1366, New York, Ed.: Elsevier.

Rosenbaum, P. R. (1995): *Observational Studies*, Springer Series in Statistics, Nueva York, Ed.: Springer-Verlag.

Rosenbaum, P. R. and D. B. Rubin (1984): "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score", *Journal of the American Statistical Association*, 79 (387): 516-524.

Rosenbaum, P. R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrica*, 70 (1): 41-55.

Roy, A. (1951): "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3 (2): 135-146.

Rubin, D. B., (1978): "Bayesian Inference for Causal Effects the Role of Randomization," *Annals of Statistics*, 6 (1): 34-58.

Rubin, D. B. (1976): "Inference and Missing Data", *Biometrika*, 63 (3): 581-592.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies", *Journal of Educational Psychology*, 66 (5): 688-701.

Stock, J. M. and Watson M. W (2003): *Introduction to Econometrics*, Boston, MA. Ed.: Addospm Wesley.