The Impact of Training Policies in Latin America and the Caribbean: The Case of "Programa Joven"

Research Report

By

Graduate Program in Economics
ILADES/Georgetown University

Index

1. Introduction

This research evaluates the "Programa Joven", a training program conducted by the Ministerio del Trabajo of Argentina. This research aims at adapting, applying and testing a non-experimental evaluation methodology, which is considered to be suitable conditional on the quality of the information available to the researchers.

We will present the main methodological definitions utilized and report empirical results obtained for some of the suggested Matching estimators.

The Matching Estimators approach was used as our basic methodology to obtain a measure of the impact of this training program. The choice of this methodological approach was based upon both the theoretical developments in the area of Program Evaluation (Heckman et al, 1995, 1997, 1998 & 1998) and the availability of relevant information. The application of this methodology presents two stages (Todd, 1999): first, the estimation of a model of program participation (Propensity Scores) and second, conditioning on the estimated propensity scores, the usage of matching estimators to obtain the impact of the Program.

To estimate the propensity scores we obtained and processed raw data for all the individuals who registered and qualified to take training programs in the period March 1996 to December 1997 (approximately 140,000 individuals). It is important to note that the samples of beneficiaries and controls we use to evaluate the Program (3,369 individuals in total) were extracted from this universe of potential trainees. We report Program impact estimates based on propensity scores estimated with the universe information and with the sample information.

Based upon cost information and the program impact estimates obtained (benefits), we performed a cost benefit analysis. This analysis was conducted under different scenarios with regard to benefit duration, discount rate and deadweight loss of taxes.

2. Description of "Programa Joven"

The "Programa Joven" offers training to facilitate the labor insertion of the beneficiaries in the formal labor market. To this purpose, the program provides intensive training for positions in the productive sector of the economy and training internships in firms.

The target population of the Program are young persons, males and females, coming from poor households, with a low educational level, without working experience, and who are unemployed, underemployed or inactive. The selection criteria for the program are: minimum 16 years of age; education level no greater than secondary education; to belong to a household considered poor and not to be integrated to the labor market.

The program comprehend the following benefits: an average of 200 hours of training, transportation expenses, a subsidy for females with young children, medical checkups, books, material and working clothing.

The criteria for selecting firms for internships are: general characteristics of the firm, tasks to be done by the trainees, personal involved in similar positions in the firm, equipment, supplies and infrastructure of the firm.

The duration of the training program varies from 14 to 20 weeks. The training is intensive and can be divided into two main activities:

a. Technical Knowledge: the beneficiaries receive knowledge and technical skills to undertake an occupation. The duration of this activity varies from 6 to 12 weeks.

b. Internships: the beneficiaries complement their technical knowledge with an applied work in firms in the occupations they have been trained. The duration of this activity is of 8 weeks.

To carry out the training the Ministerio del Trabajo hires Instituciones de Capacitación (ICAP) using an international bidding process. The distribution of the training activities at the national level is determined in accordance to the quantity of inhabitants in the regions. At the governmental level the program executor is the Ministerio del Trabajo y Seguridad Social, which uses to this end the Secretaría de Empleo and Capacitación Laboral. The program will continue its operation until the year 2,000, being jointly funded by IDB and governmental funds.

There is an evaluation and monitoring of the outcomes and impacts of the Program. The impact evaluation analyzes the degree of correspondence of the actions undertaken for the Program and the outcomes obtained. This is done with the use of a quasi experimental model with a comparison group, which tries to identify the magnitude of the changes and outcomes of the Program on social, economic, educational and attitudinal impacts on the trainees compared with non trainees.

3. The Data: Description of the Different Information Available for the Evaluation

We use two sources of information. The first data set comprehend raw data on the 139,732 individuals who registered and qualified for training programs during 03/1996 and 12/1997. The second data set comprehend two samples, one for beneficiaries and one for controls of 1.670 individuals each, used by the Ministerio del Trabajo to evaluate the impact of the fifth wave of the training program. As mentioned before, these samples were extracted from the universe of 139,732 individuals in the first data set.

Before describing these data sets, let us define some terminology. For the Programa Joven the "Acreditación" correspond to the first contact of a person with the Program, which enables them to register and to participate in a training class. Therefore, the ***"Acreditados"*** corresponds to individuals who are eligible for participating in the Program, but may or may not have taken any class. They define as ***"Beneficiaries"*** to individuals who have registered and started training classes.

a. Data on *"Acreditados"*

We obtained and process raw information on 139,732 *"Acreditados"* who registered and qualified to take training programs in the period March 1996 to December 1997. We found that 80% of them presented complete records, but for the rest 20% we had to work case by case to complete their records.[1] After this work, some of the individuals were excluded from the analysis because of missing information. This information was used to obtain the probability of program participation (Propensity Scores). Later in this Report we will extend on the information we constructed and how it was used.

b. Data on *"Beneficiaries"* and *"Comparisons"*

We have data on two samples, one for *Beneficiaries* (intervention group) and one for *Comparisons* (comparison group) of 1,670 individuals each, used by the Ministerio del Trabajo to evaluate the impact of the fifth wave of the training program. Both groups come from individuals who meet the selection criteria to be considered as a potential participant

[1] We used other data available at the Programa Joven to complete these records.

of the program *("Acreditados"*). In addition, the *Beneficiaries* are the ones who actually attended the program.

Both samples comprehend 1,670 cases (3,340 cases in total). To make both samples comparable the sample design used by the Ministerio del Trabajo controlled for the following variables: age, sex, level of education, labor insertion, socioeconomic level, and to have a children with 5 years of age or less. Therefore, the comparison group is not selected at random.

For both samples, there is information for the period covered by the first contact with the program to 12 months after the *Beneficiaries* finished the training. This allows us to construct the individual labor history for both samples.

Beneficiaries Sample

The sample was designed by the Ministerio del Trabajo to have statistical representation by gender and region of residence. The first variable was introduced to study the program impact by gender, given the different labor market conditions for males and females. The regional variable was introduced to study the differential impact of socioeconomic characteristics and regional labor markets on program outcomes. In total, they considered 11 geographic units denominated "regions".

To define the sample sizes it was considered the observed variation in the values for variables such as proportion of employed/unemployed workers and average income received by employed workers. These variables present the greatest variation among the outcomes variables. It was considered a percentage of non-response of 5%. The determination of the sample sizes was estimated under the hypothesis that a proportion of P=0.35 of unemployed wants to be estimated with a precision of 10%, with a risk level of 1%. In other words, the interval (P-0.1, P+0.1) contains the estimated "p", of the population proportion P, with probability 0.95.

Comparison Sample

Once the *Beneficiaries* sample was obtained, a comparison sample was constructed. For each beneficiary a "twin" was selected among the people who have approached the Program, satisfied the selection criteria but did not take the training program ("certified"

individuals). The "twin" was obtained at random from the universe of "certified" individuals, which presented similar socioeconomic characteristic as the person included in the *Beneficiaries* sample.

The Ministerio del Trabajo used the following variables to match the individuals: first, region, sex and age, and second, educational level, and presence of children. In the cases in which it was not possible to find an "identical" individual, a replacement was found to match as closed as possible the socioeconomic and geographic characteristics. This procedure generated a sample, which is identical in terms of region and sex, presenting some differences in terms of level of education.

In total 3,340 individuals were considered: 1,670 in the *Beneficiaries* sample and 1,670 in the comparison group.

4. Program Participation

4.1 Determinants

The estimation of the probability of program participation is one of the main elements needed to apply a cross-sectional Matching Estimator. We estimated three models of program participation: the first one, using the universe of *"Acreditados"* (139,732 individuals), that is, we estimate the conditional probability of program participation conditioning on eligibility. The second one, using the sample of 3,339 individuals, re-weighted for choice based sampling, and using the information available at the "Acreditación". Finally, a third one, using the sample of 3,339 individuals, re-weighted for choice based sampling, and using the information available at the sample.

It is important to recall the requisites for eligibility. An individual is eligible if:

- *Housing:* they don't live in a house or if the house they live in does not have a bathroom or if the house they live in is "crowded" (more than 3 person per room).
- *Income:* Per capita household income below US$ 120 per month.
- *Labor Status:* The individual is searching for a job or she/he works for a wage under US$200 a month or she/he is head of the household and her/his labor income is below US$400 a month and she/he is looking for a new job or she/he neither is working or searching for a job but she/he wishes to work.

- *Capability of Living:* Ratio head of the household to number of dependent smaller than 0.25 and level of education of head of the household at most incomplete primary.

We tried to obtain information related to the labor history of program participants, unfortunately this information was not available in the data sets. According to the authorities in charge of the Program they did not include this type of questions because individuals did not have incentives (in fact, in some cases they have disincentives) to reveal the truth and there was no readily available mechanism to check this information. For this reason, the information gathered at first contact ("Acreditación") did not contained information on labor history.

However, this information was useful for this research because it allowed us to construct the "Program History" of any individual who has been *"Acreditado"*. As mentioned before, the Programa Joven is composed of a Technical Knowledge phase and an Internships phase. Therefore an *"Acreditado"* may be in different states: she/he may or may not have started a training program; she/he may have started the Technical Knowledge phase but she/he may or may not have finished it and she/he may have started the Internships phase but she/he may or may not have finished. In fact we have information to classify the universe of *"Acreditados"* in the following mutually exclusive categories:

- *"Acreditado" only:* Individuals who are eligible for training programs but have not started the Technical Knowledge phase.
- *Incomplete Technical Knowledge phase:* Individuals who did not finish the Technical Knowledge phase because of a justified reason (family problems, pregnancy, obtained a job, etc.).
- *Deserter of Technical Knowledge phase:* Individuals who did not finish the Technical Knowledge phase and did not have a justified reason.
- *Did not approve the Technical Knowledge phase:* Individuals who did not reach the minimum standards required for approval.
- *Incomplete Internships phase:* Individuals who did not finish the Internships phase because of a justified reason (family problems, pregnancy, obtained a job, etc.).

- *Deserter of Internships phase:* Individuals who did not finish the Internships phase and did not have a justified reason.
- *Did not approve the Internship phase:* Individuals who did not reach the minimum standards required for approval.
- *Completes:* Individuals who have successfully completed both phases

Empirically, around a 52% of the *"Acreditados"* are in the *"Acreditado" only* category, and 37% of the *"Acreditados"* are in the category *Completes*. Another useful piece of information obtained is the type of training program undertaken by the *Beneficiaries*.[2] This information will be used in another research to address the issue of multi-treatment.

Given that individuals can be in different program "states" an important question for the propensity score model is how to define when an individual has taken the program (value 1) and when the individual has not taken the program (value 0). We tried several definitions: for example, we started considering a value 1 if an individual started the Technical Knowledge phase and 0 otherwise, but the adjustment was not considered "good". Our reasoning was that this definition was not useful to distinguish among the individuals. The option we took was to consider the variable ESTABTOT, which takes a value 1 if the individual has successfully completed the Technical Knowledge phase and 0 otherwise. This choice allowed us to use most of the *"Acreditado"* (around a 89% of them) and it is consistent with the way in which the Ministerio del Trabajo obtained its samples of *Beneficiaries* and *Controls.*

The following individual dimensions were used in the model of program participation:

- *Labor Status Dimension:* This variable reflects the labor status of the individual (employed, unemployed with and without labor experience, and inactive).
- *Poverty Dimension:* We use an index of unmet basic needs (NBI). This index consider an individual as poor if the person lives in an special home (minors, or unmarried mothers) or if the house they live in does not have a bathroom or if the

[2] Tertiary Sector (educative services, administration and accounting), assistant of firms and services, dental assistant, old men services, computation, gastronomy, hotel and tourism, janitor and maintenance, media and publicity, photography, hairdressing, sales, telephony, surveillance); industrial sector (construction, quality control, electronics, textiles, chemical laboratories, auto mechanic, industrial painting, plastic, refrigeration, graphic industry); agricultural, forest and mining (gardening, cultivation, watering, mining exploitation, cattle production).

house they live in is “crowded” (more than 3 person per room) or if the ratio head of the household to number of dependent smaller than 0.25 and the level of education of head of the household at most incomplete primary.

We considered also the Poverty Line criteria, using as reference the income level of the individuals at the moment of “Acreditación” and as poverty line $120 per month.

- *Sociodemographic Dimension:* We use gender and age.
- *Education and Marital Status Dimension:* We use several indicators of years of education completed, as well as school attendance at the moment of “Acreditación”. The Marital dimension was considered by measuring whether the individual was married or single, whether he/she had children (specially young children) and whether the individual was or not the head of the household.
- *Geographical Dimension:* We worked with the same 11 regions, which were used by the evaluation samples considered by the Ministerio del Trabajo.

Table 4.1

Regions	**Participation (%)**
GBA	36.2
Sur	6.0
Nea	1.4
Centro	8.7
Litoral	5.4
Cuyo	4.1
Noa	7.1
Córdoba	9.3
Mendoza	9.9
Sta.Fe	8.0
Tucumán	3.9
Total	***100***

Finally, we considered four groups in our estimations based upon gender and age. The groups were:

1. Adult Males – ages 21 to 35.

2. Young Males – ages less than 21.
3. Adult Females – ages 21 to 35.
4. Young Females – ages less than 21.

4.2 Strategic Behavior

According to the authorities in charge of the Program they suspected that individuals followed strategic behavior in order to became eligible for the Program. However, the authorities did not have a readily available mechanism to check the information provided by the individuals at the "acreditación". To address this issue we compare information available at the "acreditación" with some information revealed at the survey by the 3,339 individuals in the beneficiaries and comparisons groups twelve months after the Program. The questions refer to their labor status at the "acreditación".[3]

The Tables 4.2 to 4.4 present cross-information about unemployment status (1 unemployed and 0 otherwise) at both "acreditación" and survey, for all the individuals, for beneficiaries only and for comparisons only. The information related to "acreditación" is presented in the rows while the information related to the survey is presented in the columns.

Table 4.2

(All individuals)

Survey

		0	**1**	
"Acreditación"	**0**	34	145	179
	1	542	2,619	3,160
		576	2,763	3,339

Table 4.3

(Beneficiaries)

Survey

[3] The "acreditación" information that was asked was rather limited and refers mainly to their labor status previous to the Program.

		0	1	
“Acreditación”	**0**	34	145	179
	1	542	2,619	3,160
		576	2,763	3,339

Table 4.4

(Comparisons)

Survey

		0	1	
“Acreditación”	**0**	34	145	179
	1	542	2,619	3,160
		576	2,763	3,339

We could consider that the individuals who declare to be unemployed at the “acreditación” but revealed not to be unemployed at the survey were the ones who behaved strategically at the “acreditación”. Using this as an indicator we have in Table 4.2 that 542 individuals out of 3,160 (17,2% of this individuals) who declared to be unemployed at the “acreditación” were “misbehaving”. Separating between beneficiaries and comparisons, we have that the percentages of “misbehaviors” are 17,9% and 16,5% respectively. Based upon this (rather limited) indicator we don’t have evidence of strategic behavior at “acreditación” and we don’t perceive also differences in this issue between the two groups.

5. Estimation of Program Participation (Propensity Scores)

We estimated different models for each of the four subgroups: Young Males, Young Females, Adult Males and Adult Females. The main variables and its descriptive statistics are presented in Appendix 1.

As previously mentioned, we conducted three estimations for the Propensity Scores. The first one uses the individuals and the information available at “acreditación” (139,732 cases). The second uses the information available at “acreditación” but only considers the

individuals in the survey (3,339 cases). The third uses both the individuals and the information available at the survey (3,339 cases).

In the second and third cases we have to re-weight the sample previous to the econometric work. This is because the sample by design contains equal percentages of beneficiaries and comparisons.[4]

5.1 Universe

The main econometric results for the binary Logits and their prediction tables, for the four groups can be seen in Appendix 2. The majority of the estimated coefficients are negative, which is something that captures our attention, but for which we could not obtain a convincing explanation. A possible one is that when we considered the probability of being a Beneficiary rather than an "Acreditado Only", the same condition that increases the likelihood of becoming eligible is negatively correlated with the likelihood of completing the technical knowledge phase of the training Program.

The fit of the model for the different groups is presented in the following Table:

Table 5.1

Group	R-squared	Prediction Evaluation
Young Males	0.0295	% Correct goes from 50,29 to 52,28
Adult Males	0.0347	% Correct goes from 50,28 to 52,63
Young Females	0.0220	% Correct goes from 52,02 to 53,44
Adult Females	0.0297	% Correct goes from 51,96 to 53,88

The predicted Propensity Scores go from a minimum value of 0.2298 to a maximum value of 0.7880, showing a wide range of dispersion.

5.2 Universe and Sample

In this case we considered the 3,339 cases in the survey but we use the information available for them at the "Acreditación". The main econometric results for the binary Logits and their prediction tables, for the four groups can be seen in Appendix 3. As

[4] To re-weight we followed Manski and Lerman, 1977.

mentioned before the sample was re-weighted prior to estimation to correct for Choice-Based Sampling. Following Manski and Lerman (1977) we reweighted each observation by the ratio of the proportion of beneficiaries in a random population divided by the proportion of beneficiaries in our sample. The latter was approached by the sample proportion of beneficiaries, while for the former we used the universe information to estimate the proportion of beneficiaries in this universe.

The fit of the model for the different groups is presented in the following Table:

Table 5.2

Group	R-squared	Prediction Evaluation
Young Males	0.079	% Correct goes from 54,7 to 66,0
Adult Males	0.012	% Correct goes from 51,4 to 53,8
Young Females	0.057	% Correct goes from 53,6 to 63,3
Adult Females	0.011	% Correct goes from 53,6 to 55,7

The predicted Propensity Scores go from a minimum value of 0.18 to a maximum value of 0.91, showing a wide range of dispersion as before.

5.3 Sample

Finally, we considered the 3,339 cases in the survey but we use the information available at the survey. The main econometric results for the binary Logits and their prediction tables, for the four groups can be seen in Appendix 4. As mentioned before the sample was re-weighted prior to estimation to correct for Choice-Based Sampling.

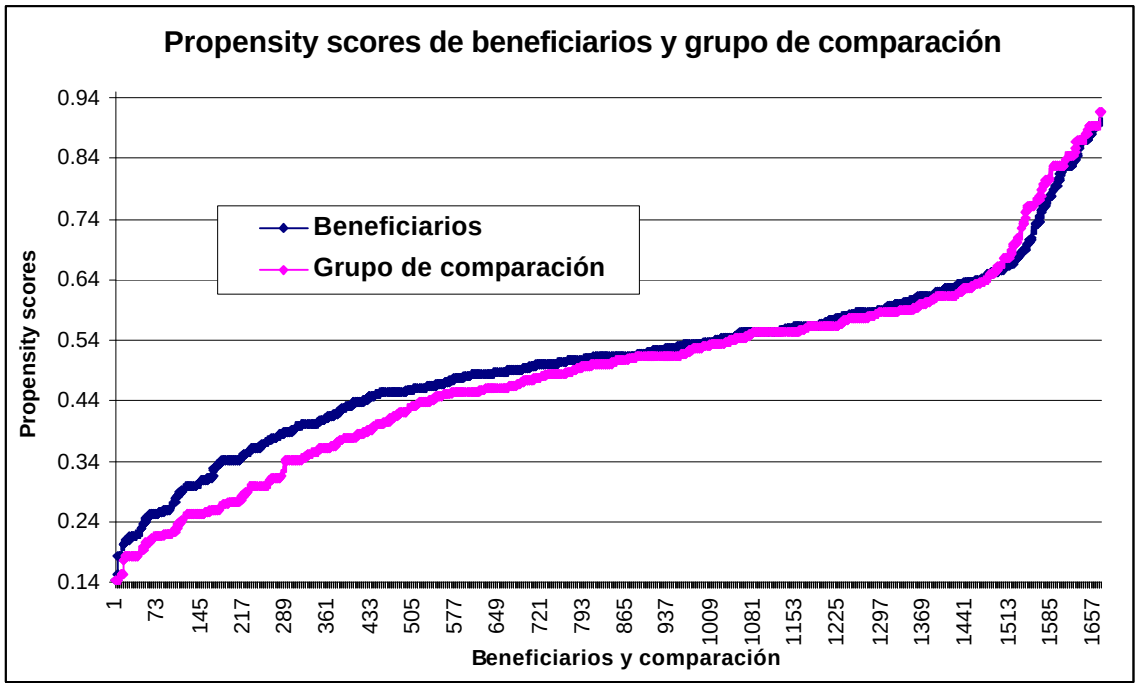
Table 5.3

Group	R-squared	Prediction Evaluation
Young Males	0.078	% Correct goes from 51,7 to 65,2
Adult Males	0.010	% Correct goes from 55,0 to 56,9
Young Females	0.049	% Correct goes from 52,8 to 61,2
Adult Females	0.003	% Correct goes from 53,3 to 54,1

The predicted Propensity Scores go from a minimum value of 0.15 to a maximum value of 0.88, showing a wide range of dispersion.

5.4 Determining a common support

The application of nonparametric estimators requires that exists propensity scores values for the *Comparisons* in the vicinity of each of the propensity scores for the *Beneficiaries*. In order to analyze whether this was a problem for some of the propensity scores values for the *Beneficiaries* sample we plotted the histograms of the propensity scores for both groups. The next figure presents the histograms.



Visually we do not observe values for the propensity scores in the sample of *Beneficiaries* for which it will not be able to find in its vicinity propensity scores in the sample of *Comparisons*. For this reason we did not excluded any observation in the *Beneficiaries* sample from the analysis. It is important to remember that in the Programa Joven they considered several criteria to select the Comparisons, controlling by some variables such that age, gender, labor status, marital status and existence of Children and controlling also by their distribution. This makes more likely that the two populations present similar propensity scores.

6. Impact estimates

Once the models of Program Participation were estimated then the estimated functional form were used to predict the propensity scores for each of the individual who were in the two samples used by the Ministerio del Trabajo (3,339 cases in total with 1,700 beneficiaries and 1,669 controls).

The outcome variables considered were: the mean earnings impact on the treated and the probability of employment on the treated.

We worked with a cross-sectional (CS) matching estimator, given that this methodology compares the results for the *Beneficiaries* and *Comparisons* at the same period after the program. The information available allows us to apply this methodology.

The specific cross-sectional matching estimators used were:

- ***Nearest Neighbor Matching Estimator :***[5] This is the simplest method to implement and its specific formulas can be seen in Todd (1999). The number of neighbors to include from the *Comparisons* sample for each *Beneficiary* is taken as given. For each Beneficiary we include only income information of the specified number of Comparisons with the lowest Euclidean distance to the ith *Beneficiary* propensity scores.
- ***Epanechnikov Kernel:*** We implemented the Epanechnikov kernel and its specific formulas can be seen in Silverman (1986). In this kernel the window width can be optimally chosen with reference to a standard normal distribution. To fix the window width we followed the adaptative estimate of spread suggested by Silverman (1986, page 47).

The technique of **bootstrapping** was used to obtain the sample variance of the impact estimates. The Appendix 5 presents the Matlab (version 5.3.1) code which were used in estimation.

6.1 Earnings Results

[5] These estimators assume that $E[Y_0 \mid P(X), D=1] = E[Y_0 \mid P(X), D=0]$ and $0 < Pr(D=1 \mid X) < 1$.

The main results for the program impact estimates on earnings are presented in Tables 6.1 and 6.2. In Table 6.1 we present impact estimates for 5, 10, 20, 30, 40 and 50 neighbors, for the four sub-groups and for the whole sample. We also report program impacts using the three estimated Propensity Scores: 1) using the universe individuals and information (PSTOT); 2) using the universe information but the individuals in the survey (PSUN); and 3) using the individuals and information from the survey (PSMU).

From this Table we can see that intra-groups it does make a difference what type of information and individuals we consider estimating the Propensity Scores, as program impact vary substantially specially for adult males and young females. However, for the whole sample the mean impacts obtained using PSTOT and PSMU are similar.

In the case of the Epanechnikov kernel (Table 6.2) the estimated program impact presents substantial variation in the case of adult males and young females. However, for the whole sample the mean impacts are similar regardless of the pair information-individual used to estimate the Propensity Scores.

Table 6.1

Nearest Matching Estimators

A. Impact Estimates

1. Usando PSTOT

Neighbors	H-Joven	H-Adulto	M-Joven	M-Adulta	Todos
5	$19.5580	$8.7916	$9.4172	$28.7714	$15.9891
10	$20.7947	$8.9829	$4.4825	$26.3506	$14.1260
20	$16.7581	$8.2552	$5.5668	$28.2364	$15.1815
30	$18.3655	$7.1672	$4.7808	$27.1712	$14.2402
40	$22.4229	$7.4626	$4.5025	$27.4691	$14.9518
50	$21.5508	$8.5232	$5.4694	$27.4820	$14.7643

2. Usando PSUN

Neighbors	H-Joven	H-Adulto	M-Joven	M-Adulta	Todos
5	$31.6918	$1.5785	$10.5613	$28.1383	$25.3236
10	$33.5754	-$5.2291	$14.9890	$27.7776	$29.2675
20	$31.0609	$0.2809	$11.1461	$32.5890	$25.1187
30	$27.9784	-$1.0621	$9.3960	$33.3182	$24.6145

40	$18.8649	-$0.3656	$9.7609	$35.4693	$23.8325
50	$22.1197	-$1.3799	$9.7000	$38.4664	$24.8189

3. Usando PSMU

Neighbors	H-Joven	H-Adulto	M-Joven	M-Adulta	Todos
5	$28.9379	-$1.2083	$8.6468	$28.7565	$15.9841
10	$32.1767	$1.0928	$12.8176	$28.7564	$14.9506
20	$25.2899	$2.6456	$9.1977	$30.4410	$14.3877
30	$25.2714	$5.2481	$7.5491	$30.4410	$14.4088
40	$16.4604	$1.8064	$7.8928	$30.4410	$15.0288
50	$20.5334	$3.5606	$7.9202	$30.4410	$15.0694

Table 6.2

Epanechnikov Kernel

B. Impact Estimates: Epanechnikov Kernel

	H-Joven	H-Adulto	M-Joven	M-Adulta	Todos
PSTOT	$13.5584	-$1.9703	$2.5922	$27.4500	$11.5502
PSUN	$17.7575	$1.5754	$5.1911	$29.8294	$11.8813
PSMU	$16.9398	$4.7204	$3.4269	$29.9882	$11.2295

Table 6.3 presents the estimated sample standard deviation of the estimators. In this Table we present the estimations only for the Nearest Matching Estimators and for the four sub-groups.[6]

Table 6.3

Bootstrapping Nearest Matching Estimators

A. Bootstraping: Nearest Matching Estimators

1. Usando PSTOT

Neighbors	H-Joven	H-Adulto	M-Joven	M-Adulta
5	$125.2746	$123.9793	$119.8345	$114.8197
10	$121.5326	$106.2886	$117.0516	$92.6762
20	$114.3734	$103.9255	$107.8788	$86.6682
30	$109.3103	$97.7808	$116.2953	$82.1266

[6] The estimated sample standard deviation for both the Epanechnikov kernel and the Nearest Matching estimator-whole sample can be obtained from the author upon request.

40	$109.6701	$100.6736	$110.3542	$80.8879
50	$108.2400	$104.8661	$114.2149	$84.3539

2. Usando PSMU

Neighbors	H-Joven	H-Adulto	M-Joven	M-Adulta
5	$114.0201	$148.4520	$132.0237	$85.4383
10	$117.0048	$134.7877	$118.8571	$82.1893
20	$113.5029	$126.8228	$132.6265	$80.8407
30	$112.4752	$157.4682	$136.7748	$80.6926
40	$114.7574	$127.1546	$129.2419	$78.3214
50	$115.9355	$129.3163	$122.3637	$83.5173

Computing the variation coefficient (Table 6.4) we can see that some of the estimated program impact are not very precise. In general, when we use PSTOT the variation coefficient present a smaller range of variation than when we use PSMU. In general, we can say that we have a much precise estimation for the cases of young males and adult females.

Table 6.4

Variation Coefficient

PSTOT

H-Joven	H-Adulto	M-Joven	M-Adulta
$6.41	$14.10	$12.73	$3.99
$5.84	$11.83	$26.11	$3.52
$6.82	$12.59	$19.38	$3.07
$5.95	$13.64	$24.33	$3.02
$4.89	$13.49	$24.51	$2.94
$5.02	$12.30	$20.88	$3.07

PSMU

H-Joven	H-Adulto	M-Joven	M-Adulta
$3.94	-$122.86	$15.27	$2.97
$3.64	$123.34	$9.27	$2.86
$4.49	$47.94	$14.42	$2.66
$4.45	$30.00	$18.12	$2.65
$6.97	$70.39	$16.37	$2.57
$5.65	$36.32	$15.45	$2.74

6.2 Employment Results

In this section we report the main Program impact on the probability of being employed. The main results are presented in Tables 6.5 and 6.6.

Table 6.5

Nearest Matching Estimators

A. Impact Estimates

1. PSTOT

Neighbors	M-Young	M-Adult	F-Young	F-Adult	All
5	0.0024	-0.0010	-0.0011	0.1340	0.0297
10	0.0088	-0.0119	-0.0161	0.1264	0.0218
20	-0.0022	-0.0153	-0.0084	0.1216	0.0284
30	0.00071	-0.0224	-0.0118	0.1210	0.0250
40	0.0094	-0.0237	-0.0126	0.1176	0.0249
50	0.0098	-0.0227	-0.0083	0.1185	0.0253

2. PSMU

Neighbors	M-Young	M-Adult	F-Young	F-Adult	All
5	0.0358	-0.0329	0.0080	0.1187	0.0322
10	0.0437	-0.0288	0.011	0.1187	0.0279
20	0.0263	-0.0257	0.0112	0.1223	0.0253
30	0.0235	-0.0217	0.0085	0.1220	0.0228
40	-0.0060	-0.0250	0.0064	0.1225	0.0266
50	0.0069	-0.0263	0.0059	0.1221	0.0272

Table 6.6

A. Bootstrapping: Nearest Matching Estimators

PSMU

Neighbors	M-Young	M-Adult	F-Young	F-Adult
5	0.4232	0.2669	0.6363	0.3038
10	0.3794	0.3065	0.4048	0.2915
20	0.3827	0.2862	0.3093	0.3025
30	0.3813	0.3107	0.3764	0.3268
40	0.3972	0.2974	0.3395	0.3301

50	0.3787	0.2919	0.5229	0.3272

From these tables it can be seen that the program impact results on the probability of employment vary across the sub-groups and they present substantial variability as judged by their sample standard error.

7. Cost Benefit Analysis

Based on the identification and quantification of the outcome measures, it is possible to estimate the benefits of Programa Joven, for the time period considered. This information is used in this section, together with information on costs of the program, to conduct a cost-benefit analysis and to calculate a rate of return to dollars spent on the Program.

We have information on:[1]

- **Direct Cost of Training:** This includes the cost of the training services rendered by the ICAP, the insurance for short-term stays in firms, and the fellowships and subsidies to the program beneficiaries for children.
- **Indirect Costs:** This includes the costs of personal, infrastructure, inputs and operational expenses of the unit that carries out the Program. It also includes, among others, information on bidding costs, promotion, computer services and supervision. The problem with the information is that the unit that carries out the Program has also other projects, however the Programa Joven is the most important in terms of expenditure.[2] This means that we will have to distribute these costs among the different projects (components) to have a reliable estimate of administrative costs of the Programa Joven.

The accumulated total cost within the period second semester 1993 to December 1998 has the following composition:

[1] This information was provided by the Programa Joven.
[2] Other components include Proyecto Microempresas, Proyecto Imagen y Fortalecimiento Institucional.

Table 7.1

Cumulative Budget Execution

Category	Cumulative	%
Direct Costs	$152,504,951.33	75.34%
Administration	$31,407,058.68	15.52%
Concurrent Costs	$5,417,166.29	2.68%
Financial Costs	$13,083,500.00	6.46%
Total	**$202,412,676.30**	**100%**

We did not have access to the detailed costs information needed to separate and allocate the Administrative costs among its several components (Programa Joven, Proyecto Microempresas, Proyecto Imagen and Fortalecimiento Institucional). As a compromise we assumed that the administrative, concurrent and financial costs maintain a constant proportionality with the direct costs (we call it constant c). Thus, we assume that the direct costs represent 3.055 times the indirect costs (3.055=Direct Costs/administration+concurrent+financial).

The Programa Joven has estimated the Direct Cost of the courses in the fifth bidding wave of the training program. They estimated in US$ 1,342 the direct cost per student graduated at least from the technical knowledge phase. Given the assumption of a constant proportionality between indirect to direct cost, this means that we have an indirect cost of US$ 483.83 per student graduated from at least the technical knowledge phase (this gives a total cost of US$ 1,780.83 per student).

We conducted the cost benefit analysis under different scenarios for the duration of benefits, discount rate, deadweight loss of taxes and ratio direct to indirect costs. The different values assumed for these variables are presented in the following Table:

Table 7.2

Duration of Benefits	Deadweight Loss of Taxes	Discount Rate	Ratio Indirect/Direct Costs
1 year	0	5%	0
3 years	0.5	10%	0.15
7 years	1.0	12%	0.327
8 years			

9 years			
Indefinite			

Some of the main cost benefits results obtained are presented in Table 7.3.[3] In the Table we present the Net Present Value (NPV) and the Internal Rate of Return (IRR) for several combinations of assumptions.

[3] Other results can be obtained from the autor upon request.

Table 7.3
Cost Benefit Analysis

Duration of Benefits	Deadweight Loss of Taxes	Discount Rate	Ratio Indirect/Direct Cost	Net Benefits	NPV	IRR
1 year	0	5%	0.00	$25.495	-$1,050.6	Negative
1 year	0	10%	0.00	$25.495	-$1,502.7	Negative
1 year	0	12%	0.327	$25.495	-$1,507.7	Negative
1 year	1	12%	0.00	$25.495	-$2,410.0	Negative
1 year	1	12%	0.327	$25.495	-$3,288.0	Negative
3 years	0	5%	0.00	$25.495	-$508.8	Negative
3 years	0.5	5%	0.00	$25.495	-$1,179.8	Negative
5 years	0	5%	0.00	$25.495	-$17.4	4.52
5 years	0	5%	0.327	$25.495	-$947.7	Negative
7 years	0	5%	0.00	$25.495	$428.3	13.25
7 years	0	5%	0.327	$25.495	-$10.6	4.85
8 years	0	5%	0.00	$25.495	$635.4	15.69
8 years	0	5%	0.327	$25.495	$196.5	7.66
8 years	0	10%	0.00	$25.495	$290.2	15.69
8 years	0	10%	0.327	$25.495	-$148.6	7.66
8 years	0	12%	0.327	$25.495	-$261.0	7.66
8 years	0.5	5%	0.00	$25.495	-$35.6	4.56
9 years	0	5%	0.00	$25.495	$832.6	17.42
9 years	0	10%	0.15	$25.495	$218.6	13.46
9 years	0	10%	0.327	$25.495	-$18.9	9.73
Infinite	0.5	10%	0.327	$25.495	$388.2	11.45

As it can be seen the NPV (IRR) can be positive or negative. Ceteris Paribus, the longer the time period for the benefits, the smaller the discount rate, the lower the ratio indirect to direct cost, and the smaller the deadweight loss of taxes, the greater the Net Present Value of the Programa Joven.

References

Hardle, W. 1992. “Applied Nonparametric Regression”.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. 1998. “Characterizing Selection Bias Using Experimental Data”. Econometrica. 66 (5).

Heckman, J., Ichimura, H., Smith, J., and Todd, P. 1997. “Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”. Review of Economic Studies. 64 (4).

Heckman, J., Ichimura, H., Smith, J., and Todd, P. 1998. “Matching As an Econometric Evaluation Estimator”. Review of Economic Studies. 65 (2): 605-654.

Heckman, J., LaLonde, R., and Smith, J. 1999. “The Economics and Econometrics of Active Labor Market Programs”. In: O. Ashenfelter and D. Card, editors. *Handbook of Labor Economics, Volume III*. Amsterdam: North-Holland. Forthcoming.

Heckman, J. and Smith, J. 1998. “Evaluating the Welfare State”. In: Steiner Strom, editor. *Econometrics and Economic Theory in the 20 th Century: The Ragnar Frisch Centennial*. Cambridge, United Kingdom: Cambridge University Press for Econometric Society.

Heckman, J. and Smith, J. 1999. “The Pre-Program Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies”. Economic Journal. 109 (143): 1-37.

Heckman, J., Smith, J., and Clements, N. 1997. “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts”. Review of Economic Studies. 64 (4): 487-536.

Jones, Marron & Sheather. 1996. “A Brief Survey of Bandwidth Selection for Density Estimation”, JASA.

LaLonde, Robert. 1986. “Evaluating the Econometric Evaluations of Training Programs with Experimental Data”. American Economic Review. 76 (4): 604-620.

Manski Charles F. 1995. “Learning about Social Programs from Experiments with Random Assignment of Treatments”. Madison, University of Wisconsin, Department of Economics, Institute for Research on Poverty.

Manski, C. F., and S.R. Lerman, 1977. “The Estimation of the Choice Probabilities from Choice-Based Samples”. Econometrica 45: 1977-1988.

Todd, P. 1999. “A Practical Guide to Implementing Matching Estimators”, mimeo presented at the IADB meeting in Santiago Chile, October.

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

“Base de datos e información del Programa Joven”. Ministerio de Trabajo y Seguridad Social de la República Argentina. Secretaría de Empleo y formación Profesional. Gerencia de Evaluación de Impacto y Estudios Especiales.

Appendix 1: Variables and descriptive statistics

Statistics

	N				
	Valid	Missing	Mean	Minimum	Maximum
EDAD	139584	152	23.0216	15.00	95.00
ESTADO	135651	4085	.4685	.00	1.00
SEXO	139584	152	.5590	.00	1.00
EDAD35	127644	12092	21.2118	16.00	35.00
HIJOS	139584	152	.2242	.00	1.00
HMENOR	139584	152	.1518	.00	1.00
VAESCU	139522	214	8.1E-02	.00	1.00
JEFE	139523	213	.1320	.00	1.00
ENPAREJA	139522	214	.2217	.00	1.00
PRINOCOM	139467	269	6.3E-02	.00	1.00
PRICOM	139047	689	.4129	.00	1.00
SENOCOM	139073	663	.4304	.00	1.00
SECOM	139073	663	9.0E-02	.00	1.00
DESOCUPA	130183	9553	.9503	.00	1.00
OCUPADO	139514	222	7.1E-02	.00	1.00
DESOEXP	129856	9880	.7769	.00	1.00
DESONEXP	129856	9880	.1734	.00	1.00
INACTIVO	130183	9553	4.5E-02	.00	1.00
POBRELP	138974	762	.7311	.00	1.00
GBA	133440	6296	.3610	.00	1.00
SUR	133440	6296	6.0E-02	.00	1.00
NEA	133440	6296	1.5E-02	.00	1.00
CENTRO	133440	6296	8.7E-02	.00	1.00
LITORAL	133440	6296	5.4E-02	.00	1.00
CUYO	133440	6296	4.1E-02	.00	1.00
NOA	133440	6296	7.1E-02	.00	1.00
CORDOBA	133440	6296	9.3E-02	.00	1.00
MENDOZA	133440	6296	9.8E-02	.00	1.00
STAFE	133440	6296	8.0E-02	.00	1.00
TUCUMAN	133440	6296	3.9E-02	.00	1.00
MUESTRA	3339	136397	1.0000	1.00	1.00
GRUPO	3339	136397	1.50	1	2

Variable Description

Variable	Description
EDAD	Age
ESTADO	1=Beneficiary, 0=Comparison
SEXO	1=Male, 0=Female
EDAD35	Dummy Age between 16 and 35 years of age
HIJOS	Children, 1=Yes, 0=No
HMENOR	Children younger than 5 years of age, 1=Yes, 0=No
VAESCU	School Attendance, 1=Yes, 0=No
JEFE	Head of the Household, 1=Yes, 0=No
ENPAREJA	Married, 1=Yes, 0=No
PRINOCOM	Primary Education Incompleted, 1=Yes, 0=No
PRICOM	Primary Education Completed, 1=Yes, 0=No
SENOCOM	Secondary Education Incompleted, 1=Yes, 0=No
SECOM	Secondary Education Completed, 1=Yes, 0=No
DESOCUPA	Unemployed, 1=Yes, 0=No
OCUPADO	Employed, 1=Yes, 0=No
DESOEXP	Unemployed with labor experience, 1=Yes, 0=No
DESONEXP	Unemployed without labor experience, 1=Yes, 0=No
INACTIVO	Out of the Labor Force, 1=Yes, 0=No
POBRELP	Poor by Income line, 1=Yes, 0=No
GBA	Reside in GBA, 1=Yes, 0=No
SUR	Reside in the South, 1=Yes, 0=No
NEA	Reside in the North East (NEA), 1=Yes, 0=No
CENTRO	Reside in the Center, 1=Yes, 0=No
LITORAL	Reside in the Coast, 1=Yes, 0=No
CUYO	Reside in Cuyo, 1=Yes, 0=No
NOA	Reside in the North West (NOA), 1=Yes, 0=No
CORDOBA	Reside in Córdoba, 1=Yes, 0=No
MENDOZA	Reside in Mendoza, 1=Yes, 0=No
STAFE	Reside in Santa Fe, 1=Yes, 0=No
TUCUMAN	Reside in Tucumán, 1=Yes, 0=No
MUESTRA	Internal Control Variable
GRUPO	Internal Control Variable

Appendix 2: Propensity Scores Estimation Universe

Young Males

Dependent Variable: ESTADO
Method: ML - Binary Logit
Date: 08/02/00 Time: 07:47
Sample(adjusted): 7 139732 IF SEXO=1 AND EDAD <21
Included observations: 39223
Excluded observations: 5715 after adjusting endpoints
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.803877	0.150947	11.95040	0.0000
EDAD	-0.071828	0.007738	-9.282715	0.0000
DESOCUPA	-0.255246	0.046744	-5.460558	0.0000
ENPAREJA	-0.224885	0.052757	-4.262654	0.0000
PRICOM	0.100371	0.037704	2.662087	0.0078
SENOCOM	0.125670	0.037495	3.351686	0.0008
GBA	-0.683742	0.028065	-24.36264	0.0000
CORDOBA	-0.369482	0.039529	-9.347130	0.0000
STAFE	-0.100391	0.037592	-2.670555	0.0076
TUCUMAN	0.262708	0.053469	4.913289	0.0000
MENDOZA	0.181558	0.038845	4.673859	0.0000
CUYO	0.604585	0.053702	11.25811	0.0000
SUR	-0.555716	0.048467	-11.46575	0.0000

Mean dependent var	0.537848	S.D. dependent var	0.498572
S.E. of regression	0.488527	Akaike info criterion	1.340416
Sum squared resid	9357.811	Schwarz criterion	1.343259
Log likelihood	-26274.57	Hannan-Quinn criter.	1.341316
Restr. log likelihood	-27074.83	Avg. log likelihood	-0.669876
LR statistic (12 df)	1600.538	McFadden R-squared	0.029558
Probability(LR stat)	0.000000		

Obs with Dep=0	18127	Total obs	39223
Obs with Dep=1	21096		

Dependent Variable: ESTADO
Method: ML - Binary Logit
Date: 08/02/00 Time: 07:47
Sample(adjusted): 7 139732 IF SEXO=1 AND EDAD <21
Included observations: 39223
Excluded observations: 5715 after adjusting endpoints
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	8754	6686	15440	0	0	0
P(Dep=1)>C	9373	14410	23783	18127	21096	39223
Total	18127	21096	39223	18127	21096	39223
Correct	8754	14410	23164	0	21096	21096
% Correct	48.29	68.31	59.06	0.00	100.00	53.78
% Incorrect	51.71	31.69	40.94	100.00	0.00	46.22
Total Gain*	48.29	-31.69	5.27			
Percent Gain**	48.29	NA	11.41			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	8769.19	9357.81	18127.00	8377.43	9749.57	18127.00
E(# of Dep=1)	9357.81	11738.19	21096.00	9749.57	11346.43	21096.00
Total	18127.00	21096.00	39223.00	18127.00	21096.00	39223.00
Correct	8769.19	11738.19	20507.37	8377.43	11346.43	19723.87
% Correct	48.38	55.64	52.28	46.22	53.78	50.29
% Incorrect	51.62	44.36	47.72	53.78	46.22	49.71
Total Gain*	2.16	1.86	2.00			
Percent Gain**	4.02	4.02	4.02			

*Change in "% Correct" from default (constant probability) specification

Adult Males

Method: ML - Binary Logit
Date: 08/02/00 Time: 08:11

Included observations: 27141
Excluded observations: 5928 after adjusting endpoints

Covariance matrix computed using second derivatives

Variable		Std. Error	z-Statistic	
C	0.543806	0.069319	7.844926	0.0000
EDAD	0.005235	0.001886	2.776131	0.0055
DESOEXP	-0.461252	0.046753	-9.865696	0.0000
ENPAREJA	-0.199429	0.029927	-6.663934	0.0000
HIJOS	-0.328496	0.048001	-6.843450	0.0000
SECOM	-0.174883	0.044582	-3.922745	0.0001
POBRELP	0.076984	0.028982	2.656263	0.0079
GBA	-0.730446	0.031747	-23.00811	0.0000
CORDOBA	-0.492786	0.046939	-10.49854	0.0000
STAFE	-0.113038	0.055298	-2.044147	0.0409
CUYO	0.425864	0.061929	6.876611	0.0000
SUR	-0.616546	0.060186	-10.24396	0.0000
LITORAL	-0.114879	0.054708	-2.099870	0.0357

Mean dependent var	0.462474	S.D. dependent var	0.498599
S.E. of regression	0.486813	Akaike info criterion	1.333704
Sum squared resid	6428.985	Schwarz criterion	1.337636
Log likelihood	-18086.03	Hannan-Quinn criter.	1.334971
Restr. log likelihood	-18736.19	Avg. log likelihood	-0.666373
LR statistic (12 df)	1300.334	McFadden R-squared	0.034701
Probability(LR stat)	0.000000		

Obs with Dep=0	14589	Total obs	27141
Obs with Dep=1	12552		

Dependent Variable: ESTADO

Date: 08/02/00 Time: 08:11
Sample(adjusted): 5 139734 IF SEXO=1 AND EDAD =>21

Excluded observations: 5928 after adjusting endpoints
Prediction Evaluation (success cutoff C = 0.5)

				Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	9562	5948	15510	14589	12552	27141
P(Dep=1)>C	5027	6604	11631	0	0	0
Total	14589	12552	27141	14589	12552	27141
Correct	9562	6604	16166	14589	0	14589
% Correct	65.54	52.61	59.56	100.00	0.00	53.75
% Incorrect	34.46	47.39	40.44	0.00	100.00	46.25
Total Gain*	-34.46	52.61	5.81			
Percent Gain**	NA	52.61	12.56			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	8160.08	6428.92	14589.00	7841.97	6747.03	14589.00
E(# of Dep=1)	6428.92	6123.08	12552.00	6747.03	5804.97	12552.00
Total	14589.00	12552.00	27141.00	14589.00	12552.00	27141.00
Correct	8160.08	6123.08	14283.15	7841.97	5804.97	13646.94
% Correct	55.93	48.78	52.63	53.75	46.25	50.28
% Incorrect	44.07	51.22	47.37	46.25	53.75	49.72
Total Gain*	2.18	2.53	2.34			
Percent Gain**	4.71	4.71	4.71			

*Change in "% Correct" from default (constant probability)

Dependent Variable: ESTADO
Method: ML - Binary Logit

Sample(adjusted): 16 139719 IF SEXO=0 AND EDAD <21
Included observations: 23760

Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

	Coefficient	Std. Error		Prob.
C	0.897656	0.195621	4.588746	0.0000
EDAD	-0.046777	0.010335	-4.526224	0.0000
DESOCUPA	-0.230151	0.059671	-3.856985	0.0001
ENPAREJA	-0.161329	0.049460	-3.261821	0.0011
HMENOR	0.394578	0.037022	10.65782	0.0000
VAESCU	-0.150914	0.047834	-3.154936	0.0016
SENOCOM	0.061177	0.027738	2.205563	0.0274
POBRELP	-0.057969	0.032472	-1.785222	0.0742
GBA	-0.555402	0.035853	-15.49104	0.0000
CORDOBA	-0.575251	0.053702	-10.71189	0.0000
MENDOZA	0.146671	0.048324	3.035153	0.0024
STAFE	-0.270651	0.051156	-5.290718	0.0000
TUCUMAN	0.264557	0.069825	3.788853	0.0002
SUR	-0.604611	0.064336	-9.397653	0.0000

Mean dependent var	0.399411	S.D. dependent var	0.489788
S.E. of regression	0.482600	Akaike info criterion	1.317109
Sum squared resid	5530.503	Schwarz criterion	1.321868
Log likelihood	-15633.26	Hannan-Quinn criter.	1.318653
Restr. log likelihood	-15985.06	Avg. log likelihood	-0.657965
LR statistic (13 df)	703.6103	McFadden R-squared	0.022008
Probability(LR stat)	0.000000		

Obs with Dep=0	14270	Total obs	23760
Obs with Dep=1	9490		

Method: ML - Binary Logit
Date: 08/02/00 Time: 07:57

Included observations: 23760
Excluded observations: 2835 after adjusting endpoints

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	12737	7593	20330	14270	9490	23760
P(Dep=1)>C	1533	1897	3430	0	0	0
Total	14270	9490	23760	14270	9490	23760
Correct	12737	1897	14634	14270	0	14270
% Correct	89.26	19.99	61.59	100.00	0.00	60.06
% Incorrect	10.74	80.01	38.41	0.00	100.00	39.94
Total Gain*	-10.74	19.99	1.53			
Percent Gain**	NA	19.99	3.84			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	8739.25	5530.75	14270.00	8570.41	5699.59	14270.00
E(# of Dep=1)	5530.75	3959.25	9490.00	5699.59	3790.41	9490.00
Total	14270.00	9490.00	23760.00	14270.00	9490.00	23760.00
Correct	8739.25	3959.25	12698.49	8570.41	3790.41	12360.82
% Correct	61.24	41.72	53.44	60.06	39.94	52.02
% Incorrect	38.76	58.28	46.56	39.94	60.06	47.98
Total Gain*	1.18	1.78	1.42			
Percent Gain**	2.96	2.96	2.96			

*Change in "% Correct" from default (constant probability) specification

Adult Females

Dependent Variable: ESTADO
Method: ML - Binary Logit
Date: 08/03/00 Time: 13:45
Sample(adjusted): 2 139736 IF SEXO=0 AND EDAD=>21
Included observations: 30285
Excluded observations: 4648 after adjusting endpoints
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.208297	0.063231	3.294210	0.0010
EDAD	-0.003668	0.001501	-2.444254	0.0145
DESOEXP	-0.200770	0.036979	-5.429268	0.0000
INACTIVO	0.150062	0.069234	2.167452	0.0302
HIJOS	0.246900	0.025801	9.569472	0.0000
SECOM	-0.084376	0.037290	-2.262667	0.0237
POBRELP	-0.054372	0.029652	-1.833693	0.0667
GBA	-0.764998	0.033466	-22.85918	0.0000
CORDOBA	-0.555462	0.046354	-11.98318	0.0000
STAFE	-0.288149	0.059562	-4.837836	0.0000
TUCUMAN	0.237946	0.075483	3.152330	0.0016
CENTRO	-0.189875	0.050503	-3.759706	0.0002
SUR	-1.002194	0.059135	-16.94763	0.0000
LITORAL	-0.153855	0.056331	-2.731264	0.0063
CUYO	0.188002	0.069233	2.715507	0.0066

Mean dependent var	0.400892	S.D. dependent var	0.490087
S.E. of regression	0.480277	Akaike info criterion	1.307728
Sum squared resid	6982.270	Schwarz criterion	1.311848
Log likelihood	-19787.26	Hannan-Quinn criter.	1.309049
Restr. log likelihood	-20393.06	Avg. log likelihood	-0.653368
LR statistic (14 df)	1211.582	McFadden R-squared	0.029706
Probability(LR stat)	0.000000		

Obs with Dep=0	18144	Total obs	30285
Obs with Dep=1	12141		

Dependent Variable: ESTADO

Date: 08/03/00 Time: 13:45
Sample(adjusted): 2 139736 IF SEXO=0 AND EDAD=>21

Excluded observations: 4648 after adjusting endpoints
Prediction Evaluation (success cutoff C = 0.5)

				Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	15321	8578	23899	18144	12141	30285
P(Dep=1)>C	2823	3563	6386	0	0	0
Total	18144	12141	30285	18144	12141	30285
Correct	15321	3563	18884	18144	0	18144
% Correct	84.44	29.35	62.35	100.00	0.00	59.91
% Incorrect	15.56	70.65	37.65	0.00	100.00	40.09
Total Gain*	-15.56	29.35	2.44			
Percent Gain**	NA	29.35	6.10			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	11161.02	6982.98	18144.00	10870.22	7273.78	18144.00
E(# of Dep=1)	6982.98	5158.02	12141.00	7273.78	4867.22	12141.00
Total	18144.00	12141.00	30285.00	18144.00	12141.00	30285.00
Correct	11161.02	5158.02	16319.05	10870.22	4867.22	15737.45
% Correct	61.51	42.48	53.88	59.91	40.09	51.96
% Incorrect	38.49	57.52	46.12	40.09	59.91	48.04
Total Gain*	1.60	2.40	1.92			
Percent Gain**	4.00	4.00	4.00			

*Change in "% Correct" from default (constant probability) specification
**corrected by equation

1. Young Males

```
      Total number of cases:      914 (Unweighted)
      Number of selected cases:   914
      Number of unselected cases: 0

      Number of selected cases:                 914
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 914


Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1261.4883

* Constant is included in the model.


 -2 Log Likelihood     1161.756
 Goodness of Fit        932.780

                     Chi-Square    df Significance

 Model Chi-Square        99.732     4        .0000
 Improvement             99.732     4        .0000

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed         +-------+-------+
   .00      0    I  215  I  200  I   51.92%
                 +-------+-------+
   1.00     1    I  111  I  390  I   77.92%
                 +-------+-------+
                            Overall  66.14%

---------------------- Variables in the Equation ------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

EDAD          -.5615     .0613  83.9248     1    .0000  -.2548     .5704
HMENOR        -.6147     .3412   3.2455     1    .0716  -.0314     .5408
VAESCU        -.3265     .1586   4.2369     1    .0396  -.0421     .7214
ENPAREJA       .6360     .2916   4.7572     1    .0292   .0468    1.8889
Constant     10.7092    1.1498  86.7448     1    .0000
```

2. Adult Males

```
      Total number of cases:      807 (Unweighted)
      Number of selected cases:   807
      Number of unselected cases: 0

      Number of selected cases:                 807
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 807

Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1115.6072

* Constant is included in the model.



 -2 Log Likelihood     1102.413
 Goodness of Fit        804.369

                     Chi-Square    df Significance

 Model Chi-Square        13.194     5        .0216
 Improvement             13.194     5        .0216

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed        +-------+-------+
   .00      0   I  255  I  137  I   65.07%
               +-------+-------+
   1.00     1   I  235  I  179  I   43.19%
               +-------+-------+
                          Overall  53.83%


---------------------- Variables in the Equation -----------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

PRICOM         .4706     .3293   2.0425     1    .1530   .0062    1.6010
SENOCOM        .5728     .3196   3.2114     1    .0731   .0330    1.7732
SECOM          .6529     .3387   3.7168     1    .0539   .0392    1.9211
OCUPADO        .4126     .1835   5.0537     1    .0246   .0523    1.5107
NEA            .6060     .3001   4.0763     1    .0435   .0431    1.8330
Constant      -.5903     .3068   3.7028     1    .0543
```

3. Mujeres Jóvenes

```
      Total number of cases:      587 (Unweighted)
      Number of selected cases:   587
      Number of unselected cases: 0

      Number of selected cases:                 587
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 587

Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   811.17762

* Constant is included in the model.


 -2 Log Likelihood      765.037
 Goodness of Fit        589.987

                    Chi-Square    df Significance

 Model Chi-Square       46.141     8        .0000
 Improvement            46.141     8        .0000

Classification Table for ESTADO
                   Predicted
                  .00    1.00     Percent Correct
                    0  I    1
Observed        +-------+-------+
   .00      0   I  134  I  139  I   49.14%
               +-------+-------+
   1.00     1   I   77  I  238  I   75.68%
               +-------+-------+
                         Overall  63.36%


---------------------- Variables in the Equation ---------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

EDAD          -.3696     .0803  21.2106     1    .0000  -.1539     .6910
VAESCU        -.4897     .2107   5.4037     1    .0201  -.0648     .6128
JEFE           .5733     .2999   3.6535     1    .0560   .0451    1.7741
SENOCOM        .3426     .2130   2.5857     1    .1078   .0269    1.4086
SECOM          .7220     .2792   6.6855     1    .0097   .0760    2.0585
DESOCUPA      -.5326     .2318   5.2783     1    .0216  -.0636     .5871
GBA           -.4980     .2816   3.1274     1    .0770  -.0373     .6078
NEA            .6789     .4110   2.7288     1    .0986   .0300    1.9717
Constant      7.2638    1.4923  23.6920     1    .0000
```

4. Mujeres Adultas

```
      Total number of cases:      844 (Unweighted)
      Number of selected cases:   844
      Number of unselected cases: 0

      Number of selected cases:                 844
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 844


Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1166.2809

* Constant is included in the model.


 -2 Log Likelihood     1153.692
 Goodness of Fit        844.660

                     Chi-Square    df Significance

 Model Chi-Square        12.589     3        .0056
 Improvement             12.589     3        .0056

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed        +-------+-------+
   .00      0   I   47  I  345  I   11.96%
               +-------+-------+
   1.00     1   I   29  I  424  I   93.66%
               +-------+-------+
                         Overall  55.78%

---------------------- Variables in the Equation ------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

JEFE           .3453     .1959   3.1082     1    .0779   .0308    1.4125
GBA           -.4251     .2162   3.8675     1    .0492  -.0400     .6537
NEA            .7667     .3473   4.8728     1    .0273   .0496    2.1526
Constant       .1076     .0804   1.7926     1    .1806
```

Expectation prediction tables

1. Young Males

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	215	111	326	0	0	0
P(Dep=1)>C	200	390	590	415	501	916
Total	415	501	916	415	501	916
Correct	215	390	605	0	501	501
% Correct	51.8%	77.84%	66.0%	0.0%	100.00%	54.7%
% Incorrect	48.2%	22.16%	33.95%	100.0%	0.00%	45.31%
Total Gain*	51.8%	-22.16%	11.35%			

*Change in "% Correct" from default (constant probability) specification

2. Adult Males

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	255	235	490	0	0	0
P(Dep=1)>C	137	179	316	392	414	806
Total	392	414	806	392	414	806
Correct	255	179	434	0	414	414
% Correct	65.1%	43.24%	53.8%	0.0%	100.00%	51.4%
% Incorrect	34.9%	56.76%	46.15%	100.0%	0.00%	48.64%
Total Gain*	65.1%	-56.76%	2.48%			

*Change in "% Correct" from default (constant probability) specification

3. Young Females

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	134	77	211	0	0	0
P(Dep=1)>C	139	238	377	273	315	588
Total	273	315	588	273	315	588
Correct	134	238	372	0	315	315
% Correct	49.1%	75.56%	63.3%	0.0%	100.00%	53.6%
% Incorrect	50.9%	24.44%	36.73%	100.0%	0.00%	46.43%
Total Gain*	49.1%	-24.44%	9.69%			

*Change in "% Correct" from default (constant probability) specification

4. Adult Females

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	47	29	76	0	0	0
P(Dep=1)>C	345	424	769	392	453	845
Total	392	453	845	392	453	845
Correct	47	424	471	0	453	453
% Correct	12.0%	93.60%	55.7%	0.0%	100.00%	53.6%
% Incorrect	88.0%	6.40%	44.26%	100.0%	0.00%	46.39%
Total Gain*	12.0%	-6.40%	2.13%			

*Change in "% Correct" from default (constant probability) specification

Appendix 4: Survey

1. Young Males

```
      Total number of cases:      915 (Unweighted)
      Number of selected cases:   915
      Number of unselected cases: 0

      Number of selected cases:                 915
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 915

Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1265.0636

* Constant is included in the model.


 -2 Log Likelihood     1166.141
 Goodness of Fit        920.862

                     Chi-Square    df Significance

 Model Chi-Square        98.923     4        .0000
 Improvement             98.923     4        .0000

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed         +-------+-------+
   .00      0    I  332  I  140  I   70.27%
                 +-------+-------+
   1.00     1    I  178  I  263  I   59.66%
                 +-------+-------+
                           Overall  65.15%

---------------------- Variables in the Equation ----------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

ZEDAD         -.5515     .0601  84.1487     1    .0000  -.2548     .5761
ZHMENOR       -.6161     .3416   3.2534     1    .0713  -.0315     .5400
ZENPARE        .6144     .2897   4.4963     1    .0340   .0444    1.8485
ZVAESC        -.3192     .1585   4.0557     1    .0440  -.0403     .7268
Constant     10.2690    1.1257  83.2126     1    .0000
```

2. Adult Males

```
      Total number of cases:      806 (Unweighted)
      Number of selected cases:   806
      Number of unselected cases: 0

      Number of selected cases:                 806
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 806

Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1112.0848

* Constant is included in the model.


 -2 Log Likelihood     1100.542
 Goodness of Fit        808.110

                     Chi-Square    df Significance

 Model Chi-Square        11.543     7        .1166
 Improvement             11.543     7        .1166

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed        +-------+-------+
   .00      0   I  379  I   65  I   85.37%
               +-------+-------+
   1.00     1   I  283  I   81  I   22.37%
               +-------+-------+
                          Overall  56.95%


---------------------- Variables in the Equation ------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

ZEDAD          -.0134     .0100   1.7871     1    .1813   .0000     .9867
ZOCUPA          .6600     .5600   1.3889     1    .2386   .0000    1.9347
ZDESOCU         .2204     .5411    .1659     1    .6838   .0000    1.2466
ZJEFE           .1860     .1562   1.4183     1    .2337   .0000    1.2044
ZPRINOC        -.4505     .3327   1.8338     1    .1757   .0000     .6373
ZSEINC          .0890     .1736    .2626     1    .6084   .0000    1.0931
ZSECOM          .1898     .2075    .8368     1    .3603   .0000    1.2090
Constant       -.2739     .6132    .1995     1    .6552
```

3. Young Females

```
   Total number of cases:      587 (Unweighted)
   Number of selected cases:   587
   Number of unselected cases: 0

   Number of selected cases:                 587
   Number rejected because of missing data:  0
   Number of cases included in the analysis: 587


Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   811.56545

* Constant is included in the model.

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      772.093
 Goodness of Fit        588.422

                     Chi-Square    df Significance

 Model Chi-Square        39.472     6        .0000
 Improvement             39.472     6        .0000

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed        +-------+-------+
   .00      0   I  212  I   98  I   68.38%
               +-------+-------+
   1.00     1   I  130  I  147  I   53.04%
               +-------+-------+
                          Overall  61.13%


---------------------- Variables in the Equation --------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

ZEDAD         -.3767     .0789  22.7646     1    .0000  -.1600     .6862
ZDESOCU       -.5693     .2284   6.2148     1    .0127  -.0721     .5659
ZJEFE          .6127     .2951   4.3120     1    .0378   .0534    1.8454
ZVAESC        -.5207     .2094   6.1847     1    .0129  -.0718     .5941
ZSEINC         .3522     .2125   2.7465     1    .0975   .0303    1.4222
ZSECOM         .7092     .2775   6.5303     1    .0106   .0747    2.0324
Constant      7.1545    1.4675  23.7702     1    .0000
```

4. Adult Females

```
      Total number of cases:      1031 (Unweighted)
      Number of selected cases:   1031
      Number of unselected cases: 0

      Number of selected cases:                 1031
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 1031


Dependent Variable..   ESTADO

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   1425.033

* Constant is included in the model.


 -2 Log Likelihood     1420.241
 Goodness of Fit       1031.232

                     Chi-Square    df Significance

 Model Chi-Square         4.792     5        .4418
 Improvement              4.792     5        .4418

Classification Table for ESTADO
                    Predicted
                   .00    1.00     Percent Correct
                     0  I    1
Observed        +-------+-------+
   .00      0   I  442  I  107  I   80.46%
               +-------+-------+
   1.00     1   I  366  I  115  I   23.93%
               +-------+-------+
                          Overall  54.07%


---------------------- Variables in the Equation ----------------------

Variable           B      S.E.     Wald    df      Sig       R    Exp(B)

ZJEFE          .1114     .1590    .4907     1    .4836   .0000    1.1178
ZHMENOR        .2201     .1266   3.0213     1    .0822   .0268    1.2462
ZSEINC        -.2744     .2448   1.2569     1    .2622   .0000     .7600
ZPRICO        -.1780     .2571    .4795     1    .4886   .0000     .8369
ZSECOM        -.2566     .2551   1.0113     1    .3146   .0000     .7737
Constant      -.0409     .2375    .0297     1    .8632
```

Expectation prediction tables

1. Young Males

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	332	178	510	472	441	913
P(Dep=1)>C	140	263	403	0	0	0
Total	472	441	913	472	441	913
Correct	332	263	595	472	0	472
% Correct	70.3%	59.64%	65.2%	100.0%	0.00%	51.7%
% Incorrect	29.7%	40.36%	34.83%	0.0%	100.00%	48.30%
Total Gain*	-29.7%	59.64%	13.47%			

*Change in "% Correct" from default (constant probability) specification

2. Adult Males

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	379	283	662	444	364	808
P(Dep=1)>C	65	81	146	0	0	0
Total	444	364	808	444	364	808
Correct	379	81	460	444	0	444
% Correct	85.4%	22.25%	56.9%	100.0%	0.00%	55.0%
% Incorrect	14.6%	77.75%	43.07%	0.0%	100.00%	45.05%
Total Gain*	-14.6%	22.25%	1.98%			

*Change in "% Correct" from default (constant probability) specification

3. Young Females

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	212	130	342	310	277	587
P(Dep=1)>C	98	147	245	0	0	0
Total	310	277	587	310	277	587
Correct	212	147	359	310	0	310
% Correct	68.4%	53.07%	61.2%	100.0%	0.00%	52.8%
% Incorrect	31.6%	46.93%	38.84%	0.0%	100.00%	47.19%
Total Gain*	-31.6%	53.07%	8.35%			

*Change in "% Correct" from default (constant probability) specification

4. Adult Females

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	442	366	808	549	481	1030
P(Dep=1)>C	107	115	222	0	0	0
Total	549	481	1030	549	481	1030
Correct	442	115	557	549	0	549
% Correct	80.5%	23.91%	54.1%	100.0%	0.00%	53.3%
% Incorrect	19.5%	76.09%	45.92%	0.0%	100.00%	46.70%
Total Gain*	-19.5%	23.91%	0.78%			

*Change in "% Correct" from default (constant probability) specification

Appendix 5: MATLAB Codes

Nearest Matching Estimators

```
%PROGRAM NEAREST MATCHING
%
%Developed by: Cristian Aedo (caedo@uahurtado.cl)
%
%Date: August 1, 2000
%Last Update: August 10, 2000
%
%Purpose: To estimate program impact using the Nearest
%Matching Estimator Approach.
%Subjects: Whole sample
%

%
%Loading and defining information matrix
%

clear;
load tresps2.dat;
m=tresps2;

%
%Defining Number of observations and location of
%beneficiaries and comparissons in the
%sample. Data set is ordered: first the beneficiaries and
%then the comparisons
%

n=3339;
n1=1670;
n2=1671;

%
%Transfering information matrix data into column vectors
%

dniclave=m(1:n,1);
grupos=m(1:n,2);
sexo=m(1:n,3);
subgrupo=m(1:n,4);
ing0=m(1:n,5);
ing1=m(1:n,6);
```

```
ing2=m(1:n,7);
ing3=m(1:n,8);
ing4=m(1:n,9);
ing5=m(1:n,10);
ing6=m(1:n,11);
ing7=m(1:n,12);
ing8=m(1:n,13);
ing9=m(1:n,14);
ocupa5=m(1:n,15);
desocu5=m(1:n,16);
inact5=m(1:n,17);
ocupa9=m(1:n,18);
desocu9=m(1:n,19);
inact9=m(1:n,20);
ocupa0=m(1:n,21);
desocu0=m(1:n,22);
inact0=m(1:n,23);
joven=m(1:n,24);
pstot=m(1:n,25);
psun=m(1:n,26);
psmu=m(1:n,27);

%
%Defining income data and propensity scores
%

yb=m(1:n1,5);
yc=m(n2:n,5);
pstotb=m(1:n1,25)/10000;
pstotc=m(n2:n,25)/10000;
psunb=m(1:n1,26)/10000;
psunc=m(n2:n,26)/10000;
psmub=m(1:n1,27)/10000;
psmuc=m(n2:n,27)/10000;

%
%Defining number of neighbors
%

neighbor=50;

%
%The following loop defines the comparissons which are
%going to be used for each beneficiaries. Then it
%calculates the average earnings for the number of
%neighbors considered.
%
```

```
for i=1:length(yb);

  difp=abs(pstotb(i)-pstotc);
  sortdifp=sort(difp);
  dist=sortdifp(neighbor);

  r=0;
  ycc=0;

  for j=1:length(yc);

     if difp(j) <= dist;

        r=r+1;
        ycc=ycc+yc(j);

     end;

  end;

     ycp(i)=ycc/r;
     ybb(i)=yb(i);

 end;

%
%Finally, we calculate the mean Program impact
%

imp=mean(ybb-ycp);
imp
```

Epanechnikov Kernel

```
%
%PROGRAM EPANECHNIKOV
%
%Developed by: Cristian Aedo (caedo@uahurtado.cl)
%
%Date: August 1, 2000
%Last Update: August 10, 2000
%
%Purpose: To estimate program impat estimates using the
%Epanechnikov Kernel
%Subjects: Whole sample
%

%
%Loading and defining information matrix
%

clear;
load tresps2.dat;
m=tresps2;

%
%Defining Number of observations and location of
%beneficiaries and comparissons in the
%sample. Data set is ordered: first the benefitiaries and
%then the comparissons
%

n=3339;
n1=1670;
n2=1671;

%
%Transfering information matrix data into column vectors
%

dniclave=m(1:n,1);
grupos=m(1:n,2);
sexo=m(1:n,3);
subgrupo=m(1:n,4);
ing0=m(1:n,5);
ing1=m(1:n,6);
ing2=m(1:n,7);
ing3=m(1:n,8);
ing4=m(1:n,9);
```

```
ing5=m(1:n,10);
ing6=m(1:n,11);
ing7=m(1:n,12);
ing8=m(1:n,13);
ing9=m(1:n,14);
ocupa5=m(1:n,15);
desocu5=m(1:n,16);
inact5=m(1:n,17);
ocupa9=m(1:n,18);
desocu9=m(1:n,19);
inact9=m(1:n,20);
ocupa0=m(1:n,21);
desocu0=m(1:n,22);
inact0=m(1:n,23);
joven=m(1:n,24);
pstot=m(1:n,25);
psun=m(1:n,26);
psmu=m(1:n,27);

%
%Defining income data and propensity scores
%

yb=m(1:n1,5);
yc=m(n2:n,5);
pstotb=m(1:n1,25)/10000;
pstotc=m(n2:n,25)/10000;
psunb=m(1:n1,26)/10000;
psunc=m(n2:n,26)/10000;
psmub=m(1:n1,27)/10000;
psmuc=m(n2:n,27)/10000;

%
%We use the Kernel with the adaptative estimate of spread
%(reference: Silverman,"Density Estimation for Statistics
%and Data Analysis", page 47).
%

for i=1:length(yb);

  difp=(pstotb(i)-pstotc);
  T=length(difp);
  s=std(difp);
  qr=(max(difp)-min(difp))/4;
  h=0.9*(T^(-1/5))*min([s (qr/1.34)]);
  sortdifp=sort(difp);
  f1=zeros(T,1);
```

```
  j=1;

  while j<T+1;
     au=abs(sortdifp(j,1)-sortdifp)/h;
     e1=au < sqrt(5);
     f1=e1.*(0.75*(1-(0.2*au.*au))/((5^0.5)*h*T))+f1;
     j=j+1;
  end;

  f1=f1 ./ sum(f1);
  matchyc(i)=f1'*yc;

end;

%
%Finally, we calculate the mean Program impact
%

imp=mean(yb-matchyc');
imp
```

Nearest Matching Estimator: Bootstrapping

```
%
%PROGRAM BOOTSTRAPPING FOR THE NEAREST MATCHING ESTIMATOR
%
%Developed by: Cristian Aedo (caedo@uahurtado.cl)
%
%Date: August 10, 2000
%Last Update: August 18, 2000
%
%Purpose: The Program will generate 100 paired samples of
beneficiaries and of comparissons (each of the samples will
%be of equal size as the original samples). For each of
%these 100 paired samples a Program Impact estimate will be
%obtained. The variance of the Mean Impact estimates will
%be computed as the sample analog using as a mean the
%original estimate of the Program Impact.
%
%Subjects: Whole sample
%

%
%Loading and defining information matrix
%

load madulta.dat;
m=madulta;

%
%Defining Number of observations and location of
beneficiaries and comparissons in the sample. Data set
%is ordered: first the benefitiaries and then the
%comparissons
%

n=3339;
n1=1670;
n2=1671;
nn1=1670;
nn2=1669;

%
%Transfering information matrix data into column vectors
%

dniclave=m(1:n,1);
grupos=m(1:n,2);
```

```
sexo=m(1:n,3);
subgrupo=m(1:n,4);
ing0=m(1:n,5);
ing1=m(1:n,6);
ing2=m(1:n,7);
ing3=m(1:n,8);
ing4=m(1:n,9);
ing5=m(1:n,10);
ing6=m(1:n,11);
ing7=m(1:n,12);
ing8=m(1:n,13);
ing9=m(1:n,14);
ocupa5=m(1:n,15);
desocu5=m(1:n,16);
inact5=m(1:n,17);
ocupa9=m(1:n,18);
desocu9=m(1:n,19);
inact9=m(1:n,20);
ocupa0=m(1:n,21);
desocu0=m(1:n,22);
inact0=m(1:n,23);
joven=m(1:n,24);
pstot=m(1:n,25);
psun=m(1:n,26);
psmu=m(1:n,27);

%
%Define some constant terms for the Random Number
Generator, number of neighbors and
%the mean for the bootstrap (these need to be changed).
%

p=2147483647.0;
q=2147483655.0;
r=16807.0;
neighbor=30;
meaneffe=0.1223;

%
%Obtain 200 seeds to initialize each random sample
%

nseeds = 200;
seed=20;

for i=1:nseeds;
```

```
   seed=MOD(r*seed,p);
   x(i,1)=seed/q;

end;

%
%Now iterate over each of these paired samples to obtain
the Program Estimate for each
%

for i=1:100;

   seed1=x(i,1);
   seed2=x(100+i,1);

   for j=1:nn1;

      seed1=MOD(r*seed1,p);
      x1=seed1/q;
      rut=round(x1*nn1+0.5);
      yb(j)=m(rut,21);
      pstotb(j)=m(rut,25)/10000;
      psunb(j)=m(rut,26)/10000;
      psmub(j)=m(rut,27)/10000;

   end;

   for j=1:nn2;

      seed2=MOD(r*seed2,p);
      x2=seed2/q;
      rut=n1+round(x2*nn2+0.5);
      yc(j)=m(rut,21);
      pstotc(j)=m(rut,25)/10000;
      psunc(j)=m(rut,26)/10000;
      psmuc(j)=m(rut,27)/10000;

   end;

%
%The following loop defines the comparissons which are
%going to be used for each beneficiaries. Then it
%calculates the average earnings for the
%number of neighbors considered.
%

for k=1:length(yb);
```

```
  difp=abs(psmub(k)-psmuc);
  sortdifp=sort(difp);
  dist=sortdifp(neighbor);

  s=0;
  ycc=0;

  for j=1:length(yc);

    if difp(j) <= dist;

      s=s+1;
      ycc=ycc+yc(j);

     end;

  end;

      ycp(k)=ycc/s;
      ybb(k)=yb(k);

 end;

 %
 %Calculate the mean Program impact
 %

 imp(i)=mean(ybb-ycp);

 end;

%
%Now we calculate the variance and the standard deviation
%of the mean Program Impact
%

 rimp=imp-meaneffe;
 boot=mean(rimp*rimp');
 neighbor, sqrt(boot)
```