Inter-American Development Bank
Banco Interamericano de Desarrollo
Latin American Research Network
Red de Centros de Investigación
Research Network Working Paper #R-483

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

by

Cristián Aedo*
Sergio Núñez**

*INACAP
**Graduate Program in Economics
ILADES/Georgetown University

November 2004

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Aedo, Cristián.
The impact of training policies in Latin America and the Caribbean : the case of Programa Joven / by Cristián Aedo, Sergio Núñez.

p. cm. (Research Network Working papers ; R-483)
Includes bibliographical references.

1. Youth—Employment—Argentina—Case study. 2. Programa Joven (Argentina)--Evaluation. I. Núñez, Sergio. II. Inter-American Development Bank. Research Dept. III. Latin American Research Network. IV. Series.

301.431 A772--dc21

©2003
Inter-American Development Bank
1300 New York Avenue, N.W.
Washington, DC 20577

The views and interpretations in this document are those of the authors and should not be attributed to the Inter-American Development Bank, or to any individual acting on its behalf.

The Research Department (RES) produces a quarterly newsletter, *IDEA (Ideas for Development in the Americas)*, as well as working papers and books on diverse economic issues. To obtain a complete list of RES publications, and read or download them please visit our web site at:
http://www.iadb.org/res

Contents

Abstract*

This paper evaluates Programa Joven, a training program conducted by Argentina's Ministerio del Trabajo. The paper adapts and applies a non-experimental evaluation methodology to answer the following questions: (1) Did Programa Joven increase the labor income of the trainees? (2) Did Programa Joven increase the probability of employment? (3) What was the rate of return to dollars spent on Programa Joven? The basic methodology used was the Matching Estimators approach. The application of this methodology requires two steps: first, the estimation of a model of program participation (propensity scores), and second, conditional upon the estimated propensity scores, the use of matching estimators to calculate the impact of the program. Three different information sources were used to estimate the propensity scores. These different information sources permitted the analysis of an additional question: how sensitive are program impact estimates to different propensity score specifications? This question has not been addressed by the previous literature but is addressed here. The paper hypothesizes that impact estimates are in fact sensitive to different propensity score specifications. Additionally, the paper reports and compares the propensity scores estimated from each of these data sources, and then estimates the program impact on earnings and employment based upon these propensity scores. Finally, the authors carry out a cost-benefit analysis of Programa Joven based upon cost information and program impact estimates (benefits).

* Financial support from the Interamerican Development Bank is gratefully acknowledged. This paper has been presented at the IADB research network seminars in Washington D.C. and Rio de Janeiro. We thank participants for helpful comments. We are particularly grateful to Jeffrey Smith, Petra Todd, James Heckman, and Gustavo Marquez for many helpful comments and suggestions. All remaining errors and omissions are our own. We acknowledge the cooperation of the Unidad de Estadísticas y Evaluación de Impacto, in providing the information used in this research. In particular, we are grateful to Mónica Muscolino, consultant of the Unidad de Estadística, and Sergio Diba, Coordinator of Programa Joven.

1. Introduction

Latin American countries invest a significant amount of resources in training programs. Thorough evaluations of these efforts are needed to gauge their strengths and weaknesses and facilitate the design of more effective programs in the future. Effective program evaluation, however, faces many challenges (Heckman, LaLonde and Smith, 1999). First, due to the varying impacts that programs produce, there are many parameters of interest in their evaluation. Second, no standard method for conducting program evaluations exists. The choice of an appropriate program impact estimate depends on the question to be answered and the availability of data. Third, "good" evaluations need "good data." In most cases, econometric methods must be used to correct problems with data. Fourth, program impact estimates require a pool of comparable individuals, which complicates the program evaluation process. It is necessary to reduce possible biases by administering similar questionnaires to participants and non-participants, by using similar time frames, and by drawing the samples of participants and non-participants from similar labor markets. Fifth, non-experimental program impact estimates solve the selection problem under different assumptions, which generates variability in their results. An experimental evaluation provides an important reference framework to analyze the performance of alternative non-experimental evaluation methodologies. Sixth, social programs at the national or regional levels have an impact on both participants and non-participants. The usual method of accounting for such program "contamination" is to assume that the impact on non-participants is not significant.

This paper evaluates Programa Joven, a training program conducted by Argentina's Ministerio del Trabajo. We adapt and apply a non-experimental evaluation methodology to answer the following questions: (1) Did Programa Joven increase the labor income of the trainees? (2) Did Programa Joven increase the probability of employment? (3) What was the rate of return to dollars spent on Programa Joven?

The Matching Estimators approach was used as the basic methodology to answer these questions. This choice was based upon both the theoretical developments in the area of Program Evaluation[1] and the availability and quality of relevant information. As described in Todd (1999), the application of this methodology requires two steps: first, the estimation of a model of

[1] See Heckman, Smith and Clements (1997); Heckman, Ichimura and Todd (1997); Heckman and Smith (1998); Heckman, Ichimura and Todd (1998); and Heckman, Ichimura, Smith and Todd (1998).

program participation (propensity scores), and second, conditional upon the estimated propensity scores, the use of matching estimators to calculate the impact of the program.

In order to estimate the propensity scores, three sources of information were used: first, the data for all individuals (approximately 140,000) who registered and qualified to take training programs in the period from March 1996 to December 1997; second, a sample of beneficiaries and comparisons used by the Ministerio del Trabajo to evaluate the program (3,340 individuals in total)[2] with information gathered at their registration; and finally, the same sample of 3,340 individuals, but with information gathered at a survey conducted one year after completion of the program. These different information sources permitted the analysis of an additional question: how sensitive are program impact estimates to different propensity score specifications? This question has not been addressed by the previous literature but is addressed here. The authors hypothesize that impact estimates are in fact sensitive to different propensity score specifications.

Additionally, the paper reports and compares the propensity scores estimated from each one of these data sources, and then estimates the program impact on earnings and employment based upon these propensity scores.

Finally, the authors carry out a cost-benefit analysis of the Programa Joven based upon cost information and the program impact estimates (benefits). This analysis was conducted under different scenarios with regard to benefit duration, discount rate and the ratio of indirect to direct cost.

2. Description of Programa Joven

Programa Joven offered training to improve recipients' participation in the formal labor market. To this end, the program provided both intensive training for positions in the productive sector of the economy and internships in firms.

The target population of the program was young people from poor households, with a low education level, no working experience, and who were unemployed, underemployed or inactive. To qualify for Programa Joven, an applicant must have been at least 16 years of age, completed no more than secondary school, been a member of a household considered poor,

[2] These individuals were drawn from the universe of potential trainees (first data source).

and not been a participant in the labor market.[3] The program provided the following benefits: an average of 200 hours of training, transportation expenses, a subsidy for females with young children, medical checkups, books, material and work clothing.

The duration of the training program varied from 14 to 20 weeks. The training was intensive and divided into two main activities: *Technical Knowledge Phase*, lasting between 6 to 12 weeks, in which the beneficiaries were taught knowledge and technical skills for a particular occupation, and *Internships Phase*, lasting eight weeks, in which the beneficiaries complemented their technical knowledge with work in the fields in which they received training.

The criteria for selecting firms for internships were: general characteristics of the firm; tasks to be performed by the trainees; personnel involved in similar positions in the firm; equipment; and supplies and infrastructure.

To carry out the training, the Ministerio del Trabajo hired Instituciones de Capacitación (ICAP) through an international bidding process. The distribution of the training activities at the national level was determined in accordance with regional populations. The program administrator was the Ministerio del Trabajo y Seguridad Social (now the Ministerio de Trabajo, Empleo, y Seguridad Social) through the Secretaría de Empleo y Capacitación Laboral. The program operated until the year 2000 through joint funding from the Inter-American Development Bank (IDB) and the government of Argentina.

Programa Joven's impact has been evaluated previously. The Unidad de Estadísticas y Evaluación de Impacto of the Programa Joven undertook a 1993 study on the impact of a group of courses taught in the first round of the program; a 1996 study of second- and third-round courses; and a 1998 study (with a sample of 3,340 individuals)[4] of fifth-round courses. These evaluations compared beneficiaries and "comparisons" by assuming that they were random samples of the population. Although the evaluations attempted to perform some matching on observables, they did not adjust properly for sample selection problems.

[3] See below for a more detailed description of what constitutes a poor household.
[4] In Section 3, we describe this sample in more detail.

3. Description of the Data

Two data sets were used in the evaluation. The first data set comprises raw data on the 139,732 individuals who registered and qualified for training programs during the period from March 1996 to December 1997. The second data set consists of two samples, one for beneficiaries and one for controls, of 1,670 individuals each, which the Ministerio del Trabajo used to evaluate the impact of the fifth round of the training program. These samples were extracted from the universe of 139,732 individuals in the first data set.

The "Acreditación" was an applicants' first contact with Programa Joven, which enabled them to register and participate in a training class. Thus, the "Acreditado" category consists of individuals who were eligible for participation in the program but may or may not have taken a class. The Beneficiary category corresponds to individuals who completed at least the Technical Knowledge phase.

3.1 Universe Information: Data on "Acreditados"

The authors obtained and processed raw information on 139,732 "Acreditados" who registered and qualified to take training programs in the period from March 1996 to December 1997. Eighty percent of the "Acreditados" presented complete records, but the remaining 20 percent required that information be completed on a case-by-case basis.[5] Some individuals were excluded from the analysis because of missing information. The information was used to obtain the probability of program participation (Propensity Scores). Later in this paper, we will expand on the information we constructed and the manner in which it was used.

3.2 Sample Information: Data on Beneficiaries and Comparisons

We have data on two samples of 1,670 individuals each, one for Beneficiaries (participant group) and one for Comparisons (comparison group). These were used by the Ministerio del Trabajo to evaluate the impact of the fifth generation of the training program. Both groups consist of individuals who meet the selection criteria for consideration as potential participants in the program *(*"Acreditados"). In addition, the Beneficiaries are those who actually completed the Technical Knowledge phase.

[5] We used other data available from Programa Joven to complete these records.

The comparison group was not selected at random. 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 force participation, socioeconomic level, and parent of children 5 years old or younger. For both samples, there is information for the period from the *Acreditacion* to 12 months after the Beneficiaries finished the training (the follow-up information for the comparisons was obtained at the same time as the information for the beneficiaries). This information allows us to construct the individual labor history for both samples.

3.2.1 Beneficiaries Sample

This sample was designed by the Ministerio del Trabajo to include statistical representation by gender and region of residence. The first variable was introduced to study the program's impact given different labor market conditions for males and females. The second variable was introduced to study the different impacts of socioeconomic characteristics and regional labor markets on program outcomes. In total, the Ministerio del Trabajo considered 11 geographic units designated as "regions."

To define the sample size, the observed variations in the values for variables such as proportion of employed to unemployed workers and employed workers' average income were considered. These variables present the greatest variation among the outcomes variables. A percentage of non-response of 5 percent was considered. 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 percent, with a risk level of 1 percent. In other words, the interval (P-0.1, P+0.1) contains the estimated "p," of the population proportion P, with probability 0.95.

3.2.2 Comparison Sample

Once the Beneficiaries sample was obtained, a comparison sample was constructed. For each beneficiary, a "twin" was selected from among people who 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 who presented similar socioeconomic characteristics 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; second, education level and children. In cases in which an "identical" individual could not be found, a replacement that most closely matched the beneficiaries' socioeconomic and geographic characteristics was used. This procedure generated a sample that is identical in terms of region and sex, but presents some differences in terms of education level.

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 propensity score Matching Estimator methodology. Three models of program participation were estimated. The first model estimated the probability of program participation conditioned on eligibility by using the universe of 139,732 "Acreditados." These estimated propensity scores are denoted by PSTOT. The second model was restricted to the sample of 3,340 individuals and used the information provided at the "Acreditación." These are denoted as PSUN. Finally, the third model used the sample of 3,340 individuals and the information available at the end of the period. These are denoted as PSMU.

It is important to define the eligibility requirements in greater detail. An individual was eligible if:

- *Housing*: The individual did not live in a house *or* if the house he/she lived in did not have a bathroom *or* if the house he/she lived in was "crowded" (more than 3 people per room).
- *Income*: Per capita household income was below US$120 per month.
- *Labor Status*: The individual was searching for a job *or* she/he worked for a wage under US$200 a month *or* she/he was head of the household and her/his labor income was below US$400 a month and she/he was looking for a new job *or* she/he was neither working nor searching for a job but she/he wished to work.
- *Living Situation*: A ratio of head of the household to number of dependents less than 0.25 in houses in which the head did not complete primary education.

The authors attempted to obtain information related to the "pre-acreditación" 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 these types of questions because individuals did not have incentives (in fact, in some cases they had disincentives) to answer truthfully and there was no readily available mechanism to verify their answers. For this reason, the information gathered at the "Acreditación" did not include information on labor history.

The information on the universe of "Acreditados" was useful because it allowed us to construct the "Program History" of every "Acreditado." As previously mentioned, the Programa Joven comprised a Technical Knowledge phase and an Internships phase. Consequently, an "Acreditado" may have been in different circumstances: 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, and she/he may have started the Internships phase but she/he may or may not have finished. In fact, there was enough information to classify the universe of "Acreditados" by the following mutually exclusive categories:

- *"Acreditado" only*: Individuals who were eligible for training programs but had not started the Technical Knowledge phase.
- *Incomplete Technical Knowledge phase*: Individuals who did not complete the Technical Knowledge phase because of a justifiable reason (family problems, pregnancy, found employment, etc.).
- *Technical Knowledge phase Drop-out*: Individuals who quit the Technical Knowledge phase without a justifiable reason.
- *Did not reach the Technical Knowledge phase*: Individuals who did not meet the minimum standards required for approval.
- *Incomplete Internships phase*: Individuals who did not finish the Internships phase because of a justifiable reason (family problems, pregnancy, obtained a job, etc.).
- *Internships phase Drop-out*: Individuals who quit the Internships phase without a justifiable reason.

- *Did not reach the Internship phase*: Individuals who did not meet the minimum standards required for approval.
- *Completes*: Individuals who successfully completed both phases.

Empirically, around 52 percent of the "Acreditados" were in the *"Acreditado" only* category, and 37 percent of the "Acreditados" were in the *Completes* category. Another useful piece of information is the type of training program undertaken by the Beneficiaries.[6] This information will be used in future research to address the issues of recent participation and trainees' progression in the program.

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). An individual was considered a Beneficiary if she/he had successfully completed the Technical Knowledge phase and 0 otherwise. This choice allowed for the use of most of the "Acreditados" (around 89%), and is consistent with the way in which the Ministerio del Trabajo obtained its samples of Beneficiaries and Controls.

The following individual dimensions, which were measured at the time of accreditation, 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:* An index of unmet basic needs. This index considers an individual poor if the person lives in a special home (minors, or unmarried mothers) *or* if the house they live in does not have a bathroom *or* if the house they live in is "crowded" (more than 3 people per room) *or* if the ratio of head of the household to number of dependents is smaller than 0.25 and the head of the household did not complete primary school education. The Poverty Line criterion,

[6] Tertiary Sector (educational services, administration and accounting, assistant of firms and services, dental assistant, retirement services, computation, food, hotel and tourism, janitor and maintenance, media and publicity, photography, hairdressing, sales, telephony, surveillance); industrial sector (construction, quality control,

using as reference the income level of the individuals at the moment of "Acreditación" and a poverty line of $120 per month, was also considered.

- *Sociodemographic Dimension:* Gender and age were used.
- *Education and Marital Status Dimension:* Several indicators measuring years of education completed and school attendance at the moment of "Acreditación" were used. Marital status was measured by whether the individual was married or single, whether he/she had children (especially young children) and whether the individual was or was not the head of the household.
- *Geographical Dimension:* The same 11 regions were used as in the Ministerio del Trabajo's evaluation samples.

Table 4.1. Participation by Region

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: Adult Males, ages 21 to 35;[7] Young Males, younger than 21; Adult Females, ages 21 to 35; Young Females, younger than 21.

electronics, textiles, chemical laboratories, auto mechanic, industrial painting, plastic, refrigeration, graphic industry); agricultural, forest and mining (gardening, cultivation, watering, mining, cattle production).

[7] We had some cases of beneficiaries older than 35 years of age. These were included as adult males and females.

4.2 Strategic Behavior

The authorities in charge of the Program suspected that individuals followed strategic behavior to become eligible for the Program. However, the authorities did not have a readily available mechanism to verify the information provided by the individuals at the "Acreditación." To address this issue, the authors compared information available at the "Acreditación" with some information provided in the survey of the 3,340 individuals in the beneficiaries and comparisons groups conducted twelve months after the Program's end.[8] The questions refer to their labor status at the time of the "Acreditación."[9]

Tables 4.2 to 4.4 present cross-information about unemployment status (1 is unemployed and 0 is 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 rows while the information related to "Survey" is presented in columns.

Table 4.2. Unemployment for All Individuals*

		"Survey"		
		0	**1**	
"Acreditación"	**0**	37 (20.7%)	142 (79.3%)	179 (100%)
	1	545 (17.2%)	2,616 (82.8%)	3,161 (100%)
		582 (17.4%)	2,758 (82.6%)	3,340 (100%)

**Row percentages are presented.*

Table 4.3. Unemployment for Beneficiaries*

		"Survey"		
		0	**1**	
"Acreditación"	**0**	26 (29.5%)	62 (70.5%)	88 (100%)
	1	285 (18.0%)	1,299 (82.0%)	1,584 (100%)
		311 (18.6%)	1,361 (81.4%)	1,672 (100%)

**Row percentages are presented.*

[8] For the rest of the paper, the information obtained in the final survey of the 3,340 individuals will be referred to as "Survey."

[9] The requested "Acreditación" information was limited and refers primarily to labor status prior to the program.

Table 4.4. Unemployment for Comparisons*

		"Survey" 0	1	
"Acreditación"	0	11 (12.1%)	80 (87.9%)	91 (100%)
	1	260 (16.5%)	1,317 (83.5%)	1,578 (100%)
		271 (16.2%)	1,397 (83.8%)	1,668 (100%)

**Row percentages are presented.*

Individuals who declared themselves unemployed at the "Acreditación" but not unemployed at the "Survey" were considered to have behaved strategically at the "Acreditación." Using this as an indicator, Table 4.2 shows that 542 out of 3,160 (17.2 percent) individuals declared themselves unemployed at the "Acreditación," and were thus "misbehaving." Separating between beneficiaries and comparisons (Tables 4.3 and 4.4), the percentages of "misbehaviors" were 18.0 percent and 16.5 percent, respectively.

However, this evidence is inconclusive because it could simply indicate measurement error. What might suggest strategic behavior is asymmetric measurement error, in which one type of disagreement between the two values is more common than the other type. In particular, one would expect to find the difference occurring in the direction consistent with the strategic incentives facing the agent. This was not the case: 79.3 percent of those who declared themselves employed at the "Acreditación" declared themselves unemployed at the "Survey" (70.5 percent and 87.9 percent for beneficiaries and comparisons, respectively). Moreover, one might also expect the asymmetry to be stronger for beneficiaries than for comparisons, which was not the case.

Based upon these (admittedly rather limited) indicators, there is no evidence of strategic behavior at the "Acreditación."

5. Estimation of Program Participation (Propensity Scores)

The authors estimated different logit models for each of the four subgroups: Young Males, Young Females, Adult Males and Adult Females. A description of the main variables is presented in Appendix 1. The variables related to an individual's eligibility for the Program were

always controlled for in these estimations (regardless of their statistical significance).[10] Statistical significance was used to determine if the other explanatory variables remained in the logit estimations.

As previously mentioned, the authors conducted three estimations for the Propensity Scores. The first one used the individuals and the information available at "Acreditación" (139,732 cases). The second used the information provided at "Acreditación," but only considered the individuals in the "Survey" (3,340 cases). The third used both the individuals and the information available at the "Survey" (3,340 cases).

In the second and third cases it was necessary to re-weight the sample prior to the econometric work since the sample, by design, contained equal percentages of beneficiaries and comparisons.[11]

5.1 Universe Information (PSTOT)

The main econometric results for the binary Logits and their prediction tables for the four groups can be seen in Tables 5.1 and 5.2.

There will be no discussion of the estimated coefficients of the variables included in the regressions because of eligibility considerations. With regard to the others, it seems that age and the presence of a spouse or companion are, in general, related to a lower likelihood of program participation. The presence of children was positively related to program participation for females. The presence of children was positively related to program participation for adult males, but for males this variable was not significant. The variables related to unemployed with working experience and school attendance were significant and negatively related to program participation only in the cases of young females and adult females, respectively. The regional dummies for GBA, Cordoba, Santa Fe, Sur, Litoral (significant only in the cases of adult males and females) and Centro (significant only for adult females) were negatively related to propensity scores. Tucuman, Mendoza and Cuyo dummies were positively related to propensity scores (although they were not significant in all the sub-samples).

[10] These variables are: Pobrelp, Desocupa, Inactivo, Jefe and the educational dummies.
[11] To re-weight, we followed Manski and Lerman (1977).

Table 5.1. PSTOT Propensity Scores*

	Young Males	Adult Males	Young Females	Adult Females
Constant	1.6333 (0.2307)	0.1951 (0.1639)	1.2739 (0.3658)	0.0338 (0.1806)
Pobrelp**	0.0190 (0.0239)	0.0791 (0.0290)	-0.0599 (0.0326)	-0.0553 (0.0297)
Desocupa**	-0.1827 (0.1662)	-0.2250 (0.1506)	-0.5210 (0.2997)	0.0857 (0.1730)
Inactivo**	0.0745 (0.1727)	0.3597 (0.1681)	-0.3000 (0.3052)	0.2299 (0.1797)
Jefe**	-0.0673 (0.0710)	-0.0182 (0.0460)	0.0709 (0.0883)	-0.0845 (0.0353)
Prinocom**	0.0866 (0.0701)	0.2087 (0.0646)	0.0051 (0.0848)	0.0432 (0.0576)
Pricom**	0.1545 (0.0551)	0.1516 (0.0465)	-0.0541 (0.0516)	0.1146 (0.0399)
Senocom**	0.1799 (0.0548)	0.1585 (0.0467)	0.0219 (0.0500)	0.0614 (0.0396)
Edad	-0.0701 (0.0078)	0.0045 (0.0019)	-0.0493 (0.0106)	-0.0031 (0.0015)
Enpareja	-0.1890 (0.0652)	-0.1976 (0.0441)	-0.1623 (0.0495)	
Hijos		-0.3264 (0.0483)	0.3937 (0.0376)	0.2536 (0.0261)
Desoexp				-0.2031 (0.0377)
Vaescu			-0.1542 (0.0479)	
Gba	-0.6834 (0.0284)	-0.7433 (0.0317)	-0.5546 (0.0359)	-0.7666 (0.0335)
Cordoba	-0.3667 (0.0404)	-0.0495 (0.0469)	-0.5767 (0.0538)	-0.5579 (0.0464)
Stafe	-0.0992 (0.0385)	-0.1145 (0.0554)	-0.2702 (0.0511)	-0.2863 (0.0596)
Tucuman	0.2581 (0.0535)		0.2708 (0.0699)	0.2314 (0.0755)
Mendoza	0.1826 (0.0393)		0.1499 (0.0485)	
Cuyo	0.6025 (0.0537)	0.4201 (0.0619)		0.1846 (0.0693)
Sur	-0.5519 (0.0488)	-0.6313 (0.0602)	-0.6070 (0.0644)	-0.9940 (0.0593)
Litoral		-0.1135 (0.0549)		-0.1473 (0.0566)
Centro				-0.1918 (0.0506)
Observations	39,119	27,215	23,758	30,278
Log Likelihood	-26,202.45	-18,151.11	-15,630.66	-19,776.20
Restr. Log Lik.	-27,004.27	-18.786.79	-15,984.04	-20,387.66

** Standard Errors in parentheses.*

*** Variables related to eligibility criteria.*

To analyze the fit of the model, the authors used the MacFadden R-Squared and the prediction evaluation of the estimated equation (using a success cut-off of 50 percent) versus a constant probability model. The following table presents this evaluation for the four sub-samples:

Table 5.2. PSTOT Prediction Evaluation

Group	R-squared	Prediction Evaluation
Young Males	0.0299	% Correct ranges from 53.76 to 59.06
Adult Males	0.0338	% Correct ranges from 53.76 to 59.57
Young Females	0.0221	% Correct ranges from 60.06 to 61.64
Adult Females	0.0300	% Correct ranges from 59.91 to 62.41

The predicted Propensity Scores vary 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 (PSUN)

This case considers the 3,340 individuals who participated in the "Survey" but uses the information provided by them at the "Acreditación."[12] As mentioned before, the sample was re-weighted prior to estimation to correct for Choice-Based Sampling. Following Manski and Lerman (1977), the authors re-weighted each observation by the ratio of the proportion of beneficiaries in a random population divided by the proportion of beneficiaries in the sample. For the former, the authors used the universe information to estimate the proportion of beneficiaries in the universe, and for the latter, the authors used the sample proportion of beneficiaries.

The main econometric results for the binary Logits, and their prediction tables for the four groups, can be seen in Tables 5.3 and 5.4.

[12] Given that the information at "Acreditación" was taken around two years before the "Survey," the distribution of the individuals by gender in the young and adult groups do not match the classification in the sample. This explains the different sizes of the groups in this information and the "Survey" information.

Table 5.3. PSUN Propensity Scores*

	Young Males	Adult Males	Young Females	Adult Females
Constant	11.3384 (1.2763)	0.4506 (0.2403)	6.4998 (1.4969)	0.0578 (0.2292)
Pobrelp**	-0.0126 (0.1427)	-0.0991 (0.1484)	0.0866 (0.1781)	-0.0176 (0.1374)
Desocupa**	-0.1654 (0.2098)	-0.3939 (0.1848)	-0.5446 (0.3622)	0.0199 (0.1977)
Inactivo**	-0.2352 (0.3388)	-0.6296 (0.5545)	0.0173 (0.4385)	0.1221 (0.3060)
Jefe**	-0.0908 (0.1976)	0.1310 (0.1484)	0.5552 (0.2986)	0.1086 (0.1597)
Prinocom**	-0.4856 (0.4064)	-0.6572 (0.3447)	-0.5280 (0.5318)	0.0738 (0.2696)
Pricom**	-0.3008 (0.2668)	-0.1827 (0.2086)	-0.2845 (0.2987)	0.0700 (0.1779)
Senocom**	-0.1063 (0.2268)	-0.0769 (0.1890)	-0.0161 (0.1765)	-0.0113 (0.1571)
Edad	-0.5771 (0.0634)		-0.3138 (0.0768)	
Enpareja	0.7099 (0.3126)			
Hijos	-0.6201 (0.3455)			
Vaescu	-0.3995 (0.1679)		-0.3498 (0.2021)	
Nea		0.6099 (0.3010)	0.6320 (0.4096)	0.7889 (0.3082)
Gba			-0.5082 (0.2809)	-0.2215 (0.1908)
Observations	914	807	587	1031
Log Likelihood	-579.327	-550.585	-385.333	-707.496
Restr. Log Lik.	-630.744	-557.8036	-405.589	-712.633

** Standard Errors in parentheses.*
*** Variables related to eligibility criteria.*

As before, there will be no discussion of the estimated coefficients of the variables included in the regressions because of eligibility considerations. With regard to the others, it seems that age, the presence of children and school attendance are related to a lower likelihood of program participation (although these effects are not significant for all the subgroups). The presence of a spouse or companion was positively related to program participation only for young males. The

regional dummies for Nea and Gba were positively and negatively related to Propensity Scores, respectively.

Table 5.4 presents the MacFadden R-Squared and the prediction evaluation of the estimated equation (using a success cut-off of 50 percent) versus a constant probability model:

Table 5.4. PSUN Prediction Evaluation

Group	R-squared	Prediction Evaluation
Young Males	0.0815	% Correct ranges from 54.8 to 67.1
Adult Males	0.0129	% Correct ranges from 51.1 to 54.4
Young Females	0.0499	% Correct ranges from 53.7 to 61.7
Adult Females	0.0072	% Correct ranges from 52.9 to 53.9

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 (PSMU)

Finally, the authors considered the 3,340 "Survey" cases and the information provided at the "Survey." Tables 5.5 and 5.6 present the main econometric results for the binary Logits and their prediction tables. As mentioned before, the sample was re-weighted prior to estimation to correct for Choice-Based Sampling.

Table 5.5. PSMU Propensity Scores*

	Young Males	Adult Males	Young Females	Adult Females
Constant	6.911 (1.004)	0.420 (0.366)	4.240 (1.240)	-0.240 (0.236)
Pobrelp**	-0.064 (0.130)	0.177 (0.158)	0.178 (0.160)	0.011 (0.147)
Desocupa**	-0.236 (0.190)	-0.550 (0.329)	-0.529 (0.332)	-0.030 (0.202)
Inactivo**	-0.286 (0.314)	-0.058 (0.397)	-0.085 (0.401)	0.137 (0.322)
Jefe**	0.122 (0.163)	0.194 (0.253)	0.250 (0.256)	0.200 (0.165)
Prinocom**	-0.400 (0.363)	-0.372 (0.476)	-0.625 (0.485)	0.173 (0.275)
Pricom**	-0.229 (0.225)	-0.290 (0.233)	-0.586 (0.249)	0.119 (0.190)
Senocom**	-0.137 (0.198)	-0.150 (0.194)	-0.327 (0.202)	-0.003 (0.170)
Edad	-0.349 (0.049)		-0.187 (0.060)	
Vaescu			-0.443 (0.188)	
Observations	1026	695	709	909
Log Likelihood	-707.61	-473.597	-458.599	-626.293
Restr. Log Lik.	-718.24	-475.356	-465.587	-630.765

** Standard Errors in parentheses.*
*** Variables related to eligibility criteria.*

Besides the variables related to eligibility criteria, only age and school attendance are significant (for some subgroups) and negatively related to program participation. Table 5.6 presents the MacFadden R-Squared and the prediction evaluation of the estimated equation (using a success cut-off of 50 percent) versus a constant probability model:

Table 5.6. PSMU Prediction Evaluation

Group	R-squared	Prediction Evaluation
Young Males	0.014	% Correct goes from 50.9 to 62.4
Adult Males	0.004	% Correct goes from 56.5 to 58.7
Young Females	0.015	% Correct goes from 51.8 to 58.1
Adult Females	0.007	% Correct goes from 54.1 to 54.9

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

5.4 A Comparison of the Propensity Score Estimates

Table 5.7 presents simple correlation coefficients among the estimated propensity scores for each of the four subgroups.

Table 5.7. Correlation Coefficients

Young Males	**PSTOT**	**PSUN**	**PSMU**
PSTOT	1	0.6507	0.5937
PSUN		1	0.5669
PSMU			1
Adult Males	**PSTOT**	**PSUN**	**PSMU**
PSTOT	1	0.2229	0.2418
PSUN		1	0.2521
PSMU			1
Young Females	**PSTOT**	**PSUN**	**PSMU**
PSTOT	1	0.3940	0.4213
PSUN		1	0.3363
PSMU			1
Adult Females	**PSTOT**	**PSUN**	**PSMU**
PSTOT	1	0.2471	0.2729
PSUN		1	0.2512
PSMU			1

From Table 5.7, it is apparent that the Propensity Scores do not present, in general, high correlation coefficients among the different data sources used in estimation.

Tables 5.8 to 5.11 present the predicted Propensity Scores for the estimated functional forms obtained from each of the three sources of information to the average information for each of the variables included in the models arising from these sources. The functional form in the Tables is kept fixed along a row and the source of the information for the independent variables is kept fixed along the columns.

Table 5.8. Young Males

Functional Form/Source of Information	Universe	Survey	Combined
Universe Functional Form	0.5389	0.5571	0.5593
Survey Functional Form	0.5635	0.4921	0.5183
Combined Functional Form	0.6500	0.5112	0.5548

Table 5.9. Adult Males

Functional Form/Source of Information	Universe	Survey	Combined
Universe Functional Form	0.4598	0.5279	0.4794
Survey Functional Form	0.3936	0.4326	0.4276
Combined Functional Form	0.4689	0.5184	0.5143

Table 5.10. Young Females

Functional Form/Source of Information	Universe	Survey	Combined
Universe Functional Form	0.3890	0.4447	0.4436
Survey Functional Form	0.4821	0.4793	0.4973
Combined Functional Form	0.4986	0.5112	0.5391

Table 5.11. Adult Females

Functional Form/Source of Information	Universe	Survey	Combined
Universe Functional Form	0.4112	0.4801	0.4761
Survey Functional Form	0.4592	0.4589	0.4582
Combined Functional Form	0.5076	0.5367	0.5311

From these tables a decomposition of the main differences among the average propensity scores into functional form (Aβ) and independent variables (AX) can be attempted. Table 5.12 presents these decompositions for two alternative methods:[13]

Method 1: $PS(\beta_i,X_i) - PS(\beta_j,X_j) = [PS(\beta_i,X_i) - PS(\beta_i,X_j)] + [PS(\beta_i,X_j) - PS(\beta_j,X_j)]$

$= [AX\ ] + [A\beta\ ]$

Method 2: $PS(\beta_i,X_i) - PS(\beta_j,X_j) = [PS(\beta_i,X_i) - PS(\beta_j,X_i)] + [PS(\beta_j,X_i) - PS(\beta_j,X_j)]$

$= [A\beta\ ] + [AX\ ]$

where $PS(\beta_k,X_s)$ denotes the predicted Propensity Score using the functional form obtained by using the k data source (k = universe, survey, universe-survey combined), and s denotes the source of information for the independent variables.

Table 5.12. Propensity Scores Decomposition

	Universe Survey	Versus Info	Universe Combined	Versus Info	Combined Survey	Versus Info
	Method 1	**Method 2**	**Method 1**	**Method 2**	**Method 1**	**Method 2**
Young Males						
AX	-0.0714	-0.0182	-0.0952	-0.0045	0.0436	0.0262
Aβ	0.0246	-0.0650	0.1111	0.0204	0.0191	0.0365
Total	-0.0468	-0.0468	0.0159	0.0159	0.0627	0.0627
Adult Males						
AX	0.039	0.0681	0.0454	0.0196	-0.0041	-0.005
Aβ	-0.0662	-0.0953	0.0091	0.0349	0.0858	0.0867
Total	-0.0272	-0.0272	0.0545	0.0545	0.0817	0.0817
Young Females						
AX	-0.0027	0.0557	0.0405	0.0546	0.0279	0.0279
Aβ	0.0903	0.0346	0.1096	0.0955	0.0319	0.0418
Total	0.0903	0.0903	0.1501	0.1501	0.0598	0.0598
Adult Females						
AX	-0.0003	-0.0212	0.0235	0.0649	-0.006	-0.000
Aβ	0.048	0.069	0.0964	0.055	0.077	0.073
Total	0.048	0.048	0.1199	0.1199	0.0722	0.0722

Although there are mixed results across methods, it appears that the change in functional is more important than the changes in average values of the independent variables as determinants of the differences in propensity scores across data sources.

5.5 Determining a Common Support

The application of propensity scores matching estimators requires that there exist 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, the authors plotted the histograms of the propensity scores for both groups for each of the three estimated propensity scores. Appendix 2 presents these figures for each of the four subgroups.

In the cases of young males, adult males and young females, values of propensity scores for Beneficiaries were not observed for which comparable propensity scores in the sample of Comparisons could not be found. In the case of adult females, there were 20 adult female

[13] There is no way of decomposing the total difference.

beneficiaries in the combined universe-survey data for whom no close matches existed in the comparison sample.

The results reported in the next section assume a common support. It is important to mention that the Programa Joven considered several criteria to select the Comparisons, controlling by variables such as age, gender, labor status, marital status and existence of children, and controlling also by the distribution of these variables. This increases the likelihood that the two populations present similar propensity scores.

6. Impact Estimates: Labor Earnings and Employment

The parameter being estimated is the impact of the program on its recipients. The outcome variables considered were earnings and probability of employment in the twelfth month after the program.

The authors worked with a cross-sectional (CS) matching estimator, given that this methodology compares the results for the Beneficiaries and Comparisons in the same period after the program. The available information makes it possible to apply this methodology. The specific cross-sectional matching estimator was the *Nearest Neighbor Matching Estimator*.[14] 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 included only income information of the specified number of Comparisons with the lowest Euclidean distance to the ith Beneficiary propensity scores.

The technique of bootstrapping was used to obtain the sample variance of the impact estimates. Appendix 2 presents the Matlab (version 5.3.1) codes, which were used in estimation. Table 6.1 presents some descriptive statistics for Beneficiaries and Comparisons in the four subgroups on labor earnings and employment in the twelfth month after the program.

[14] As in the case of any propensity score matching estimators, it is necessary to assume that $E[Y_0 \mid P(X), D=1] = E[Y_0 \mid P(X), D=0]$ and $0 < Pr(D=1 \mid X) < 1$.

Table 6.1. Descriptive Statistics

Young Males	**Mean**	**Std. Dev.**
Beneficiaries Income at 12 months	$138.38	$163.28
Beneficiaries Employment at 12 months	0.677	0.4681
Comparisons Income at 12 months	$127.51	$152.63
Comparisons Employment at 12 months	0.6499	0.4775
Adult Males	**Mean**	**Std. Dev**
Beneficiaries Income at 12 months	$180.61	$192.07
Beneficiaries Employment at 12 months	0.7062	0.4559
Comparisons Income at 12 months	$161.68	$173.84
Comparisons Employment at 12 months	0.7234	0.4477
Young Females	**Mean**	**Std. Dev.**
Beneficiaries Income at 12 months	$86.24	$141.41
Beneficiaries Employment at 12 months	0.4650	0.4993
Comparisons Income at 12 months	$76.47	$130.35
Comparisons Employment at 12 months	0.4333	0.4960
Adult Females	**Mean**	**Std. Dev.**
Beneficiaries Income at 12 months	$111.82	$115.46
Beneficiaries Employment at 12 months	0.57	0.49
Comparisons Income at 12 months	$87.18	$133.44
Comparisons Employment at 12 months	0.45	0.48

6.1 Labor Earnings

The main results for the program impact estimates on earnings are presented in Table 6.2. Impact estimates for 5, 10, 20 and 30 neighbors, for the four subgroups, and for the whole sample are also presented.[15] The authors report program impact estimates using the three estimated Propensity Scores: 1) using the universe individuals and information; 2) using the universe information but the individuals in the "Survey"; and 3) using the individuals and the information from the "Survey."

Table 6.2 shows that program results on earnings are statistically significant for young males and adult females. However, program results on earnings were not statistically significant for adult males and young females. Given that all groups undergo the same type of training programs and that no other aspect of the program differs among the groups, this result is likely

[15] The estimate for the whole sample is constructed by weighting the individual results by sample proportions in the Beneficiary sample of 1,670 individuals.

related to labor market differences for the different groups. For this reason, the training program is more valuable for young males and adult females than for adult males and young females.

Table 6.2. Impact Estimators on Earnings ($ per month)*

1. Universe Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	$19.626 (10.555)	$7.102 (13.029)	$11.098 (11.395)	$31.075 (8.493)	$18.721
10	$19.755 (10.707)	$1.451 (12.574)	$4.955 (11.486)	$28.869 (8.830)	$15.673
20	$18.938 (10.669)	$4.522 (12.275)	$7.341 (11.639)	$29.128 (9.241)	$16.661
30	$22.176 (10.401)	$2.484 (12.278)	$8.020 (11.446)	$26.244 (8.989)	$16.332
2. "Survey" with Universe Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	$23.748 (13.029)	-$0.545 (14.871)	$13.632 (12.546)	$25.494 (9.262)	$16.834
10	$17.508 (11.381)	$9.565 (13.967)	$20.693 (12.747)	$23.402 (8.777)	$18.047
20	$20.061 (11.469)	$7.908 (14.435)	$21.376 (11.690)	$32.396 (8.615)	$21.259
30	$17.771 (11.450)	$13.837 (13.077)	$16.803 (12.757)	$29.661 (8.727)	$20.343
3. "Survey" Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	$22.993 (11.754)	-$9.652 (14.899)	$23.885 (10.430)	$30.696 (8.099)	$17.918
10	$17.167 (11.644)	-$5.2026 (13.670)	$19.653 (9.758)	$31.606 (8.194)	$16.841
20	$19.778 (12.314)	$0.301 (13.660)	$18.813 (9.948)	$29.563 (8.347)	$18.082
30	$21.717 (11.4318)	-$1.137 (13.548)	$17.998 (11.161)	$28.160 (8.379)	$17.717

** Bootstrapping estimated sample standard deviation of the estimators is presented in parentheses.*

Controlling for the source of information used and for the statistically significant groups, program impact estimates on earnings were not very sensitive to the number of nearest

neighbors.[16] In addition, impact estimates for the different propensity score specifications were very similar despite the low correlations among the different scores reported earlier.

6.2 Employment

This section reports the main program impact on the probability of employment. The main results are presented in Table 6.3. This table demonstrates that the estimated program impact on employment was statistically significant for adult females only. For this group, the estimated impact was not sensitive to the number of nearest neighbors. Additionally, the impact estimates for the different propensity score specifications were very similar. For the non-statistically significant groups, there was observed a greater sensitivity of the estimates to the number of nearest neighbors and to the different sources of information used to estimate the propensity scores.

[16] For adult males, results were sensitive to the number of nearest neighbors but not statistically significant.

Table 6.3. Impact Estimators Employment

1. Universe Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	0.0095 (0.034)	-0.0238 (0.039)	-0.0181 (0.049)	0.1318 (0.035)	0.0345
10	0.0034 (0.035)	-0.0257 (0.029)	-0.0410 (0.048)	0.1194 (0.034)	0.0274
20	0.0072 (0.033)	-0.0292 (0.027)	-0.0244 (0.045)	0.1202 (0.035)	0.0279
30	0.0087 (0.033)	-0.0238 (0.027)	-0.0171 (0.047)	0.1130 (0.035)	0.0287
2. "Survey" with Universe Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	0.0487 (0.032)	-0.0280 (0.031)	0.0019 (0.053)	0.0987 (0.037)	0.0379
10	0.0303 (0.036)	-0.0198 (0.030)	0.0255 (0.050)	0.1035 (0.033)	0.0403
20	0.0226 (0.036)	-0.0181 (0.031)	0.0107 (0.049)	0.1341 (0.035)	0.0453
30	0.0198 (0.035)	-0.0122 (0.029)	0.0213 (0.047)	0.1222 (0.034)	0.0441
3. "Survey" Information					
Neighbors	**Young Male**	**Adult Male**	**Young Female**	**Adult Female**	**All**
5	0.0323 (0.048)	-0.0295 (0.028)	0.0023 (0.035)	0.1282 (0.039)	0.0421
10	0.0276 (0.093)	-0.0357 (0.027)	-0.0095 (0.048)	0.1346 (0.035)	0.0392
20	0.0251 (0.039)	-0.0293 (0.029)	-0.0153 (0.049)	0.1277 (0.032)	0.0368
30	0.0209 (0.037)	-0.0395 (0.029)	-0.0137 (0.052)	0.1252 (0.046)	0.0328

** Bootstrapping estimated sample standard deviation of the estimators is presented in parentheses.*

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, together with data on the costs of the program, is used to conduct a cost-benefit analysis of the program and to calculate its rate of return to dollars spent. Information is available about:[17]

[17] This information was provided by Programa Joven.

- *Direct Cost of Training*: Costs include training services provided by ICAP, insurance for short-term stays in firms, and fellowships and subsidies to program beneficiaries with children.
- *Indirect Costs*: These include personal, infrastructure, inputs and operational expenses of the ministerial department that carried out the program. It also includes, among others, information on bidding costs, promotion, computer services and supervision. The unit that carried out the program also had other projects, although the Programa Joven was the most important in terms of expenditure.[18] This means that in order to have a reliable estimate of Programa Joven's administrative costs, it was necessary to distribute these costs among the other projects.

The accumulated total cost from the second semester of 1993 to December 1998 breaks down as follows:

Table 7.1. Cumulative Budget as of 12/31/1998

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%**

The information needed to separate and allocate the department's administrative costs among its other programs was not available. As a compromise, it was assumed that the administrative, concurrent and financial costs maintain a constant proportionality with the direct costs. Thus, direct costs were assumed to represent 3.055 times indirect costs (3.055=Direct Costs/Administration + Concurrent + Financial).

Programa Joven estimated the Direct Cost of the courses in the fifth round of the training program. The direct cost for every student who graduated from the Technical Knowledge phase was estimated at US$1,342. Given the assumption of constant proportionality between direct and indirect cost, it is possible to estimate an indirect cost of US$483.83 per graduate of the

[18] Other components include Proyecto Microempresas, Proyecto Imagen and Fortalecimiento Institucional.

Technical Knowledge phase. Adding the direct and indirect costs, the authors estimate a total cost of US$1,825.83 per participant.

In addition, the authors assumed zero opportunity cost while the recipient is undertaking training and a constant impact of training on labor income. The authors conducted the cost-benefit analysis under different scenarios for the duration of benefits, discount rate and ratio of direct to indirect costs. The specific formula used to obtain the Net Present Value (NPV) is as follows:

$$NPV = \sum_{t=0}^{T} \left(\frac{E_t - C_t}{(1+r)^t} \right)$$

where E_t denotes the mean earnings effect on the recipients,[19] C_t denotes the costs of Programa Joven (assuming they take place at time zero), T denotes the duration of benefits and r denotes the opportunity cost of capital.

The different values assumed for these variables are presented in Table 7.2. The yearly real discount rate was set at two values, 5 percent and 10 percent, with the latter corresponding to the social rate of discount. The authors considered two values for the ratio of indirect to direct costs: the figure of 0.327 mentioned above and 0.15, a figure consistent with indirect costs representing 13 percent of total costs.[20]

Table 7.2. Simulation Scenarios

Variable	Values
Duration of Benefits (years)	1, 3, 6, 9, 12, 15 and infinite
Discount Rate (%)	5 and 10
Ratio of Indirect/Direct Costs	0.15 and 0.327

The main cost-benefits results obtained are presented in Table 7.3. Table 7.3 presents the NPV calculations for two benefit figures: (i) US$17.87 per month, which corresponds to the average impact estimator on earnings for all the groups and across the three sources of information (see Table 6.2); and (ii) US$24.67 per month, which corresponds to the average

[19] Opportunity cost while the individual is undertaking training is assumed to be zero.
[20] Programa Joven officials suggested this figure informally.

impact estimator on earnings for young males and adult females only. As Table 6.2 shows, there are statistically significant program impacts on earnings among young males and adult females. This is equivalent to performing a cost-benefit analysis for a program (with similar costs) targeted to young males and adult females only.

Table 7.3. Cost-Benefit Analysis (in US$)

Duration of Benefits	Discount Rate	Ratio Indirect/ Direct Cost	Net Benefits (US$)	Net Present Value (US$)	Net Benefits (US$)	Net Present Value (US$)
1 year	5%	0.15	17.87	-1,339.1	24.67	-1,261.3
1 year	5%	0.327	17.87	-1,576.6	24.67	-1,498.9
1 year	10%	0.15	17.87	-1,348.4	24.67	-1,274.2
1 year	10%	0.327	17.87	-1,585.9	24.67	-1,511.7
3 years	5%	0.15	17.87	-959.4	24.67	-737.1
3 years	5%	0.327	17.87	-1,196.9	24.67	-974.6
3 years	10%	0.15	17.87	-1,010.1	24.67	-807.0
3 years	10%	0.327	17.87	-1,247.6	24.67	-1,044.6
6 years	5%	0.15	17.87	-454.9	24.67	-40.6
6 years	5%	0.327	17.87	-692.5	24.67	-278.1
6 years	10%	0.15	17.87	-609.4	24.67	-253.9
6 years	10%	0.327	17.87	-846.9	24.67	-491.4
9 years	5%	0.15	17.87	-19.2	24.67	561.0
9 years	5%	0.327	17.87	-256.7	24.67	323.5
9 years	10%	0.15	17.87	-308.4	24.67	161.7
9 years	10%	0.327	17.87	-545.9	24.67	-75.82
12 years	5%	0.15	17.87	357.2	24.67	1,080.8
12 years	5%	0.327	17.87	119.7	24.67	843.2
12 years	10%	0.15	17.87	-82.3	24.67	473.9
12 years	10%	0.327	17.87	-319.8	24.67	236.4
15 years	5%	0.15	17.87	682.4	24.67	1,529.7
15 years	5%	0.327	17.87	444.86	24.67	1,292.2
15 years	10%	0.15	17.87	87.7	24.67	708.6
15 years	10%	0.327	17.87	-149.9	24.67	471.0
Infinite	5%	0.15	17.87	2,744.3	24.67	4,377.9
Infinite	5%	0.327	17.87	2,507.7	24.67	4,140.4
Infinite	10%	0.15	17.87	600.9	24.67	1,417.3
Infinite	10%	0.327	17.87	363.5	24.67	1,179.8

Therefore, the NPV can be positive or negative. Ceteris paribus, the longer the time period for the benefits, the smaller the discount rate and the lower the ratio of indirect to direct costs, the greater the NPV of the Programa Joven.[21]

8. Conclusion

This paper aimed to answer the following questions: (1) Does Programa Joven increase the income of the trainees? (2) Does Programa Joven increase the probability of employment? (3) How sensitive are program impact estimates to different propensity score specifications? and (4) What is the rate of return to dollars spent on Programa Joven?

First, the results indicate that the program's impact on earnings was statistically significant only for young males and adult females. In the opinion of the authors, this result, which makes the training program more valuable for these specific groups, is related more to the different labor market conditions these subgroups face than to program-specific components. Program impact estimates on earnings were not sensitive to the number of nearest neighbors.

Second, the estimated program impact on employment was statistically significant only for adult females. For this group, the estimated impact was not sensitive to the number of nearest neighbors. A greater sensitivity of the estimates to the number of nearest neighbors was observed for the other groups.

Third, impact estimates on earnings and employment for the different propensity score specifications and for the statistically significant groups were very similar despite the low correlations among the different scores reported in the article. For non-statistically significant groups, the authors observed a greater sensitivity of the estimates to the different sources of information used to estimate the propensity scores. This was a surprising result; greater variability in the impact results across different propensity score specifications was expected.

Finally, the cost-benefit exercise suggested that, ceteris paribus, the longer the time period for benefits, the smaller the discount rate and the lower the ratio of indirect to direct costs, the greater the NPV of the Programa Joven. Young males and adult females, which present higher and statistically significant earning impacts, required only 9 years of program benefits to achieve a positive NPV. After 12 years, all beneficiaries had reached a positive NPV.

[21] The NPV assumes a zero deadweight loss of the resources used to fund Programa Joven. If a 50 percent deadweight loss is assumed, for instance, $890 must be subtracted from the figures reported in the table.

References

Hardle, W. 1992. *Applied Nonparametric Regression*. Cambridge, United Kingdom: Cambridge University Press.

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

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

---. 1998. "Matching As an Econometric Evaluation Estimator." *Review of Economic Studies* 65 (2): 261-294.

Heckman, J., R. LaLonde, and J. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In: O. Ashenfelter and D. Card, editors. *Handbook of Labor Economics, Volume IIIA*. Amsterdam, the Netherlands: North-Holland.

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

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

Jones, M.C., J.S. Marron, and S.J. Sheather. 1996. "A Brief Survey of Bandwidth Selection for Density Estimation." *Journal of the American Statistical Association* 91: 401-407.

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

Manski, C. F. 1996. "Learning about Treatment Effects from Experiments with Random Assignment of Treatments." *Journal of Human Resources* 31(4): 709-733.

Manski, C. F., and S.R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples." *Econometrica* 45(8): 1977-1988.

Ministerio de Trabajo y Seguridad Social de la República Argentina, Secretaría de Empleo y Formación Profesional, and Gerencia de Evaluación de Impacto y Estudios Especiales. "Base de datos e información del Programa Joven." Buenos Aires, Argentina: Ministerio de Trabajo y Seguridad Social.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London, United Kingdom: Chapman and Hall.

Todd, P. 1999. "A Practical Guide to Implementing Matching Estimators." Mimeographed document presented at Inter-American Development Bank meeting in Santiago, Chile.

Appendix 1: Description of Variables

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 Incomplete, 1=Yes, 0=No
PRICOM	Primary Education Completed, 1=Yes, 0=No
SENOCOM	Secondary Education Incomplete, 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 Cordoba, 1=Yes, 0=No
MENDOZA	Reside in Mendoza, 1=Yes, 0=No
STAFE	Reside in Santa Fe, 1=Yes, 0=No
TUCUMAN	Reside in Tucuman, 1=Yes, 0=No
MUESTRA	Internal Control Variable
GRUPO	Internal Control Variable

Appendix 2: Common Support

Figure A.2.1. Young Males Universe Information

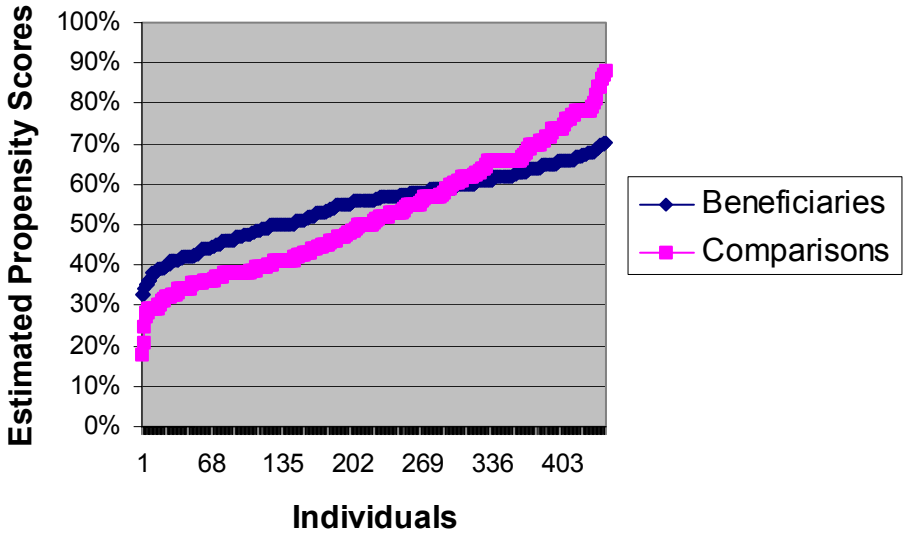


Figure A.2.2. Young Males Combined Information

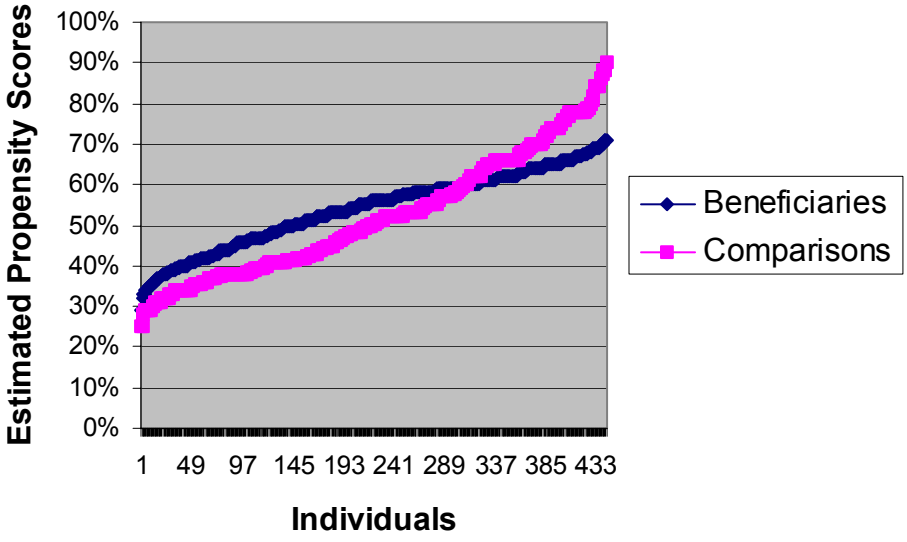


Figure A.2.3. Young Males Survey Information

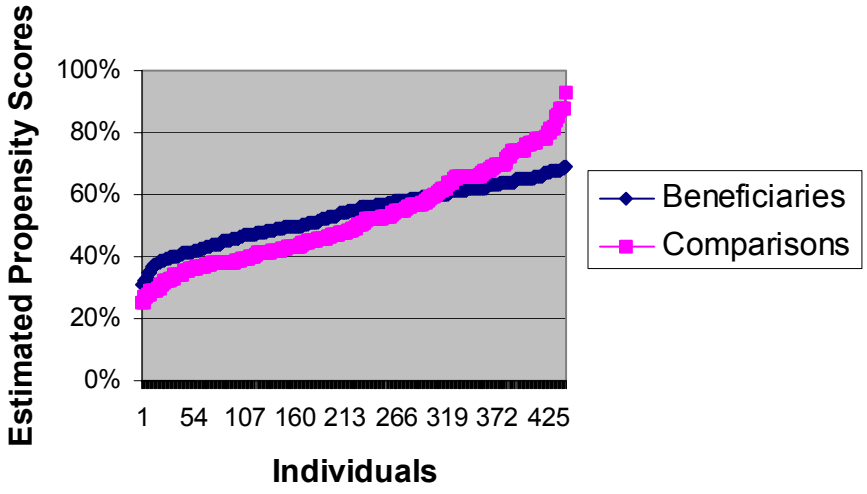


Figure A.2.4. Adult Males Universe Information

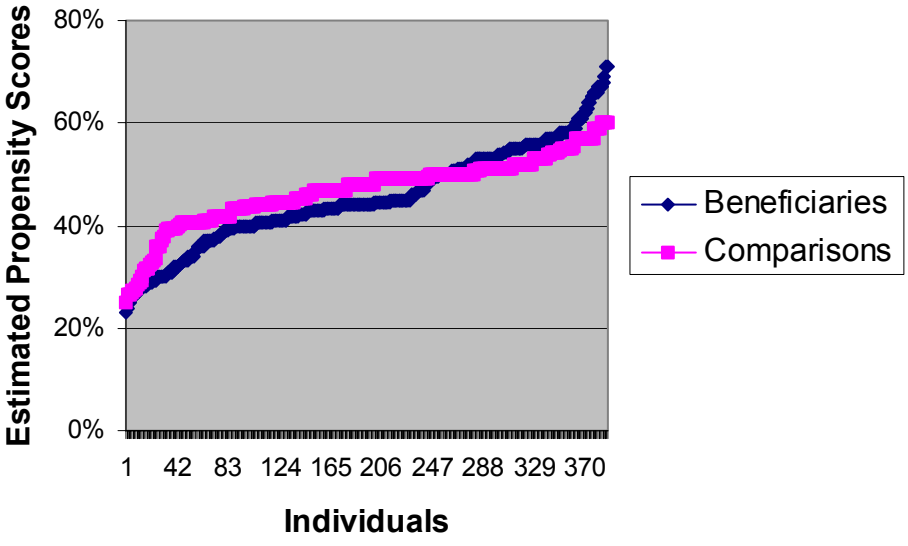


Figure A.2.5. Adult Males Combined Information

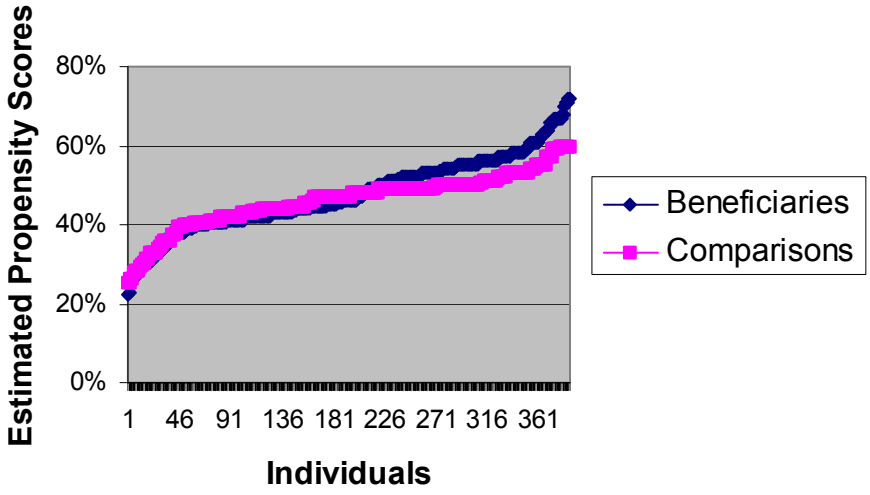


Figure A.2.6. Adult Males Survey Information

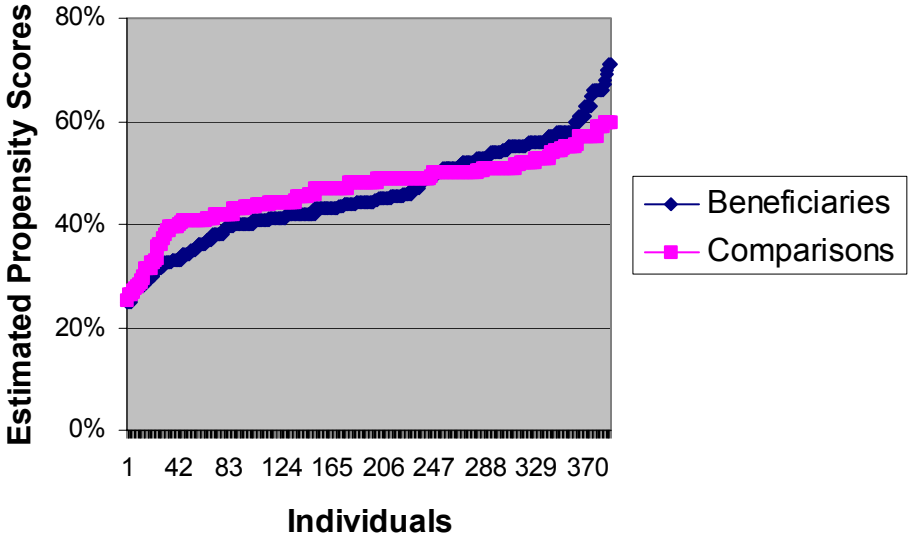


Figure A.2.7. Young Females Universe Information

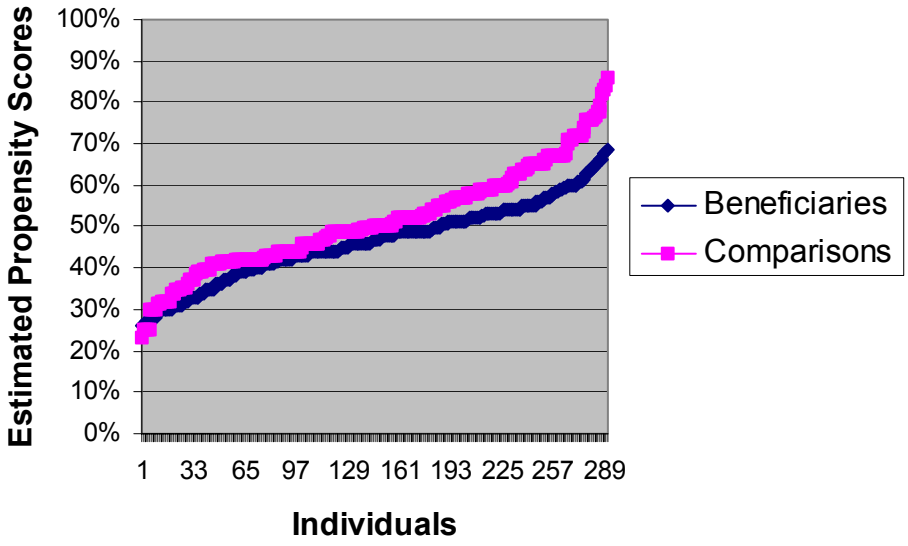


Figure A.2.8. Young Females Combined Information

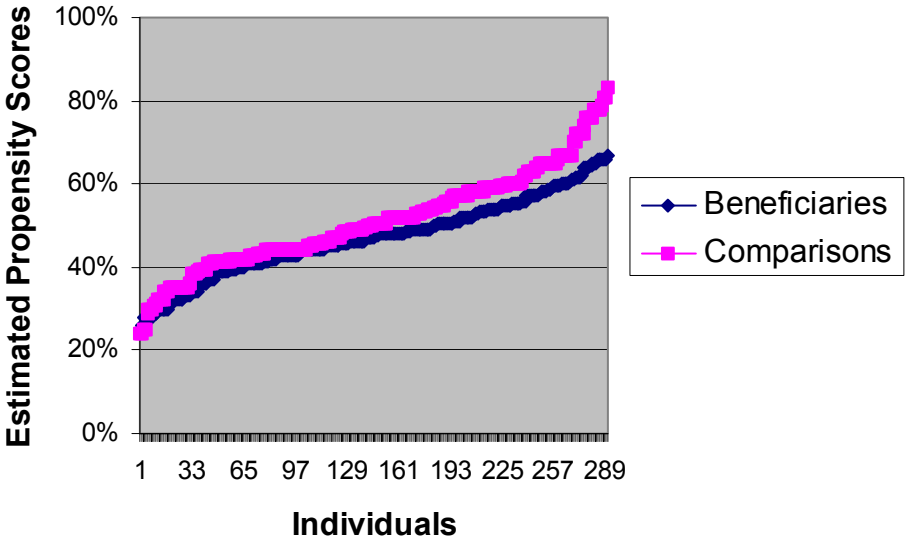


Figure A.2.9. Young Females Survey Information

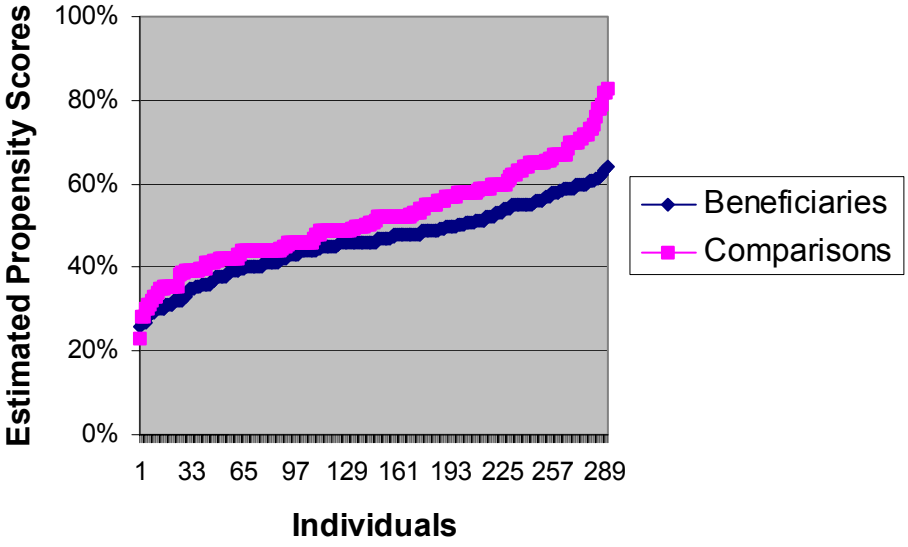


Figure A.2.10. Adult Females Universe Information

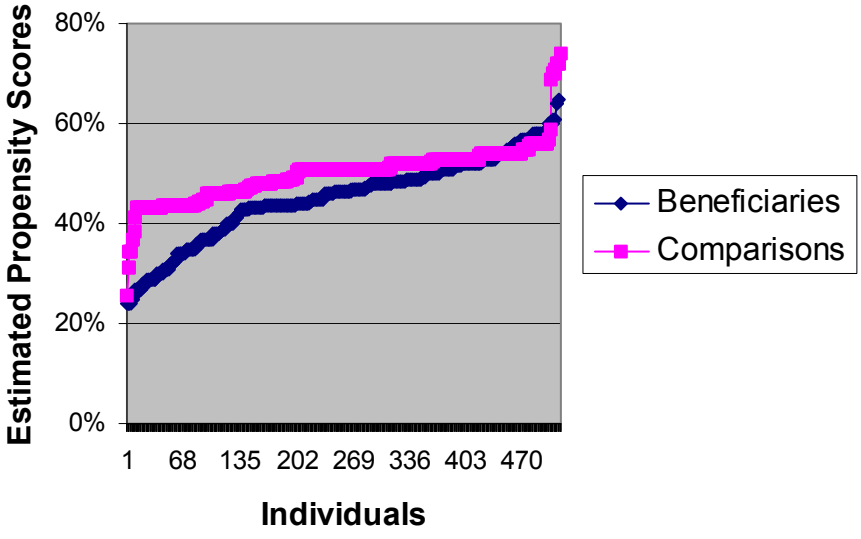


Figure A.2.11. Adult Females Combined Information

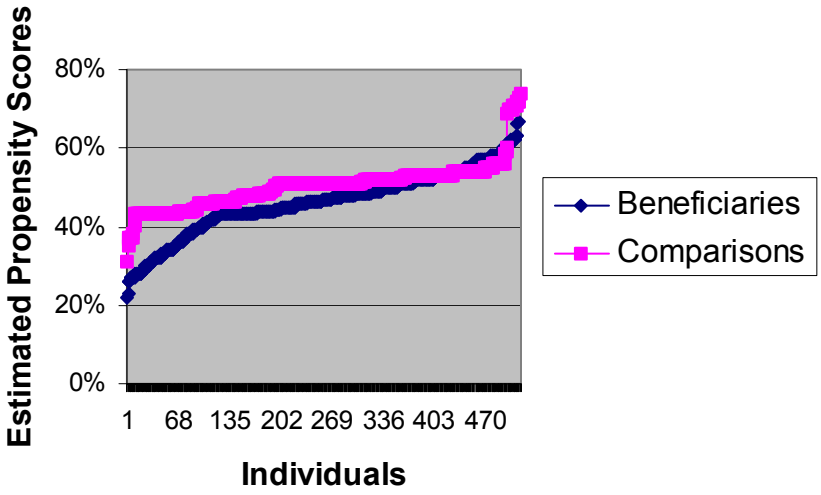
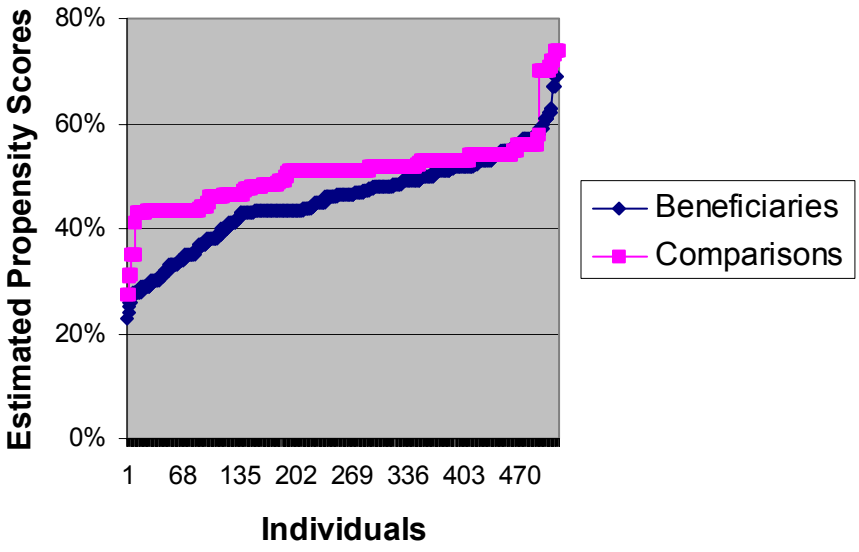


Figure A.2.12. Adult Females Survey Information



Appendix 3: MATLAB Codes

a) 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 comparisons in the sample. Data set is ordered: first the
beneficiaries and then the comparisons
%

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

%
%Transferring 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 comparisons which are going to
be used for each of the 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
```

b) 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 comparisons (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 comparisons in the sample. Data set is ordered: first the
beneficiaries and then the comparisons
%

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

%
%Transferring 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
%

p=2147483647.0;
q=2147483655.0;
r=16807.0;

%
%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;

%
%Define the number of neighbors
%

neighbor=30;

%
%The following loop defines the comparisons that are going to be
used for each of the beneficiaries. Then it calculates the
average earnings for the 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
%

 meaneffe=mean(imp);
 boot=sqrt(var(imp));
 meaneffe, boot
```