University of Lausanne: Applied econometrics II

Prof. Marius Brülhard

Ass. Frederica Sbergami

Treatment effects in adult education

Evaluation of a school training low skilled labour force

In a contra factual framework, I compute average treatment effects under several assumptions in order to evaluate the performance of the training program at the Centre de Formation à l'Intendance CFI. I show that training is can increase wage in the mean by 24% and that the self-selection process plays an important role for low skilled workers.

Catherine Pfeifer, MSE student

April 2005

catherine.pfeifer@unil.ch

Contents

Tables

1 Introduction

In knowledge society, unskilled labour force earns low wages compared to skilled labour force. This increasing difference is potentially a source of social tension. In his book "The post-capitalist society" Peter Drucker states that life long training is the only way to give equal chances to low-skilled labour force and preserve social peace.

Currently the whole Swiss educational system is revised on the academic as well as on the practical level and life long training should become an integral part of it. The programs for life-long training program for unskilled labour are underdeveloped, as these people usually cannot afford to pay for education. Is it nevertheless efficient to develop training programs for unskilled labour? Do these persons gain from training program?

To answer this question, labour economists have developed a method based on the contra factual framework that allows not only evaluation of programs thanks to wages but also analyses selection processes into programs.

The objectives of this work are first to try to evaluate training programs for low-skilled labour force with methods calculating treatment effects, based on the contra factual framework. The school used for the evaluation is the CFI (Centre de formation à l'intendance), which trains low skilled labour force in cleaning and home economy sector. Therefore the second objective is to better understand the motivation of unskilled labour to get trained. In the end, recommendations based on the quantitative result will be made and should allow a new strategic orientation for the CFI.

2 Method using treatment effects

2.1 General framework[1]

Treatment effects are a special case of average partial effect, an effect measured for a binary explanatory variable. Originally treatment effects have been used in test on medicine before their introduction on the market. The binary variable represented the medical treatment, which takes the value 1 if the person is treated and 0 if the person got a placebo. The main difference with training programs is, that for test on medicine people are randomly selected into the treated or the non-treated (placebo) group. But the decision to join in a training program is not random. Indeed, people

[1] this section is mainly based on WOODRIDGE J.- econometric analysis of cross section and panel data –MIT2002 p.603-607 which itself is based on Heckmann(1997) and Imbens and Angrist (1994)

decide themselves whether or not they get enrolled. Therefore new way to estimated treatment effect have been developed trying to face the limits by introducing the contra factual framework and self-selection (Rosenbaum & Rubin 1983, Heckmann 1992,1997).
The modern literature on treatment effects begins with a contra factual, where each individual has an outcome with and without treatment. But because an individual cannot be observed in both states at a given time, we cannot observe the value of the explanatory variable in both states. So in fact we face a problem of missing data.

2.1.1 Mathematical computations of the treatment effects with one treatment

Let us denote y_0 : the outcome without treatment, y_1 : the outcome with treatment and w : the treatment binary variable (1= treated). Furthermore, assume that we have an independent, identically distributed sample of the population. This rules out that a treatment effect on person i affects a person j[2]. In the framework of this research, this is not very restrictive since only few people get the training compared to all people working in the cleaning and home economy sector. Then the treatment effect is given by y_1 - y_0. Since (y_1;y_0; w) is a vector of random variables, then y_1 - y_0 is random too. The first way to estimate this value is to compute the average treatment effect (ATE) as it has been done by Rosenbaum & Rubin (1983):
$ATE = E(y_1 - y_0)$, but this can also include part of the population that that is not relevant for a policy purpose, in other words, that people who would never be accepted on the training program can be included. Notice that this is not the case in the analysis of the CFI, since only people who could join the program have been included in the survey.
Another interesting measure of is the average treatment effect on treated (individuals, ATE1) that can be measured as follows: $ATE1 = E(y_1 - y_0 | w = 1)$, which measures the effect only on those people that have been treated. The main problem with the contra factual framework is that we cannot observe an individual in the two states at a given time. So the observed outcome is given by :
$y = (1 - w)y_0 + wy_1 = y_0 + w(y_1 - y_0)$. In order to be able to measure treatment effects, we need to have more information about the unobserved value, which must be consistently estimated.

2.1.2 Overview of the estimation methods

There are three main challenges that estimation methods based on a contra factual framework must be faced when they want to be adapted to evaluation of training programs. The first is the challenge to face self-selection into the program, which can be integrated by using a two-step procedure. In

[2] in the literatures this assumption is call stable unit treatment value assumption (SUTVA)

the first step the probability to take part in the program is estimated and then its result is introduced into the initial equation that has to be estimated. The second challenge is, that the treatment can influence the explicative variable, and so the effect become group specific and are therefore called heterogeneous. The third challenge occurs when the treatment variable is endogenously defined. In this case, the instrumental variable must be used to correct the endogeneity. This last case is described in the appendix 8.3.5 and 8.3.6 only and will only partly be empirically tested.
The various methods to face these problems can be summarized as follows:

Table 1 : overview of estimation methods[3]

	Random selection	Self-selection
Exogenous treatment	Case 1 :	Case 2 :
Homogeneous treatment if ignorability of treatment assumption $E(y_0 \vert x, w) = E(y_0 \vert x) = E(y \vert x)$	Ordinary least square (OLS)	Two stage estimation based on propensity score
Heterogeneous treatment	Computation based on OLS by group	Two stage estimation with selection equation by group
Endogenous treatment	Case 3	Case 4
Homogenous effects	OLS with introduction of pre treatment characteristics	Three stage estimation with selection equation (corrections on residuals) or instrumental variable
Heterogeneous effects		Three stage estimation with selection equation (corrections on residuals) or instrumental variable by group

[3] The appendix 8.3 shows how average treatment effects can be computed mathematically in each case.

2.2 *The treatments at the CFI*

The CFI (centre de formation à l'Intendance) is a school, which has been created 15 years ago and offers 1.5 years on the job training programs for low skilled labour in cleaning and housekeeping sector. The selection of the candidate into the program is based on the candidate's motivations, on the need of the institution she is working for and the previous experience. The program leads to the "certificat CFI" which is given to every candidate that passed the practical exam organised by the CFI and wrote a thesis. Furthermore, the candidate can present at the governmental exam, which leads to the "brevet fédéral" a Swiss wide recognized title, which corresponds to the next higher level after the CFC (the title one gets after the apprenticeship) in the practical career as opposed to academic.

Conditions present the exams are to possess a CFC and 3 year practical experience in the sector, or 8 years experiences if the candidate does not have any CFC. This gives a chance to people, who lack in the basic education to enter in the officially recognised Swiss educational system.

In terms of treatment, we will say that a candidate receives a *single treatment* if she works in a job only with the certificate based on the CFI training and we will say that he gets a *double treatment* if he get the brevet federal based on the exam and the certificate of the CFI training.

3 The data

In order to get the data, the CFI made a survey which had been sent to 3 categories of people : those who have the brevet and the certificate, those who have the CFI certificate and those who do not have any of them but would fulfil the conditions to join the training programs.

This survey gives us the information about the wage, the education, the civil status, nationality, the position at work, the number of children, satisfaction at work and volunteering activities.

Furthermore, series of open questions on the motivation for trainings and the effects of training have been asked to candidates at CFI. Reasons for not joining the training program have been asked to those who never followed CFI training. This survey should allow a second more qualitative evaluation, which can be compared to this quantitative study. Note that all people who got the brevet federal, got the certificate at the CFI, which allows us to speak of the double treatment as defined in the previous section.

The survey can be found in the appendix 8.1.

4 The empiric results

4.1 Basic model: Mincer model

In economics the gain of training is only measured by the increase in wage. Even though education should be evaluated on more than just its influence on the wage as for example self-esteem, it has the big advantage to not be influenceable by the school itself and reveals the real value of the education on the market. To do so a wage equation is needed. In labour economics, often the Mincer model is used (Mincer 1958). It says that the change in wage is influenced by a non-linear relationship of the time of education, of experience and age. Furthermore I added control variables for gender, nationality and number of children. I use the following specification:

$$lwage=const+edu+year\_exp+year\_exp\_2+age+age\_2+age\_3+gend+swiss+child\_wom+\varepsilon$$

where edu is the level in education, year-exp the years of experience in the cleaning sector and home economy sector, gend is the gender (1=woman), swiss is the nationality (1=swiss) and child_wom is the number of children if the individual is a woman and 0 if it is a men. This implies the assumption that the number of children is only significant for women, they stop working while men don't!

Since we do not time a person has been educated but only the highest level of education she has achieved, I created a count variable that grows when a higher education is achieved. For example if an individual has achieved obligatory school than edu = 1 and if he has a CFC then edu =2 etc… The time of experience within the sector is known and therefore the non-linear relationship is entered with the squared value (which is varname _2 for the square and _3 if ^3). The same is true for age. A complete description and construction of the variable can be found in the appendix 8.2.

The table 2, column 1 shows the estimation of this model with OLS. As we can see almost all the coefficients have the expected sign. Indeed education, year of experience will increase the wage while the number of children a woman decreases it. The non linear relationship with age have the right sign since wage should be lower for young people and for older people, while for the middle-aged people it should be maximal, but it is not significant. Nationality is not significant which leads to the positive conclusion that there is no discrimination based on nationality. But as the coefficient of the variable gender shows there is a quite high discrimination of women[4]. Surprising is also the

[4] This is the result of this estimation method, in fact if the topic of this research was discrimination of women on labour market, other more sophisticated methods as Oaxaca-Blinder should have been used in order to measure the discrimination in a better and more precise way.

non significance of number of children for a woman, which would imply that a woman does not get penalised the loss of flexibility due to the duty of education. One explanation for that is that I should have differentiated between small children and children who go to school. Unfortunately this information was lacking in the dataset. The column 2.2 shows the data adapted model in which the non-significant variables have been omitted. Note the Cook-Weisberg test on heteroskedasticity cannot reject the hypothesis that the variance is constant, and that the basic hypothesis of OLS are not violated and no correction is needed.

Table 2 : estimation based on OLS

	2.1	2.2	2.3
	Lwage	lwage	lwage
edu	0.110 (5.22)**	0.112 (5.78)**	0.031 (0.90)
year_exp_int	0.015 (2.16)*	0.005 (2.30)*	0.004 (2.09)*
year_exp_int2	0.000 (1.63)		
age	-0.035 (1.20)	0.005 (2.54)*	0.004 (1.85)
age2	0.001 (1.57)		
age3	0.000 (1.69)		
gend	-0.169 (2.74)**	-0.191 (3.53)**	-0.203 (4.15)**
swiss	0.037 (0.80)		
child_wom	-0.031 (1.49)		
brevet			0.224 (2.63)*
cert			0.098 (1.73)
Constant	8.312 (25.52)**	8.081 (73.17)**	8.247 (66.31)**
Observations	88	88	88
R-squared	0.51	0.45	0.49

4.2 *Effects assuming random selection*

Under the very strong assumption of ignorability of treatment and random selection we are in the simplest case where we can use OLS. First the homogenous effect, i.e. the measurement of the effect over the whole sample is discussed.

4.2.1 Homogenous treatment and ignorability of treatments

In this very specific case, the average treatment effect is simply an average partial effect, which is measured by the dummies. Table 2, column 2.3 show the result computed under these assumptions. It presents the model, keeping a constant and dropping the group which never attended to courses at the CFI and this implies that this group becomes the reference group. The dummies brevet and certificat therefore measure the average treatment effect. Only the dummy brevet is significant, which leads to the conclusion that only if a person gets the double treatment, the courses at CFI and passes the brevet federal exam, then the wage will be higher. But for people who only got the CFI courses leading to the certificate, it does not influence the wage. Note also that while including these dummies, education is not significant anymore, because the dummies measure a great part of the variable education. Indeed edu is the same for all people who go a brevet and the same for all those who got the certificat. This shows clearly that CFI offers a treatment that is only useful to those people who get the double treatment. This could be explained by the fact that the brevet is a title that is officially recognized.

4.2.2 Heterogeneous treatments

With the ignorability of treatment assumption, we assumed that the effect of a treatment is the same over the whole sample, so we could use a simple OLS over the whole sample. If this assumption is relaxed, then the effect can vary over the group. This is interesting information, because it allows comparing the benefit from the treatment, compared to the potential benefit of a person who did not get the treatment. The problem is, that the potential wage of people who didn't get a treatment and the wage before treatment of the treated people, is unknown, so we cannot measure them. For this reason, we need to estimate the wage in the unobserved state for each individual consistently. It turns out that this can be done, by estimating the wage equation with OLS for each group separately and predicts the wage. So we get a predicted value for each state and individual. The average treatment effect and the average treatment effect on treated can be computed by taking the average of the difference between the predicted values. A complete proof of this can be found in the appendix 8.3.2.

As the sample for each group is very small (25 observations), the OLS estimations on each group separately does not lead to consistent estimations. Furthermore, in order to be able define whether the effects differ in between the groups, the standard deviation for the treatment effect must be computed with the bootstrap method. This is not possible with such a small sample. Therefore these results are not displayed here.

4.3 *Effects with self-selection*

4.3.1 Homogenous effects based on propensity score (selection on observable elements)

Until now we have assumed that individuals are randomly selected into the training program. This is not a very realistic assumption since the students are chosen by a committee in order to be accepted on the CFI training program. Furthermore, the people who want to join a training program are those, who expect to get some benefit from the education they get at CFI ; while those who do not expect any benefit do not even apply. In order to introduce these phenomena, the selection process has to be included in the estimation. Intuitively, a selection equation defines the probability of joining one or both treatments. As we have a double treatment a new variable needed to be created : cfio, which take value 0 for the non-treated individuals, 1 for the single treated individuals and 2 for the double treated. Then the probability to be in a given group can be computed thanks to an ordered probit estimation on cfio and the other variables of the wage equation..

Table 3 : results based on the propensity score

Table 3	lwage
Cert	0.0848296 (1.75)
Brevet	0.2797816 (5.47)**
Pr(cfio==1)	-0.2796389 (-0.51)
Pr(cfio==2)	0.2685097 (1.89)
Constant	8.389394 (54.78)**
Observations	88
R-squared	0.38
Absolute value of t statistics in parentheses	
* significant at 5%; ** significant at 1%	

The model with a constant, the treatment variable and the propensity score allows to compute the average treatment effects and test whether the self-selection into the program is significant. As shown in table 3,none of the probabilities is significant. This suggests that the selection into both groups is random and leads to the conclusion that the estimations with OLS are not biased. Therefore, the conclusion about the average treatment effects measured by the dummies (cert, brevet) stay the same : only the coefficient of brevet is significant implying that only double treatments has a positive impact on wage.

But these results have to be taken with caution because the selection equation, which results are presented in the appendix, has few variable are significant and the F-test does not reject the hypothesis that all coefficients are equal to zero.

5 Endogeneity of treatment

Up to now, we have assumed that if the individuals are or randomly selected or self-selected based on a known and observed process. When this is not the case, the selection into a program is not observed and the treatment is said to be endogenously defined. In econometrical terms, this implies that the residuals of the models are not independent of the treatment. As in this case the selection is done on unobservable elements we have an omitted variable bias. Therefore the model must be adapted.

Basically, two methods have been developed to solve this problem. The first one is the method with instrumental variable that tries thanks to an instrument to correct for the bias. The second method, defines a selection equation that is endogenous to the wage equation.

5.1 *Homogenous effect based on the propensity score and instrumental variables*

As usual in econometrics, problems of endogeneity are faced with the introduction of an instrumental variable. Remind, that a good instrument is highly correlated with the variable, which is suspected to be endogenous, and not correlated with the right hand variable. In the case of the CFI it is not clear if the treatments are exogenous. Therefore, in the survey, people were asked about the volunteering activities, which can be supposed not be strongly correlated with the wage but could show an attitude to life of openness, which could lead to a higher motivation for training. The endogeneity of treatment has therefore been tested with a Hausman test on exogeneity with volunteering as an instrument. But this test could not reject the hypothesis of exogeneity. So volunteering is not needed as an instrument. As in the dataset there is no other potential instrument, this method does not allow for endogenous selection in the framework of this work. Therefore I focus on the second method.

5.2 *Homogenous effects with endogenous self-selection equation*

The second method to correct for the omitted variable bias is to define an endogenous selection equation, and therefore here again a two-step method is used. But instead of using the propensity score, a new parameter, called lambda[5] is included. Lambda, also known as the inverse mills ratio is computed from the fitted the values of the selection equation and allows to compute a parameter that correct for self-selection. The intuitive idea behind this method is to include the probability of

[5] In fact lambda is the inverse mills ratio

joining a training program that is not computed over the whole sample, but over the one that belong to the group referring to[6]. For the computation of lambda, please consult the appendix 8.3.6.2.

Various endogenous selection processes have been test. The following equation is the one that tracks the process best and is give by :

$$\text{treatment} = \text{cst} + \text{swiss} + \text{statut} + \text{age} + \text{satis} + \text{act\_rate} + \varepsilon$$

where in addition to already discussed variable we have the self evaluated satisfaction index, the act_rate which measures the activity rate in %, and statut which is an integer which measures the level in the hierarchy of the company. The variable treatment is an integer which takes the value 0 if no treatment the value 1 and if trained by the CFI only and 2 if trained at CFI and achieved the brevet fédéral. Out of this equation I could compute the predicted value and compute the tree correction factor for the residuals of the wage equation. This is quite complicated, since the correction factor is not the inverse mills ratio anymore as there are now more that just a binary state (training - no training).[7]Since I was unable to program this on stata I decided to present the results for each treatment separately. This is much easier because stata has a built-in routine[8] to compute selection equation in the framework of treatment effects. When considering each treatment separately the selection equation become :

$$certificat = \text{cst} + \text{swiss} + \text{statut} + \text{age} + \text{satis} + \text{act\_rate} + \varepsilon$$

$$brevet = \text{cst} + \text{swiss} + \text{statut} + \text{age} + \text{satis} + \text{act\_rate} + \varepsilon$$

this can be estimated with probit. With the predicted values of these models we can compute the lambda, and include them into the wage equation in order to correct for self-selection.

The table 4 shows the estimation the wage equation corrected for self-selection. Column 4.1 shows the result considering only the single treatment compared to the non treated group[9]. The selection equation shows that the probability to join the training program only is influenced by nationality. Swiss have a much higher probability to join the training program. There are two possible reasons for this : first it might be that Swiss find it easier to get into a training program maybe because of language, secondly it might be that there is a discrimination based on nationality when the responsible choose the candidates. Status has the expected sign, indeed the higher in the hierarchy the more probable it is to get into the training program. Age, satisfaction degree, and activity rate are not significant.

[6] in econometrical terms the correction term lambda is based on a truncated normal distribution, please consult the appendix for the exact computation.
[7] Since there are 3 probabilities (no training, cfi, brevet), one cannot exploit the symmetry of the normal distribution anymore (as it is done for computing the lambda).
[8] Command :treatreg
[9] note that for this estimation the double treated group has been drop from the sample

Table 4 : estimations for endogenous treatment

Table 4	4.1	4.2
	Treatment=CFI training	Treatment = brevet & CFI training
Wage equation	lwage if brevet<1	
edu	0.0280698 (0.66)	0.06564 (2.39)*
year_exp_int	0.0078602 (2.93)**	0.00466 (2.10)*
age	0.0073996 (2.91)**	0.00291 (1.36)
gend	-0.2114526 (-3.48)**	-0.2034 (-4.09)**
cert	0.2588514 (2.82)**	
brevet		0.2410 (2.88)**
_cons	8.014821 (49.76)	8.241 (65.69)**
Selection equation		
	Cert	brevet
swiss	1.253788 (2.40)*	-0.5712 (-1.29)
statut	0.9956557 (3.69)**	1.379 (3.97)**
age	-0.0445515 (-1.91)	0.05980 (2.26)*
satis	-0.2502914 (-1.36)	0.2663 (1.64)
act_rate	-0.0017647 (-0.09)	-0.04888 (-3.47)**
_cons	0.7417398 (0.32)	-4.001 (-1.89)
hazard		
Lambda	-0.1390521 (-2.62)*	-.1049041 (-2.23)*
Observations	88	88
Z-statistics in parentheses		
* significant at 5% level; ** significant at 1% level		
Computation with treatreg command in two steps		

The corrected wage equation does not really change from the table 2 with OLS without correction for self-selection, even though lambda, which measures the impact of self-selection is significant, implying that there is self-selection into the program. The binary variable cert, which measures the treatment effect on the whole sample, is significant. Recall that is was not the case with the simple OLS regression, which didn't take endogenous selection into account. This leads to the conclusion that the single treatment has an impact on the wage when it is corrected for self-selection. The main conclusion from this is that self-selection is efficient, meaning the individuals choose the right scheme for themselves. In other words, the students that apply for themselves, want to join the training because they know that they will benefit, while those who do not apply would not benefit. This implies that students that are sent by their employers or through the reintegration program of invalidity insurance would not benefit from the training. This can be explained with the fact that they do not select themselves into the program.

The column 4.2, presents the same than 4.1 but considering the double treatment only. This implies that the results will tell us something about the double treated group compared to the non-treated group and the single treated group together. This method does not allow us to separate the non-treated from the single treated group anymore and therefore needs to be considered as a single group. In this selection equation, only age and activity rate are significant. Here age has the opposite sign than in 4.1. This could be interpreted as a kind of maturity that is needed to pass the brevet fédéral exam. The activity rate has a negative sign, which lead us to the conclusion that full employed people have less time to study and are therefore less likely to pass the exam. The probability of being double treated does not depend on the nationality anymore, and therefore there is no nationality based discrimination.

First note that in both cases the lambda is significant and therefore the self-selection plays a role. This implies that the coefficients estimated with OLS are biased. All the variables in the wage equation are now significant except for age. The coefficients are similar to the OLS so the same analysis holds. Both binary variable, which measures the average treatment effect, are highly significant. So the brevet and the CFI certificate leads to a significant higher wage.
This analysis shows that when we correct for self-selection and when we consider the treatment independently, then both treatment have a significative impact on wage. This shows that the self-selection is efficient.

6 Conclusion

Average treatment effects showed that training of low skilled labour force can increase the wage when the treatment includes an officially recognised title in the Swiss educational system. The effect of the CFI certificate is less obvious. It does not influence the wage in almost all the estimated cases except for the correction for sample selection and endogeneity of the treatment. Since this is the most sophisticated method which relaxes most of the assumption, we may deduce that the treatment effect is consistently estimated and therefore conclude that even the CFI certificate has a 25% increase on wage. This shows that self-selection plays an important role, and implies that people that want to join the program without any external pressure can benefit from training. It also suggests that people who do not self-select into the program, but are sent by the employer or by the reintegration program from the invalidity insurance would not get the benefits. Therefore, CFI should not accept those candidates are sent from an institution.

One explanation for the importance of self-selection for people who get only the certificate could be that most of time these people do not have any end of apprentice title (CFC). For them, the CFI certificate is a way validate un-formal education[10], showing that they are willing to learn. In this view, the certificate can be a signal that they are ready to learn and are capable of adaptation. The brevet instead has a positive impact on wage ever without self-selection. The brevet is a guarantee that the person has reached a given level of competences, which is the same all over Switzerland. Therefore it can be understood as a signal for competences.

This research shows that officially recognize titles are more valuable than others, as it is not only a recognised signal for competences but it is also an access to the official Swiss educational system. Indeed 2/3 of the persons who got the brevet do not have an other officially recognised title. Therefore CFI should lobby for a new officially recognised exam in the cleaning sector under art. 41 in the new educational law, which wants to address the recognition of non formal education.

A more qualitative analyses of the survey can and should be done in order to understand what the factors that motivate people to join a training program are. After having understood this phenomenon, the training offer should be adapted in order to motivate low skilled labour force to get training.

[10] in German “informell erworbene Kompetenzen”

7 Bibliography

DRUCKER P,, *The Post-Capitalist Society*, Harperbusiness, 1994

IMBENS Guido W & ANGRIST Joshua D, 1994. -*Identification and Estimation of Local Average Treatment Effects*-Econometrica, Econometric Society, vol. 62(2), pages 467-75, 1994

HECKMANN J. - *Randomisation and social evaluation programs* - in Evaluating Welfare and Training Programs, ed C.T: Manski and Grefinkel, Haward University Press, pp 201-230, 1992

HECKMANN J. - *Instrumental variable: A study of Implicit Behavioural Assumption Used in Making Program Evaluation* - Journal of Human Resources 32, p 441-462, 1997

MINCER J. - *Investment in Human Capital and Personal Income Distribution* – journal of political economy, vol 66. No 4, pp 281-302, August 1958

OAXACA, R. *Male-Female Wage Differentials in Urban Labour Markets*, International Economic Review Vol. 14 (3), 693-709, 1973

ROSENBAUM Paul R & RUBIN Donald B. - *The Central Role of the Propensity Score in Observational Studies for Causal Effects* -Biometrika, Vol. 70, No. 1. pp. 41-55. Apr., 1983

WOOLDRIDGE Jeffrey M. –*Econometrics of cross section and panel data* – MIT 2002

8 Appendix:

8.1 *The survey at CFI*

Questionnaire destiné aux responsables dans le secteur de l'intendance en Suisse romande
Merci de cocher s'il y a une case et d'écrire s'il y a une ligne

Sexe : ☐ F ☐ M
Nationalité : ☐ Suisse ☐ Etrangère________________

Age : _____________ ans

Nombre d'enfants : _____________

Statut actuel ☐ employéE ☐ cadre ☐ cadre avec responsabilité jusqu'à 10 plus que d'un petit groupe subordonnéEs 10 sub.

Formation
Scolarité obligatoire :
En Suisse ☐ ☐ A l'étranger

Combien d'années : _____________

Formation professionnelle :
(plusieurs choix sont possibles)

Certificat fédéral de capacité ☐ Titre ___________________________
Certificat cfi ☐ Année ________________________
Brevet fédéral d'intendant/E ☐ Année ________________________
Autres titres : ☐ Lesquels ______________________

Expériences
dans le secteur de l'intendance ☐ Hôtellerie ☐ EMS ☐ Hôpitaux ☐ Ailleurs
(plusieurs choix sont possibles)
Nombre d'années totales _____________

hors du secteur de l'intendance ☐ Lesquelles________________

Travail bénévole ☐ _________________________
(garde d'enfants, associations, etc.)

Activité actuelle
Taux en % __________ %

Salaire net : Fr. ___________________
Salaire reçu à la fin du mois

Etes-vous imposé à la source ? ☐ oui ☐ non

Quel est votre indice de satisfaction au travail sur une échelle de 1 à 10

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 (satisfaction maximum)

SI VOUS AVEZ SUIVI UNE FORMATION EN INTENDANCE

1. Pour quelles raisons avez-vous entrepris une formation ? ____________________

2. Est-ce qu'elle vous apporté quelque chose sur le plan professionnel ?
☐ oui ☐ non
si oui, que vous a-t-elle apporté ? ____

3. Est-ce qu'elle vous apporté quelque chose sur le plan personnel ?
☐ oui ☐ non
si oui, que vous a-t-elle apporté ? ____

Qu'est-ce qu'il vous a manqué dans votre formation ? Qu'est-ce qu'il vous aurait encore été utile ? Qu'est-ce que vous auriez aimé en plus ?

____

5. Quels sont les changements positifs ou négatifs que vous avez constatés après votre formation ?

____

6. Remarques éventuelles :

____

SI VOUS N'AVEZ PAS SUIVI UNE FORMATION EN INTENDANCE

7. Pour quelles raisons n'avez-vous pas entrepris une formation ?
(plusieurs choix sont possibles)

☐ pas besoin
☐ employeur pas d'accord
☐ pas de temps
☐ pas envie
☐ pas de finance
☐ pas connaissance d'un lieu de formation
☐ pas assez confiance en moi
☐ pas d'occasion
☐ autres raisons __________________________________

8. Remarques éventuelles :

____

8.2 Raw data of the survey and description

Variable	Description	Obs	Mean	Std. Dev.
Wage	Wage in the survey (CHF)	92	4464.749	1340.435
Fwage	Wage calculated for 100%	92	4948.33	1114.954
Lwage	Log of fwage	92	8.481702	0.2266406
Age		92	43.8587	9.490981
age2	Age squared (years)	92	2012.685	774.3622
age3	Age^3	92	95379.03	52334.25
Edu	Level of education, 1-5 Count data, 1 for each level of recognised education reached Ex: obligatory school =1, cfc =2	92	2.369565	0.9575311
Nocfi	Bin =1 if no cfi training	92	0.4565217	0.5008354
Cert	Bin =1 if only certificate cfi	92	0.2826087	0.452735
Brevet	Bin = 1if brevet fédéral	92	0.2717391	0.4472937
Gend	Bin =1 if woman	92	0.8695652	0.3386266
year_exp_int	Year of experience in the sector	88	15.11364	8.641562
year_exp_I^2	The square of it	88	302.25	317.7653
other_exp	Bin =1 if other experience	92	0.5434783	0.5008354
Swiss	Bin 1 if swiss	92	0.7065217	0.4578508
Child	Number of children	92	1.521739	1.10422
child_wom	Nb of children if a woman	92	1.26087	1.098146
Statut	Position at work	92	2.163043	0.9051401
act_rate	Activity rate	92	90.03261	16.20233
Part	Bin=1 if part time job	92	0.3804348	0.488154
Volun	Bin=1 if volunteering	92	0.2173913	0.4147311
Satis	Satisfaction at work	92	7.869565	1.233394

8.3 *Computation of treatment effect with single treatment and its extension to double treatment*

Assume the following basic process that defines the output y :
$y_1 = c_1 + x\beta_1 + u_1$ if w=1
$y_0 = c_0 + x\beta_0 + u_0$ if w=0
where w is the treatment

8.3.1 Case 1.1 : homogeneous treatment and random selection

When we assume random selection and ignorability of treatment, then we have the simplest way to compute treatment effects. Then the average treatment effect is given by :

$ATE = E(y_1 - y_0) = E(y_1) - (y_0)$. An other quantity of interest is the average treatment effect on ht treated which is given by $ATE1 = E(y_1 - y_0 | w = 1)$ When w is statistically independent on the outcome (ignorability of treatment) and random selection then the average treatment effect on treated is equal to average treatment effect : $ATE1 = E(y_1 - y_0 | w = 1) = E(y_1 - y_0) = ATE$

Econometrically the problem is reduced to a simple OLS estimation in which appears the dummy variable. If the dummy is significant it measures the difference between treated and non-treated. The level of the fixed effect can be measured with a model without a constant but two symmetric dummies. So we can get a homogenous effect, an effect which is considered to be the same for the whole sample. In the case of double treatment two dummies are introduced in order to measure both treatments. The analysis stays the same.

8.3.2 Case 1.2 : heterogeneous treatment and no self-selection

In this case the treatment matters and we cannot assume anymore $E(y_0 | x, w) = E(y_0 | x) = E(y | x)$. In such a case we are interested in the average treatment effect on the treated group or on the non-treated group. In the contrary of case 1, when the effect was calculated on the whole sample, the ATE is not equal to ATE1 anymore but is given by :

$ATE1 \equiv E(y_1 - y_0 \mid w = 1) = E(y_1 \mid w = 1) - E(y_0 \mid w = 1)$

In this case we face the problem of missing data. Indeed y_0 is known only in the case the individual didn't get the treatment, so $y_0 | w=1$ is unknown. Mathematically we have :

$E(y_1 \mid x, w = 1) = x\beta_1 + E(u_1 \mid x, w = 1)$
$E(y_0 \mid x, w = 1) = x\beta_0 + E(u_0 \mid x, w = 0)$

and we are missing

$E(y_0 \mid x, w = 1 = x\beta_0 + E(u_0 \mid x, w = 1)$

So we need to estimate it consistently. Econometrically we do an OLS regression over each group, and predict the values over the whole sample.

$$ATE \equiv \bar{x}(\beta_1 - \beta_0) = \overline{x\beta_1} - \overline{x\beta_0}$$
$$ATE1 \equiv \overline{x_1}(\beta_1 - \beta_0) = \overline{x_1\beta_1} - \overline{x_1\beta_0}$$

This implies that we can simply compute the difference between the predicted value of the regression where w=1 and the predicted value of the regression where w=0. The average of this difference will give the average treatment effect and the average of this difference on the treated group only will give the average treatment on treated.

In the case of double treatment we can compute also the average treatment effect on the double treated group and compare it to the two other groups. These computation are given by :

$$ATE2(0) \equiv \bar{x}_2(\beta_2 - \beta_0) = \overline{x_2\beta_2} - \overline{x_2\beta_0}$$
$$ATE2(1) \equiv \overline{x_2}(\beta_2 - \beta_1) = \overline{x_2\beta_2} - \overline{x_2\beta_1}$$

Where ATE2(0) is the comparison between the double treated and the non-treated group and ATE2(1) is the comparison between the double treated and the single treated group.

8.3.3 Case 2.1: self-selection and homogenous treatment

Self-selection occurs when we relax the assumption of randomness into the selection program. This allows us to introduces the fact that individuals at least partly determine whether they receive treatment or not and their decision may be related to the benefit of the treatment. In other word the people who decide to get a treatment for which they have to pay do it because they expect to benefit of the treatment. So it might be that the people who do not select themselves into the treatment do it because they know that they will not benefit from it. This is a very important issue, since neglected self-selection leads to correlation of the residuals and provokes an omitted variable bias.

In such a case it turns out that ATE1 can be consistently estimated as a difference in means under the weaker assumption that w is independent of y_0, without placing any restriction on the relationship between w and y_1 (Wooldridge, 2002). To show this :

$$E(y \mid w = 1 = -E(y \mid w = 0) = E(y_0 \mid w = 1) - E(y_0 \mid w = 0) + E(y_1 - y_0 \mid w = 1)$$
$$E(y_0 \mid w = 1) - E(y_0 \mid w = 0) + ATE1$$

when y is independent of w the first two term cancel out and we can ignore the effects of treatment. The case 2.1 will assume this and correspond to case 1.1 but with self selection. It turn out that to estimate the average treatments effects a 2 step procedure is used. In the first step the propensity score, the predicted probability of a discrete response model is computed and denote p(x). This new variable is then introduced in the wage equation and corrects for the self-selection.

$$ATE = E(\frac{[w - p(x)]y}{p(x)[1 - p(x)]}$$

where p(x) represents the probability to be in the treatment. Note that under the assumption of ignorability of treatment equal ATE1.

Econometrically the 2-step procedure consists in the first step to compute the probability to join a training program with a logit or probit estimation. The response probability for treatment is called the propensity score, which is the predicted probability of a logit or probit model. In a second step this propensity score is introduced in the basic wage equation, which is therefore corrected for self-selection. The estimated model be become $y = c + \alpha w + \delta \hat{p}(x) + v$ where $\hat{p}(x)$ is the propensity score. The coefficient α is a consistent estimator of the homogenous effect with self-selection under the assumption that $E(y_1 - y_0 \mid x)$ is uncorrelated with $\mathrm{var}(w \mid x) = p(x)[1 - p(x)]$.

In the case of a double or different treatment the selection equation has to be estimated with a multinomial response model. Ordered probit will be used if the treatments can be ranked and multinomial logit will be used if there is the choice between different treatment which cannot be rank. Out of the selection equation propensity scores can be computed representing the probability to get a given treatment. Then theses propensity scores are introduced in the outcome equation and the estimated model becomes $y = c + \alpha_1 w_1 + \alpha_2 w_2 + \delta_1 \hat{p}_1(x) + \delta_2 \hat{p}_2(x) + v$. The analysis of the coefficient of the treatment stays the same than in the case of one treatment.

Note the coefficient of the propensity scores measures the impact of self-selection. If these coefficients are significantly different from zero then the selection process is important and the assumption of random selection will lead to misleading results.

8.3.4 Case 2.2 : self-selection with propensity score and heterogeneous effects

This section is basically the same case that 2.1, except that now we relax the ignorability of treatment. Like to case 1.2 the aim is to compute heterogeneous effects, this is effect for each group separately but including self-selection. The difference is that now the self-selection equation has to be estimated for each group separately and its propensity score computed. One can show that the average treatment effect and the one on treated are given by:

$$ATE = E(\frac{[w - p(x)]y}{p(x)[1 - p(x)]}$$

$$ATE1 = E(\frac{[w - p(x)]y}{[1 - p(x)]} / P(w = 1)$$

For the proof see Wooldrigde p 615.

The formula for average treatment effect stays the same than in case 2.1. In order to estimate the average treatment effect on treated we need to compute:

$$ATE1(x) \equiv E(y_1 - y_0 | x, w = 1) = E(y_1 | x, w = 1) - E(y_0 | x, w = 1)$$

The first part of the computation is simply the mean of the predicted wage for people who got the treatment. In order to compute the second part of the equation, a estimation of the selection equation is needed. The potential wage can be computed including the correction for the self-selection process, which is different for each group.

The main issue arisen for this computation is the problem of common support. Common support implies that an agent in one group should have the same propensity score that an other one in the other group so that only comparable agents would be compared. Basically the agents for which so similar one can be found in the other group are dropped out of the sample. This matching approach suggests first by Rosenbaum and Rubin is motivated by the following thought experiment. Suppose we choose propensity score p (x) at random from the population. Then, we select two agents from the population sharing the chosen propensity score where one agent receives treatment and the other does not. An estimation strategy requires estimating the propensity scores estimated the response differences for pairs matched on the basis of estimated propensity score, and the averaging over all such pair" (Wooldridge 2002 p620). But since such kind pair a quite rare grouping must be done into cells or local averaging.

8.3.5 Case 3 : endogeneity of treatment and no self-selection

Now assume that there is no self-selection but the treatment is influenced itself by the variables included in the model. In other word the ignorability of treatment assumption does not hold anymore. This can be the case when for example people are randomly made eligible for a voluntary job-training program. The ignorability of treatment assumption would imply that the participation decision would be unrelated to what people would earn in the absence of the program, which is quite unlikely. In order to be able to compute ATE and ATE1 we use the following model :

$y_0 = E(y_0) + v_0$ and $y_1 = E(y_1) + v_1$ then $y_1 - y_0 = [(E(y_1) - E(y_0)] + (v_1 - v_2)$

With this model we have: $ATE1 = ATE + E(v_1 - v_0 | w = 1)$

The last term of this expression can be interpreted as the person-specific gain from participation. So ATE1 and ATE differ. Introducing individual characteristic and relevant pre-treatment outcome can solve this problem.

8.3.6 Case 4: endogeneity of treatment and self-selection

8.3.6.1 Self-selection based on the propensity score and instrumental variable

In this case again we do not assume ignorability of treatment and the randomness of selection anymore. Like in other cases of endogeneity, the solution can be found with an appropriate instrument, a variable that is highly correlated with the endogenous explanatory variable and not with the explained variable. Basically the estimation becomes a tree stage procedure where in the first step the probability of treatment is estimated given the explanatory variable and the instrument $P(y|x,z)$, take out the propensity score and include it into the wage function, which is estimated with a two-stage procedure. It consist in regressing first the endogenous explanatory variable on the instrument then including the result in to the wage equation which is corrected for self-selection by introducing the propensity score and corrected for the endogeneity of the explanatory variable by introducing the instrument.

8.3.6.2 Self-selection based on unobserved elements

Until now we considered a

To see this assume the following model:

$$
\begin{aligned}
y_1 &= x\beta_1 + u_1 \\
y_0 &= x\beta_0 + u_0 \\
w^* &= z\delta + v \\
w &= \begin{cases} 1 \Rightarrow z\delta + v > 0 \\ 0 \Rightarrow z\delta + v < 0 \end{cases}
\end{aligned}
$$

Where w* represents the latent model behind the selection process and w is the observed treatment binary. Furthermore, assume that the residuals v are normally distributed, N(0,1).

Then the expected residuals are given by

$E(u_1|x, v > -z\delta) = \lambda_1(z\hat{\delta}) = \frac{\phi(z\hat{\delta})}{\Phi(z\hat{\delta})}$ [11] and $E(u_0|x, v < -z\delta) = \lambda_0(z\hat{\delta}) = \frac{\phi(z\hat{\delta})}{1-\Phi(z\hat{\delta})}$

where ϕ is the pdf of a normal distribution and Φ is the cdf.

The lambda is therefore the correction for the residuals u and is also called the inverse mills ratio.

So the model to estimate becomes

$y = c + \beta x + \alpha w + \gamma_0\lambda_0 + \gamma_1\lambda_1 + v$ where α is a consistent estimate of the homogenous treatment effect since λ_0 & λ_1 correct the bias due to omitted variable and therefore correct for self-selection. γ_0 & γ_1 measures the correlation of u_i and v. Indeed, $E[u_i|x, z, v > -z\delta] = \rho\lambda_i$ where $\rho = \mathrm{cov}(u_i, v)$.

The extension of this method to more than one treatment is quite complicated, and in the case of ordered probit as selection equation. The correction factor is not an inverse mills ratios anymore, as it needs to consider 3 probabilities (no training, cfi training only, brevet) . Recall that the lambda could be computed thanks to the symmetry of normal distribution, with tree probabilities, there are

[11] Using the formula for a truncated random variable given by : $f(x|x > a) = \frac{f(x)}{\Pr(x > a)}$

two cut off point of the normal distribution and therefore the correction factors cannot be computed as easy as before anymore.

8.4 Estimation results used for heterogeneous effects with random selection

A	A1 if nocfi	A2 if certificat	A3 if brevet
	lwage	lwage	lwage
edu	0.110 (1.91)	-0.027 (0.46)	0.024 (0.43)
year_exp_int	0.008 (1.87)	0.006 (1.61)	0.000 (0.01)
age	0.007 (1.99)	0.006 (1.89)	0.000 (0.10)
gend	-0.231 (2.66)*	-0.216 (2.30)*	-0.164 (1.92)
Constant	7.981 (40.01)**	8.421 (35.81)**	8.729 (30.34)**
Observations	39	25	25
R-squared	0.35	0.36	0.17
Absolute value of t-statistics in parentheses			
* significant at 5% level; ** significant at 1% level			

Table 5 : estimation of the wage equation for each group

For the equation A3, the F-test does not reject the hypothesis that all the coefficients are zero. For the A2 this rejection is done at 7%.

8.5 Results of the selection equation for the computation of the propensity score

Table 6 : result of the selection equation

cfio	
year_exp_int	0.0421387 (2.64)**
age	-0.0057619 (-0.42)
gend	-0.0293299 (-0.07)
swiss	0.5670172 (1.97)*
child_wom	-0.0014612 (-0.01)
_cut1	0.5650575
_cut2	1.363894
Absolute value of z statistics in parentheses	
* significant at 5%; ** significant at 1%	

Note that since edu was constructed on the base of the acquired title, it was almost multicolinear with the variable cfio, which is the variable that measures the probability to be in a given group.

Therefore education has been drop and the Mincer model without the non linear relationship has been estimate.

Note that the T-test does not reject the hypothesis that all the coefficients are zero at 8%.

8.6 *The do-file for the computation on stata*

Estimation of the Mincer equation:

```
reg lwage edu year_exp_int age age2 age3 gend swiss child_wom
hettest
reg lwage edu year_exp_int age age2 age3 gend swiss child_wom

reg lwage edu year_exp_int age gend
```

Introduction of the dummies

```
reg lwage edu year_exp_int age gend cert brevet
```

we can conclude that certificate does not have an impact but the brevet has one since it's significantly different from 0 and the coefficient measures the difference between the first and third group

```
reg lwage edu year_exp_int age gend nocfi cert brevet, nocons

test cert=brevet
test cert=nocfi
test nocfi=brevet
```

heterogeneous effect without propensity score

```
 reg lwage edu year_exp_int age gend
 reg lwage edu year_exp_int age gend if nocfi==1
predict y0

reg lwage edu year_exp_int age gend if cert==1
predict y1

reg lwage edu year_exp_int age gend if brevet==1
predict y2

gen ate20 = y2-y0
sum ate20
sum ate20 if nocfi==1
sum ate20 if cert ==1
sum ate20 if brevet==1

gen ate21 = y2-y1
sum ate21
sum ate21 if nocfi==1
sum ate21 if cert ==1
sum ate21 if brevet==1

gen ate10 = y1-y0
sum ate10
sum ate10 if nocfi==1
```

```
sum ate10 if cert ==1
sum ate10 if brevet==1
```

self-selection based on the propensity score with ordered probit

```
gen cfio= certificat + brevet
```

cfio take the value 2 if brevet (and therefore also cert) and value 1 if only brevet and value 0 is no training

```
sum cfio
oprobit cfio year_exp_int  age gend child_wom
predict p0 p1 p2

reg lwage cert brevet p1 p2
```

Note that 1 propensity score must be dropped

hausman test

```
reg cert edu year_exp_int age gend volun
predict certhat
reg brevet edu year_exp_int age gend volun
predict brevethat
reg lwage edu year_exp_int age gend cert brevet certhat
reg lwage edu year_exp_int age gend cert brevet brevethat
```

self-selection on unobservable elements with treatreg command

```
treatreg lwage edu year_exp_int  age gend if brevet <1 , treat(cert = swiss
statut age satis act_rate)twostep
```

Note that brevet <1 allows us to compare the non treated group with the single treated group!

```
treatreg lwage edu year_exp_int  age gend, treat(brevet = swiss statut age satis
act_rate ) twostep
```