Matched-pair analysis based on business survey data to evaluate the policy of supporting the adoption of advanced manufacturing technologies by Swiss firms

Laurent Donzé

April 2003

Department of Quantitative Economics
University of Fribourg
Av. de Beauregard 13
CH - 1700 Fribourg
Switzerland
Email: Laurent.Donze@UniFr.ch

Abstract

We apply different matched-pair methods to evaluate the policy of sustaining the adoption of advanced manufacturing technologies by Swiss firms and we compare it in decomposing the selection bias. In this aim we present the evaluation problem, the decomposition of the selection bias and the estimation techniques used. We base our study on the J. Heckman and al. (1998)'s paper. The empirical results show that the matched-pair method proposed by J. Heckman and al. (1998) is not necessarily better than the alternative methods usually used. The multiple imputations technique, which we propose too, appears to be remarkably good. The impact of the policy of sustaining the new technologies is weak.

Key words: Advanced Manufacturing Technologies, Matching, Propensity Score, Selection Bias, Multiple Imputations, Swiss Firms, Business Survey, Adoption of Technologies, Participation to a Programme

JEL Classification: C4, C5, L2, L6

1 Introduction

In 1996, the KOF ETH Zurich investigated, in the framework of its business innovation survey in the industry sector,[1] the adoption by the Swiss firms of advanced manufacturing technologies (AMT). The questionnaire had to permit the evaluation of the Swiss policy of sustaining in this area, support on the one hand from the "Commission for encouragement of the scientific research" or on the other hand in the framework of the "CIM action program". We have from the outset to emphasize that one of the main characteristics of the Swiss technological policy is that it is less oriented to direct measures than to put in place the general conditions

[1] Cf. for example Arvanitis and al. (1998).

in favour to the introduction, the development and the propagation of new products and new technologies. And indeed, this is in this way that we ought to appreciate the sustaining programs just mentioned which were initiated in 1990 and stopped in 1996.

The principle of these programs was to offer to the firms information and formation services as well as subsidies for consultation and development projects, principally in the form of joint-ventures between enterprises or between enterprises and research institutions. These elements were the core of the questionnaire[2] along with information about as instance number of AMT components introduced by the enterprise,[3] market conditions which the enterprise has to face and consequences on different parameters (turnover, employment, etc.) of the introduction of new technologies.

By using a probit simultaneous equations model, Arvanitis and al. (2002) showed that generally the policy conducted by the Swiss authorities seems to have produced from the enterprises side a more intensive adoption of AMT, specially for those who didn't use them before the beginning of the program. The acquired experience leads to our participation in 2000 to a similar study conducted for the Austria.[4] In complement to the estimation of a simultaneous model, an analysis based on the matching method was made which one gave interesting results too. Thus the idea to complete the Swiss study by such an analysis.

The matching methods appeared in the 1970. Although one can apply them in different contexts, they are very adapted to evaluate economic policy. Recently, thank to the impulsion particularly of James J. Heckman, they benefit from a renewal of attention. However, as we will see, they are not without faults. We plane to investigate a some matching methods under the scope particularly of the selection bias appearing during their utilisation. We will propose specifically an alternative matching method: the multiple imputations. We will use the Heckman's method too. These two additions made the central points of the study.

This study is structured in the following manner. First, we describe briefly in section 2 the data which we dispose as well the variables that we generated. We give in section 3 the essential of the results obtained by Arvanitis and al. (2002). We will be able then to enter veritably in the subject in posing in section 4 the evaluation problem. We present shortly in section 5 the basic principles of the matching methods. Then, in section 6, we explain how one can measure the selection bias. In section 7, we talk some words from the multiple imputations method. In section 8, we will see how one can estimate non parametrically the selection bias and, in section 9, how estimate the components of the selection bias as well as the mean selection bias. We present in section 10 the estimator of the impact of a treatment. In section 11, we describe briefly the matching with the Local linear regression-adjusted method proposed by J. Heckman. At last, the section 12 will be devoted to the empirical results. We will able to conclude our study.

2 Data description

With the aim of comparing the results of our study with those of Arvanitis and al. (2002), we use as it is their data base. The data and the variables of our study are largely described in Arvanitis and al. (2002). We give in the following only the main characteristics. First, the (business) data are issued from a survey made in 1996 in the Swiss industry sector. The response rate was at about 34 %. Although this rate may seem rather low, its level, which is in the mean for such a survey, is totally acceptable. Due to limited resources an analysis of the nonresident was not conducted at that time. Moreover, we will not apply a weighting factor

[2]The questionnaire can be downloaded in French, German and Italian from http://www.kof.gess.ethz.

[3]When we speak about the adoption of a computer manufacturing technology, we think to the introduction and utilisation by the enterprise of an AMT component. These are many and of different nature depending on the section in the firm. Arvanitis and al. (2002) give a list of them. One can note as instance the components CAD/CAE (Computer-aided design and/or engineering) for the design; CAP (Computer-aided (manufacturing) planning) for the planning; CNC/DNC (Computer numerically controlled machines) for the manufacturing, etc.

[4]Cf. Geyer and al. (2000)

in the data analysis. At last, about 80 % of the respondents pretended to use an AMT component in 1996.

The final data base with witch we made our estimations contains 463 enterprises, all of them using at least one component AMT in 1996 or scheduling to use at least one until 1999. These data are remarkably representative with respect to the economic activities and to the enterprise sizes of the initial data. Apart from the initial variables of the questionnaire, other variables have been generated, notably with factor analyses.[5] The study of Arvanitis and al. (2002) describes and justifies the variables used in the formulation of their model of Economic policy and adoption of new technologies. We adopt in this paper the same framework. The Table 1 in annex gives us the complete list of the variables used.

3 Results obtained by a simultaneous equations model

In order to estimate the impact of the government sustain on adoption of AMT, Arvanitis and al. (2002) postulate a probit simultaneous equations model, the first equation modelling the adoption of AMT components and the second one the economic policy in the field. They completely justify the use of their model. They conclude the following. In taking into account all the set of observations they have, they note no significant influences of the economic policy variable on the adoption intensity. For the sub-sample of observations relative to enterprises which have adopted for the first time AMT after 1990, the result shows that the economic policy variable in the adoption equation has a positive coefficient and is statistically significant with a 10% significant level. In this case, the participation to the governmental program could have contributed to a greater adoption of new technologies from enterprises sustained with respect to the others. Thus the governmental policy could have been effective. A similar finding has been made in the case of a set of observations reduced to the small firms (less than 200 workers). At last, according to the kind of sustain (formation projects, consultation, R & D), the same effects have been found with yet a little precaution for the R & D. However the effect for the latter is positive and significant for the sub-sample of enterprises with less than 50 workers.

4 The evaluation problem

The problem we have to manage is known in the literature as *the evaluation problem*. A lot of articles and studies have been devoted to it, particularly in relation with the labor market. Though the finality is essentially empirical, the evaluation problem poses absolute theoretical challenges in connection to the selection problems and the selection bias. Pioneer in the matter, James J. Heckman has contributed significantly to the development of the research and in this paper we will base ourselves strongly on his works and those of his collaborators.[6]

Given a unit i – one individual or in our case an enterprise – likely to be subjected to a treatment or to benefit from a sustain programme. Admitting that the treatment has been. Noting by "$D = 1$" the event has received a treatment and by "$D = 0$" the event hasn't received a treatment. For a unit i, these two events are mutually exclusive. We are interested to the repercussions of a treatment on a variable Y. Design with Y_0, the level of the variable Y in case of no treatment and by Y_1 the one recorded in case of treatment. We want to estimate $\Delta = Y_1 - Y_0$. The evaluation problem appears because only Y_0 or Y_1 is observed for each unit i, but never simultaneously: $Y = DY_1 + (1 - D)Y_0$. As a consequence, it is a priori impossible to estimate Δ. With the help of additional information – on the characteristics of the units i – nevertheless one can generally rebuilt the missing data and thus solve the problem. We will see in the following how

[5] Arvanitis and al. (2002) or Arvanitis and Hollenstein (2001).

[6] The interested reader will find an excellent survey in Heckman and al. (1999). The main theoretical reference of this study is the paper of Heckman and al. (1998).

concretely this can be made. We have to solve the counterfactual hypothesis : "How high would have been the result Y if the unit i, which in reality has not participate to the treatment, had participated ?"

We are interested essentially to the mean treatment impact on the treated units defines as:

$$\Delta(X) = E(\Delta|X, D = 1) = E(Y_1|X, D = 1) - E(Y_0|X, D = 1). \tag{4.1}$$

We suppose that the conditional distributions of X satisfy the following equality $F(X|Y_0, Y_1, D) = F(X|Y_0, Y_1)$, i.e. that conditionally to the potential and realised results, D doesn't influence X. The data relating to the participants in the treatment permit us to estimate $E(Y_1|X, D = 1)$. On the other hand, it misses the data necessary to estimate $E(Y_0|X, D = 1)$. We use the method called "Comparison groups".[7] The data of the non treated units group are used to estimate the missing data. The method supposes that conditionally to X, the results Y_0 of the non participants (non treated) give an estimate of the results that the participants (treated) would have obtained if they had not participate in the treatment. We have thus $E(Y_0|X, D = 0) \cong E(Y_0|X, D = 1)$. If the missing data are generated with this hypothesis, a selection bias will appear. We define it as

$$B(X) = E(Y_0|X, D = 1) - E(Y_0|X, D = 0) \tag{4.2}$$

This selection bias is central in the analysis of the treatment impact. One has therefore to choose a construction method of the missing data that minimises this bias. The matching methods are currently in fashion, among others those that compare the units on the base of the participation probability. They take into account on great advantages and are relatively easy to use. Let us present briefly the principle.

5 The ground principles of the matching methods

Generally the matching method is based on the hypothesis that conditionally to X, Y_0 is independent from D. We note this hypothesis as:

$$Y_0 \perp D|X, \quad X \in \chi_c \tag{5.1}$$

for a given set χ_c where "$\perp$" denotes the independence and at the right from "|" are listed the variables of conditioning. This assumption permits us particularly to pretend that conditionally to X, the distribution of Y_0 given $D = 1$ is the same as the distribution of Y_0 given $D = 0$. In particular, if the expectation exists, we have

$$E(Y_0|X, D = 1) = E(Y_0|X, D = 0) \tag{5.2}$$

in such a way that punctually in X, the bias $B(X) = 0$.

[7] For an experience (scientific, social), we talk about "control group". In this case, the treatment and control groups are selected by a stochastic process, which is not the case for non experimental data.

The matching methods exploit intensively this assumption. The idea is to pair a non participant along with a participant who is according to a suitable measure "near" in terms of X. For each unit i from the set of participants, we allocate an estimated result $\hat{Y}_{oi}$ which can be calculated in different manners, e.g. in taking a weighting mean of the units j of the comparison group. In this case, we have

$$\hat{Y}_{0\,i} = \sum_{j \in \{D=0\}} W_{ij} Y_{0\,j} \tag{5.3}$$

where $\{D = 0\}$ is the set of indices of the non participants and $\{D = 1\}$ is the one of the participants, and with $\sum_{j \in \{D=0\}} W_{ij} = 1$ for all i.

The matching estimators essentially differ in the definition of the weights. In this study, we will test the following matching methods: (1) the nearest neighbour method; (2) the caliper method; (3) the kernel method; (4) the local linear regression-adjusted method; (5) the multiple imputation. The first three methods have several variants and are commonly used.[8] The fourth method is a proposition of Heckman and al. (1998). We will present it in section 11. Concerning the last method, it is an alternative that we propose and that we in the framework of this study want to test. We will present the major elements of it in section 7.

As we noted it above, matching techniques have been recently developed based on the work of Rosenbaum and Rubin (1983, 1985) which exploit certain properties of the probability to participate in treatment ("the propensity scores"). A theorem shows indeed that if the hypothesis (3) is satisfied, then

$$Y_0 \perp D | P(X) \quad \text{pour} \quad X \in \chi_c \tag{5.4}$$

provided that $0 < P(X) < 1$ for $X \in \chi_c$, i.e. it exists a positive probability that the events "$D = 0$" or "$D = 1$" would have been append for all the elements of χ_c.Therefore, in conditioning with respect to $P(X)$ rather with X, the conditional independence is maintained. This permits us to write

$$E(Y_0|P(X), D = 1) - E(Y_0|P(X), D = 0) = B(P(X)) = 0 \tag{5.5}$$

It results from (**??**) that the selection bias is null and that one can reduce favourably the dimension of the matching problem in solving it on the base of the scalar $P(X)$. The condition (**??**) is thus essential. Heckman and al. (1998) show how one can test it statistically. Applied to theirs data the test rejects it.

The great problem underlying to the use of the matching method is due to the fact that the distributions of the characteristics are not certainly identical for the comparison group and for the treatment group, and this even if the hypothesis (**??**) is satisfied. The distribution supports of X may be different in both groups and the distribution shapes may also be different for regions of a common support. Furthermore, there is for the matching based on characteristics X, the same uncertainty for the choice of the variables to use than for the specification of a conventional econometric model. Even if a set of values of X satisfies the condition (**??**), an augmented or reduced version doesn't necessary.

[8]One can find a description in e.g. Heckman and al. (1998).

6 The measure of the selection bias

A major contribution of Heckman and al. (1998) is their attempt to measure the selection bias. With the aim of doing this they will decompose the bias in several components each of those having a particular interpretation. The selection bias will thus rigorously be redefined. One can present the proposed decomposition.

Given $S_{1X} = \{X|f(X|D=1) > 0\}$, the support of X for $D = 1$, where $f(X|D = 1)$ is the conditional density of given D; $S_{0X} = \{X|f(X|D = 0) > 0\}$, the support of X for $D = 0$; $S_X = S_{0X} \cap S_{1X}$, the intersection region of both supports. One can rewrite the traditional measure of the bias $B = E(Y_0|D = 1) - E(Y_0|D = 0)$ as the sum of three components:

$$
\begin{aligned}
& B = B_1 + B_2 + B_3, \\
& \text{where} \\
& B_1 = \int\limits_{S_{1X}\backslash S_X} E(Y_0|X, D = 1)dF(X|D = 1) - \int\limits_{S_{0X}\backslash S_X} E(Y_0|X, D = 0)dF(X|D = 0), \\
& B_2 = \int\limits_{S_X} E(Y_0|X, D = 0)\,[dF(X|D = 1) - dF(X|D = 0)], \\
& B_3 = P_X \bar{B}_{S_X}.
\end{aligned} \tag{6.1}
$$

with $\bar{B}_{S_X} = \frac{\int_{S_X} B(X)dF(X|D=1)}{\int_{S_X} dF(X|D=1)}$ the mean selection bias; $P_X = \int_{S_X} dF(X|D = 1)$, the proportion of the density of X given $D = 1$ in the intersection region S_X; $S_{1X}\backslash S_X$ is the support of X given $D = 1$ which is not in the intersection region S_X; $S_{0X}\backslash S_X$ is the support of X given $D = 0$ which is not in the intersection region S_X.

The interpretation of these three terms is the following. The first one B_1 appears if the supports $S_{0X}\backslash S_X$ and $S_{1X}\backslash S_X$ are not empty. In this case, one cannot find the counterpart of $E(Y_0|X, D = 1)$ in the set $S_{0X}\backslash S_X$ or the counterpart of $E(Y_0|X, D = 0)$ in the set $S_{1X}\backslash S_X$. The term B_2 is issue from the difference in weighting of $E(Y_0|X, D = 0)$ by the both densities of X given $D = 1$ and $D = 0$. Finally, B_3 is due to the differences in the result which remain even after controlling for observable differences. The selection bias, defined as $\bar{B}_{S_X}$ can be of different size, or even of different sign than the traditional measure of the bias B.

What happens with the matching method about this ? If the method doesn't impose a common support for the matching, the first source of bias appears. The second component of the bias is eliminated if the matching is done with the help of the probabilities P of the participants. In this case, the matching weights effectively the data of the non participants. The last source of the bias is not eliminated with the matching. Thus, the $\bar{B}_{S_X}$ is the bias associated with a matching estimator.

7 The multiple imputations technique

We propose the multiple imputations technique as an alternative method to measure the impact of a treatment. In fact, we have had the possibility to exploit in other researches the potential of this method.[9] The latter has been proposed since several years, essentially in the framework of imputing missing values, by D.

[9] Cf. Donzé (2001).

Rubin and R. Little.[10] The general principle of the method is the following. Substitute for each missing values imputed values. The final estimator will be simply the mean of the imputed values. The question is to choose for each imputation made the right method. There exists of course several possibilities to impute the better than one can do the missing values. A general method, easy to employ, proposed by Rubin, is the *Approximate Bayesian Bootstrap* (ABB) technique. This bootstrap method provides proper imputation, in the sense defined by Rubin. The participation probabilities can serve as support of the imputation too.

We have extremely used the multiple imputations in the case of missing values and we propose to test it in this context of matching. Despite the undeniable advantages for the data analysis, the multiple imputations is rarely advocated as an imputation method. Recently yet, with the development of specialised software, the method becomes again popular. On the other hand, it is surprising that this technique has not been proposed in the case of data matching for the estimation of the impact of a treatment. There exists yet a lot of several common points between the matching methods and those of imputation. We think therefore that it is not without interest to evaluate the multiple imputations techniques in the present case.

8 Non parametric estimation of selection bias

Heckman and al. (1998) describe how estimate convergently the selection bias. The main steps are the following. First, from a traditional econometric model of selection, $Y_0 = X\beta + U_0$, we have $E(Y_0|P(X), D = 1) = X\beta + E(U_0|P(X), D = 1)$ and $E(Y_0|P(X), D = 0) = X\beta + E(U_0|P(X), D = 0)$. Thus the selection bias (**??**) can be written as

$$\begin{aligned} B(P(X)) &= E(Y_0|P(X), D = 1) - E(Y_0|P(X), D = 0) \\ &= E(U_0|P(X), D = 1) - E(U_0|P(X), D = 0). \end{aligned} \tag{8.1}$$

We define, for a unit i the bias functions $K_1(P_i) = E(U_{0i}|P_i, D_i = 1)$ and $K_0(P_i) = E(U_{0i}|P_i, D_i = 0)$. These functions will be estimated by the so called "*Double residual regression technique*" and non parametrically. The first thing to do is to estimate the participation probabilities P, denoted by $\hat{P}_i$. Those can be easily obtained by estimating for example a logit regression model. The regressors are the characteristics X common to the participants and non participants. On the other hand, one can postulate the following partial linear regression model[11]

$$Y_{0\,i} = X_i\beta + D_iK_1(P_i) + (1 - D_i)K_0(P_i) + \varepsilon_i. \tag{8.2}$$

We form an adjusted version of (**??**) by subtracting from it its conditional expectation with respect to P_i and D_i. We obtain

$$Y_{0\,i} - E(Y_{0\,i}|P_i, D_i) = [X_i - E(X_i|P_i, D_i)]\,\beta + \varepsilon_i. \tag{8.3}$$

We then estimate β from equation (**??**) by ordinary least squares. In this aim we first estimate the conditional expectations $E(Y_{0i}|P_i, D_i)$ and $E(X_i|P_i, D_i)$. Those are estimated non parametrically by using separately the observations on one hand for $D_i = 1$ and on the other hand for $D_i = 0$. We also use the suggestion of

[10] Cf. e.g. Little and Rubin (1987) or Rubin (1987).

[11] A partial linear regression model is postulate because it is supposed that the bias functions K_0 and K_1 are non parametric functions of continuous variables.

Heckman and al. (1998) to eliminate a small fraction of the data (2 %) for which the estimated density function $\hat{f}(\hat{P}_i|D_i = d)$, $d \in \{0,1\}$, is small.[12] This operation permits to guarantee a parametric estimator that is uniformly convergent.

The estimator $\hat{\beta}$ of β estimated in the first step permits us to calculate the adjusted residuals $c_i = Y_{0i} - X_i\hat{\beta}$. In a second step, we estimate then by a local linear regression of the residuals c_i on the probabilities $\hat{P}_i$. The regression model is written as

$$\min_{\gamma_{1d},\gamma_{2d}} \sum_{i\in\{D=d\}} \left[c_i - \gamma_{1d} - \gamma_{2d}(\hat{P}_i - P_0)\right]^2 G\left(\frac{\hat{P}_i - P_0}{a_N}\right), \quad d \in \{0,1\}, \tag{8.4}$$

where $\hat{\gamma}_{1d}$ is the estimator of $K_d(P_0)$, i.e. of the expectation $E(U_{0i}|P = P_0)$ in P_0 and $\hat{\gamma}_{2d}$ estimates convergently the first derivative of $E(c_i|P = P_0)$; P_0 is a given point of the support of $\hat{P}_i$ for $\{D = d\}$; G is a kernel and $\{a_N\}$ is a sequence of smoothing parameters;[13] $\hat{P}_i$ is the estimated value of P for unit i.[14]

9 Estimation of the selection bias components and the mean bias selection

One can obtain a non parametric estimation of the selection bias components given in (**??**). Indeed, Heckman and al. (1998) propose to compute those in the following manner

$$\begin{aligned}
&\hat{B} = \hat{E}(Y_0|D=1) - \hat{E}(Y_0|D=0) = \hat{B}_1 + \hat{B}_2 + \hat{B}_3 \\
&\text{where} \\
&\hat{B}_1 = \tfrac{1}{N_1} \sum_{\substack{i\in\{D=1\}\\ P_i\in S_{1P}\backslash S_P}} Y_0(P_i) - \tfrac{1}{N_0} \sum_{\substack{i\in\{D=0\}\\ P_i\in S_{0P}\backslash S_P}} Y_0(P_i)\,, \\
&\hat{B}_2 = \tfrac{1}{N_1} \sum_{\substack{i\in\{D=1\}\\ P_i\in S_P}} \hat{E}(Y_{0\,i}|P_i, D_i = 0) - \tfrac{1}{N_0} \sum_{\substack{i\in\{D=0\}\\ P_i\in S_P}} Y_0(P_i)\,, \\
&\hat{B}_3 = \tfrac{1}{N_1} \sum_{\substack{i\in\{D=1\}\\ P_i\in S_P}} \left[Y_0(P_i) - \hat{E}(Y_{0\,i}|P_i, D_i = 0)\right],
\end{aligned} \tag{9.1}$$

N_1 and N_2 are respectively the number of observations $D = 1$ and $D = 0$, $Y_0(P_i)$ is the value of Y_{0i} for the unit i with probability P_i, and where the supports S_P, $S_{1P}\backslash S_P$, $S_{0P}\backslash S_P$ are defined in similar manner as supports S_X, $S_{1X}\backslash S_X$, $S_{0X}\backslash S_X$ in (**??**).[15] We will estimate $\hat{E}(Y_{0i}|P_i, D_i = 0)$ by a local linear regression

[12] The details of this operation called "trimming", is given in Heckman and al. (1998), annex A.2.

[13] Heckman and al. (1998) propose to use a quartic kernel function. If the choice of the kernel function is not a problem – it is proving in practice that the results are not much sensible to the type of function used –, on the contrary those of the smoothing parameter, the "bandwidth", is more delicate to determine. We follow the rule $a_N = 2.7768(\hat{H}/1.34)N^{-1/5}$ where $\hat{H}$ is the interquartile range of $\hat{P}_i$.

[14] One can remark in fact that the problem to solve is those of a weighted least squares regression, weights being given by the kernel.

[15] In order to estimate the region of overlapping support S_P, we estimate non parametrically the densities $f(\hat{P}_i|D_i = d)$, $d \in \{0,1\}$, and we retain as common support the region where for each group the densities are positive.

model of Y_{0i} on P_i for the observations $D = 0$, i.e. one have to solve the problem

$$\min_{\gamma_1,\gamma_2} \sum_{i \leqslant N_0} \left[Y_{0\,i} - \gamma_1 - \gamma_2(\hat{P}_i - P_0) \right]^2 G\left(\frac{\hat{P}_i - P_0}{a_N} \right), \tag{9.2}$$

which has the same characteristics as (**??**).

One can show assuming a random sample that each term of the bias is estimated convergently and that, centred on the expectation, multiplied by $\sqrt{n}$, it is asymptotically normal.

The mean selection bias (MSB) is estimated from the bias functions $K_d(P_i)$, $d \in \{0,1\}$, computed according to the steps describe above. Furthermore, it will be estimated on the common support S_P. We have

$$MSB = \hat{\hat{B}}_{S_p} = \frac{1}{N_1} \sum_{\substack{i \in \{D=1\} \\ P_i \in S_P}} (\hat{K}_{1i}(P_i) - \hat{K}_{0i}(P_i)). \tag{9.3}$$

10 The estimator of the impact of a treatment

The estimation of the impact of a treatment or a programme of economic policy, in the framework of a matching analysis, is done simply by calculating weighted differences of output variables on a domain of study K. We have

$$\hat{M}(K) = \sum_{i \in \{D=1\}} \omega_i \left[Y_{1i} - \hat{Y}_{0i} \right] \quad \text{pour} \quad X_i \in K, \tag{10.1}$$

where ω_i is a weighting factor for eventual design plan, scale or heteroskedasticity problems, and $\hat{Y}_{0i}$ is the estimator (**??**) of Y_{0i} by matching. With $\omega_i = 1/N_1$, the impact of a programme will be estimated by the mean difference of realised and supposed results.

11 Matching according the local linear adjusted-regression method

Heckman and al. (1998) propose to use the local linear regression method as matching method and to extend the matching to the adjusted values of the outcome variables, i.e. the c_i of equation (**??**). In this new strategy, the weight W_{ij} of (??) are built as

$$\begin{aligned} & W_{ij} = \frac{A-B}{C-D} \ , \\ & \text{with} \\ & A = G_{ij} \sum_{k \in \{D=0\}} G_{ik}(P_k - P_i)^2 \ , \\ & B = G_{ij}(P_j - P_i) \left[\sum_{k \in \{D=0\}} G_{ik}(P_k - P_i) \right] \quad , \\ & C = \sum_{j \in \{D=0\}} G_{ij} \sum_{k \in \{D=0\}} G_{ik}(P_k - P_i)^2 \ , \\ & D = \sum_{k \in \{D=0\}} G_{ik}(P_k - P_i)^2 \ , \\ & G_{ij} = G\left(\frac{P_k - P_i}{a_N}\right) \ . \end{aligned} \tag{11.1}$$

12 Empirical results

We have computed with 5 matching techniques the estimators $\hat{M}(K)$ of (??) along with the selection bias decomposition resulting from the matching. The output variable of interest is $AMTDIFF$, the number of AMT components introduced between 1990 and 1996. For each method the matching is founded on participation probabilities at the program. These latter have been estimated by a logit model resumed in Table 2. To guarantee a greater homogeneity of the firms that we want match, we classify the firms in 5 cells and operate separately in each cell, i.e. the matching is made for each cell. The cells are generated according to the estimated participation probabilities and account each for one quintile of the distribution. The Figure 1 gives the densities of the participation probabilities.

The first method of matching is the "*Nearest neighbour method*". The second one is the so called *Caliper* where W_{ij} of (??) are simply defined as $W_{ij} = 1/n$, where n is the number of units Y_{0j} considered. Our third method is a kernel matching. This method defines the weights W_{ij} as

$$W_{ij} = \frac{G_{ij}}{\sum_{k \in \{D=0\}} G_{ik}} \tag{12.1}$$

where $G_{ik} = G((P_i - P_k)/a_N)$ is a kernel with a_n as smoothing factor.

The fourth method is the multiple imputations. Each values of an imputation is chosen according to the "Approximate Bayesian Bootstrap" method. We set at 5 the number of multiple imputations. The final result is the mean of the estimator obtained in each imputation.[16] At last, we apply the linear local regression-adjusted method.

First, we have made the analysis with the whole data set, taking nevertheless into account a trimming of 2 %. Then we have considered successively the following two sub-sets: the enterprises which have declared

[16] Cf. Donzé (2001).

none AMT component at the beginning of the programme, i.e. the data with INT90=0, and the enterprises with a number of employees less or equal to 200, i.e. the small firms. The results are in Table 4 to 6.

As we don't have imposed a common support for the matching, the term B_1 appears naturally in the decomposition, and B_2 is not eliminated by our matching methods. Table 3 gives the detail of the number of observations, before and after the trimming process, of the overlapping and non overlapping regions. Furthermore, considering the fact that B_3 is not eliminated by the matching, the question is to find the optimal method, i.e. those which will produce the minimum bias.

Our results show for all the matching methods and the domains of study an important bias of selection, though not significant. The components B_1 and B_2 represent a great part of this bias and this suggest to impose at least a common support for the matching. We have to note that the "Local linear regression-adjusted" method provides, as expected, the minimum bias, but the multiple imputations method appears to be surprisingly in this context a relative good one too.

Finally, considering our initial problem of measuring the impact of the policy of sustaining the adoption of advanced manufacturing technologies by Swiss firms, we have to note that this impact is, on the basis of the whole set of data, negative but not significant; calculated for the firms which don't have AMT components at the starting year of the programme, the impact is positive but always not significant; at last, for the small firms, the impact appears to be negative but not significant. These results confirm in some sense the findings of Arvanitis and al. (2002).

13 Conclusion

There was two goals to our study. First, we wanted to compare different evaluation methods of economic programmes and, secondly, to measure the impact of the Swiss policy of sustaining the adoption of advanced manufacturing technologies by Swiss firms. At the methodological level, our personal contribution is the use of the multiple imputations method to the matching problem. This latter appears to be as good as the methods usually used, e.g. the kernel method. The method proposed by Heckman and al. (1998) seems to be better. Nevertheless, others analyses have to be done in order to confirm this fact. Concerning the evaluation of the impact itself, we can confirm the results of Arvanitis and al. (2002). But it is also necessary to investigate others sub-samples of data and others output variables to strengthen our judgement.

14 References

1. **Arvanitis, S. and Hollenstein, H.** (2001): "*The Determinants of Adoption of Advanced Manufacturing Technologies - An Empirical Analysis Based on Firm-level Data for Swiss Manufacturing*", Economics of Innovation and New Technology, 10, pp. 377-414.

2. **Arvanitis, S.; Hollenstein, H. and Lenz, S.** (2002): "*The Effectiveness of Government Promotion of Advanced Manufacturing Technologies (AMT): An Economic Analysis Based on Swiss Micro Data*", Small Business Economics, Publishers, K. A., pp. 1-20.

3. **Arvanitis, S.; Donzé, L.; Hollenstein, H. and Lenz, S.** (1998): "*Innovationstätigkeit in der Schweizer Wirtschaft. Teil I : Industrie. Eine Analyse der Ergebnisse der Innovationserhebung 1996* ", Studienreihe Strukturberichterstattung, Bundesamt für Wirtschaft und Arbeit, Bern, 240 p.

4. **Donzé, L.** (2001): "*L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données : l'enquête 1999 KOF/ETHZ sur l'innovation*", Schweiz. Zeitschrift für Volkswirtschaft und Statistik, 137(3), pp. 301-317.

5. **Geyer, A.; Rammer, C.; Pointer, W.; Polt, W.; Hollenstein, H.; Donzé, L. and Arvanitis, S.** (2000): "*Evaluierung des ITF-Programms FlexCIM* ", S-0102, KOF/ETHZ, Joanneum Research, Austrian Research Centers, ZEW, OEFZS, Seibersdorf, Dezember 2000, 108 p.

6. **Heckman, J. J.; Lalonde, R. J. and Smith, J. A.** (1999): "*The Economics and Econometrics of Active Labor Market Programs*", in. **Ashenfelter, A. and Card, D.**, eds.:"H*andbook of Labor Economics*", Handbooks in Economics, Arrow, K. J. et Intriligator, M. D., Elsevier Science Publishers BV, Amsterdam Lausanne New York Oxford, Shannon Singapore Tokyo, Part 7 - The supply side, pp. 1865-2097.

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

8. **Little, R. J. A. and Rubin, D. B.** (1987): "*Statistical Analysis with Missing Data* ", Wiley series in probability and mathematical statistics, Applied probability and statistics, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 278 p.

9. **Rosenbaum, P. R. and Rubin, D. B.** (1983): "*The central role of the propensity score in observational studies for causal effects*", Biometrika, 70(1), 41-55.

10. **Rosenbaum, P. R. and Rubin, D. B.** (1985): "*Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score*", The American Statistician, 39(1), 33-38.

11. **Rubin, D. B.** (1987): "*Multiple Imputation for Nonresponse in Surveys* ", Wiley series in probability and mathematical statistics, Applied probability and statistics, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 258.

Figure 1: Estimated density of probabilities of participation

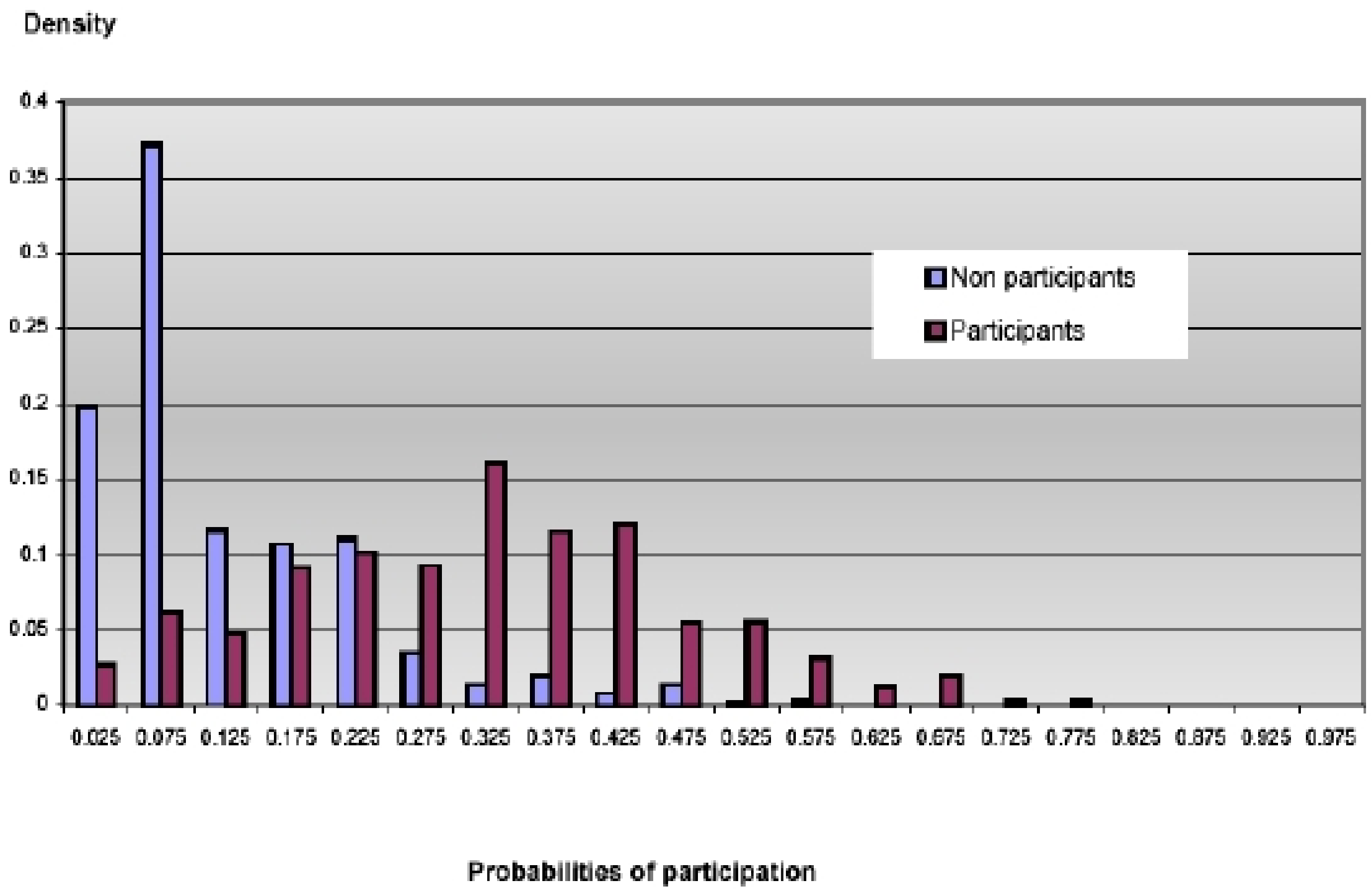

Table 1: **Description of the variables**

Variables	**Description**
Output Variables	
DAMTINT	Change of the AMT intensity (i.e. number of AMT element used) in the period 1990-1996
AMTDIF	Number of AMT components introduced in the period 1990 et 1996
Participation to the programme	
CIMPOL	Dummy variable for the support of the government
Objectives of / motives for the adoption of AMT	
FINCOMP	Favourable financial conditions; competitive pressure
COST	Cost reduction
FLEX	Higher flexibility
DEV	Improving product development
QUAL	Better product quality
BEST	Securing technological lead/"best practice"
Impediments to the adoption of AMT	
TECH	High technological costs/uncertainties
KNOWPERS	Lack of knowledge/lack of adequately qualified personnel
RESIST	Resistance to new technology within the firm
INVCOST	High investment costs
UTILIZ	Uncertainty with respect to capacity utilisation
COMPAT	Compatibility problems (e.g. with installed machinery, etc.)
Market conditions	
IPC	Intensity of price competition in the product market (five-point Likert scale)
INPC	Intensity of non-price competition in the product market (five-point Likert scale)
CONC0110	Dummy variable for the market concentration (1-10 competitors ; 50+ as reference group)
CONC1115	Dummy variable for the market concentration (11-15 competitors ; 50+ as reference group)
CONC1650	Dummy variable for the market concentration (16-50 competitors ; 50+ as reference group)
Type of production technique / products	
PDMARKET	Product differentiation
PDUSER	Products according to user specification
SBATCH	Small-batch production
LBATCH	Medium-batch/large-batch production
CONTFLOW	Continuous flow/mass production
Absorptive capacity	
HUMCAP	Percentage share of highly qualified employees
COOP	Cooperation in R&D activities (dummy variable)
Firm size	
L	Number of employees (full time equivalent)
L2	Square of number of employees (full time equivalent)
Control variables	
INT90	Intensity of AMT in 1990 (starting year of the programme)
METAL	Dummy variable for industry (Metalworking; others industries as reference group)
MACH	Dummy variable for industry (Machinery; others industries as reference group)
ELECT	Dummy variable for industry (electrical machinery/electronics; others industries as reference group)
CHEM	Dummy variable for industry (Chemicals/plastics; others industries as reference group)
AFFCOMP	Affiliate company
FOROWNER	Foreign owner
FINPROB	Financial problems
PREGOVSU	Previous government support

Table 2: **Logit Modelling of Response Probability (Dependent Variable: CIMPOL**

Model Variables	**Estimated Parameters**	**Standard values**
Constant	-0.1789	0.5563
FINCOMP	0.0527	0.0797
COST	-0.0413	0.0772
FLEX	0.0892	0.0778
DEV	-0.0455	0.0800
QUAL	-0.0100	0.0775
BEST	0.1589*	0.0778
TECH	-0.0280	0.0810
KNOWPERS	0.0567	0.0757
RESIST	-0.0243	0.0802
INVCOST	-0.0701	0.0877
UTILIZ	0.1697*	0.0798
COMPAT	-0.1101	0.0706
IPC	-0.1766*	0.0810
INPC	-0.1488	0.0796
CONC1650	0.0030	0.2358
CONC1115	-0.0194	0.2061
CONC0110	0.0868	0.1898
PDMARKET	-0.0149	0.1646
PDUSER	-0.0982	0.1866
SBATCH	0.0872	0.1658
LBATCH	-0.0286	0.1626
CONTFLOW	-0.0925	0.2166
HUMCAP	0.0126	0.00719
COOP	0.4759**	0.1710
L	0.2390	0.1838
L2	-0.0193	0.0175
INT90	0.0268	0.0238
METAL	0.2403	0.2236
MACH	0.1801	0.2508
ELECT	-0.0432	0.2624
CHEM	-0.2033	0.3423
AFFCOMP	-0.2656	0.1602
FOROWNER	0.1417	0.2303
FINPROB	0.3709*	0.1682
PREGOVSU	0.20190	0.1632
N obs.	463	
-2 Log	73.5981**	

Notes : 1) "N obs" is the number of observations, "-2 Log λ" is the likelihood ratio statistic to test the global dependency. 2) "**" significant at 1%, "*" significant at 5%.

Table 3: **Number of observations in and out the common support**

		Data before trimming		Common support estimated		Non common support estimated	
Domain of study	CIMPOL	N	In %	N	In %	N	In %
Whole data set	0	367	79.27	343	81.09	16	53.33
	1	96	20.73	80	18.91	14	46.67
	Total	463	100.00	423	100.00	30	100.00
INT90=0	0	73	87.95	26	78.79	45	95.74
	1	10	12.05	7	21.21	2	4.26
	Total	83	100.00	33	100.00	47	100.00
L<=200	0	248	81.31	168	77.78	75	91.46
	1	57	18.69	48	22.22	7	8.54
	Total	305	100.00	216	100.00	82	100.00

Table 4: **Impact and bias decomposition according to different matching methods (Output variable : AMTDIF**

Method	**Impact M**	$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	**B**	**MSB**
Nearest neighboor	-0.3333 (0.4039)	0.3845 [91.77] (0.1252)	-0.1130 [-26.97] (0.2104)	0.1475 [35.20] (0.3085)	0.4190 [-125.71] (0.2840)	0.2254 [-67.63] (0.2141)
Caliper	-0.0602 (0.2511)	0.2267 [108.31] (0.1124)	0.0116 [5.54] (0.1760)	-0.0290 [-13.86] (0.2013)	0.2093 [-347.67] (0.1485)	-0.0609 [101.16] (0.2073)
Kernel	-0.0918 (0.2512)	0.2016 [83.69] (0.1067)	0.0116 [4.82] (0.1760)	0.0276 [11.46] (0.1984)	0.2409 [-262.42] (0.1504)	-0.0113 [12.31] (0.2081)
Multiple imputations	-0.1813 (0.6162)	0.3158 [116.49] (0.243435)	-0.113 [-41.68] (0.2104)	0.06834 [25.21] (0.538626)	0.2711 [-149.53] (0.557453)	0.09072 [-50.04] (0.573304)
Local linear regression-adjusted	-0.0736 (0.2536)	0.2700 [212.10] (0.1418)	-0.1130 [-88.77] (0.2104)	-0.0295 [-23.17] (0.2913)	0.1273 [-172.96] (0.2840)	0.0125 [-16.98] (0.2143)

Notes : We give in parentheses the bootstrap standard values and in brackets for B_1, B_2 and B_3 the value in % of the bias with respect to the total bias B and for B and MSB, the value in % with respect to the mean impact M.

Table 5: **Impact and bias decomposition according to different matching methods (Output variable : AMTDIF; additional condition: INT90=0**

Method	**Impact M**	$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	**B**	**MSB**
Nearest neighboor	0.6000 (0.4039)	-2.1986 [601.20] (1.0567)	1.7885 [-489.06] (0.8544)	0.0443 [-12.11] (1.0035)	-0.3657 [-60.95] (1.8360)	-0.4774 [-79.57] (1.8683)
Caliper	1.6513 (0.2511)	-1.9946 [131.47] (1.0854)	1.7885 [-117.89] (0.8544)	-1.3110 [86.41] (0.5284)	-1.5171 [-91.87] (1.2783)	-1.7484 [-105.88] (1.6571)
Kernel	1.7410 (0.2512)	-1.9323 [126.98] (1.0961)	1.7885 [-117.53] (0.8544)	-1.3778 [90.54] (0.5453)	-1.5217 [-87.40] (1.2736)	-1.7882 [-102.71] (1.5741)
Multiple imputations	-0.1813 (0.6162)	-2.0986 [123.03] (1.178362)	1.7885 [-104.85] (0.8544)	-1.3956 [81.82] (1.24635)	-1.7057 [940.82] (1.806048)	-1.8613 [1026.6] (1.989755)
Local linear regression-adjusted	0.4151 (0.7852)	-0.7986 [-595.08] (1.4482)	1.7885 [1332.71] (0.8544)	-0.8556 [-637.6] (1.1953)	0.1342 [32.33] (1.6485)	0.1738 [41.87] (1.7446)

Notes : We give in parentheses the bootstrap standard values and in brackets for B_1, B_2 and B_3, the value in % of the bias with respect to the total bias B and for B and MSB, the value in % with respect to the mean impact M.

Table 6: **Impact and bias decomposition according to different matching methods (Output variable : AMTDIF; additional condition: L<=200**

Method	**Impact M**	$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	**B**	**MSB**
Nearest neighboor	-0.4561 (0.5562)	-0.3678 [-63.67] (0.2210)	0.7074 [122.45] (0.2846)	0.2381 [41.22] (0.4382)	0.5777 [-126.66] (0.3902)	0.0620 [-13.59] (0.3476)
Caliper	-0.0459 (0.3355)	-0.5649 [-229.63] (0.2208)	0.8594 [349.35] (0.3055)	-0.0485 [-19.72] (0.3275)	0.2460 [-535.95] (0.2132)	-0.3325 [724.40] (0.3613)
Kernel	-0.0668 (0.3339)	-0.5922 [-218.69] (0.2176)	0.8594 [317.36] (0.3055)	0.0036 [1.33] (0.3276)	0.2708 [-405.39] (0.2133)	-0.2973 [445.06] (0.3592)
Multiple imputa- tions	-0.3018 (0.8181)	-0.39238 [-84.94] (0.351653)	0.7074 [153.14] (0.2846)	0.14686 [31.79] (0.777766)	0.4619 [-153.06] (0.776121)	-0.0035 [1.16] (0.849612)
Local linear regression-adjusted	-0.2265 (0.3608)	0.2700 [212.10] (0.1418)	-0.1130 [-88.77] (0.2104)	-0.0295 [-23.17] (0.2913)	0.1273 [-56.20] (0.2840)	0.0125 [-5.52] (0.2143)

Notes : We give in parentheses the bootstrap standard values and in brackets for B_1, B_2 and B_3 the value in % of the bias with respect to the total bias B and for B and MSB, the value in % with respect to the mean impact M.