

Donzé, Laurent

Working Paper

Impact of the density support on the matching bias: A matched-pair analysis based on business survey data

KOF Working Papers, No. 98

Provided in Cooperation with:

KOF Swiss Economic Institute, ETH Zurich

Suggested Citation: Donzé, Laurent (2005) : Impact of the density support on the matching bias: A matched-pair analysis based on business survey data, KOF Working Papers, No. 98, ETH Zurich, KOF Swiss Economic Institute, Zurich,
<https://doi.org/10.3929/ethz-a-004957353>

This Version is available at:

<https://hdl.handle.net/10419/50902>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Arbeitspapiere/ Working Papers

Laurent Donzé

Impact of the Density Support on the
Matching Bias: A Matched-Pair Analysis
Based on Business Survey Data

Impact of the density support on the matching bias: a matched-pair analysis based on business survey data

Laurent Donzé

Department of Quantitative Economics

University of Fribourg

Av. de Beauregard 13

CH – 1700 Fribourg, Switzerland

Laurent.Donze@UniFr.ch

<http://www.unifr.ch/stat/>

Matching methods have been extensively used to evaluate economic policy. However, they are not without fault and, indeed, a selection bias may appear. In 1998, Heckman and al. (1998) have precisely characterize this bias by a decomposition in three parts which can be non parametrically estimated. These estimations depend among others on the support of the conditional density function of the covariates X used for the matching.

In a study of the impact of the policy of supporting the adoption of advanced manufacturing technologies by Swiss firms, we have measured the different components of the bias. We put into evidence the extreme sensibility of the results to the measure of support density.

1. The selection bias decomposition

Heckman and al. (1998) find the following bias decomposition.

Given $S_{1X} = \{X|f(X|D=1) > 0\}$, the support of X for $D = 1$, i.e. for the participation to a programme; $f(X|D = 1)$ is the conditional density of X given $D = 1$; $S_{0X} = \{X|f(X|D=0) > 0\}$, the support of X for $D = 0$, i.e. for the no participation to a programme; $S_X = S_{0X} \cap S_{1X}$, the intersection region of both supports. One can rewrite the traditional measure of the bias $B = E(Y_0|D = 1) - E(Y_0|D = 0)$, where Y_0 is the output variable in the case of no treatment, as the sum of three components:

$$(1) \quad B = B_1 + B_2 + B_3,$$

where

$$(2) \quad \begin{aligned} B_1 &= \int_{S_{1X} \setminus S_X} E(Y_0|X, D=1) dF(X|D=1) - \int_{S_{0X} \setminus S_X} E(Y_0|X, D=0) dF(X|D=0), \\ B_2 &= \int_{S_X} E(Y_0|X, D=0) [dF(X|D=1) - dF(X|D=0)], \\ B_3 &= P_X \bar{B}_{S_X}. \end{aligned}$$

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 ; $S_{1X} \setminus S_X$ is the support of X given $D = 1$ which is not in the intersection region S_X ; $S_{0X} \setminus S_X$ 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 $S_{0X} \setminus S_X$ and $S_{1X} \setminus S_X$ are not empty. In this case, one cannot find the counterpart of $E(Y_0|X, D=1)$ in the set $S_{0X} \setminus S_X$ or the counterpart of $E(Y_0|X, D=0)$ in the set $S_{1X} \setminus S_X$.

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.

2. 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 $E(Y_0|P(X), D = 0) = X\beta + E(U_0|P(X), D = 0)$, where $P(X)$ is the probability that the events " $D = 0$ " or " $D = 1$ " would have been append for all $X \in \chi_c$, a set of covariates. The selection bias can be written as

$$(3) \quad \begin{aligned} B(P(X)) &= E(Y_0|P(X), D = 1) - E(Y_0|P(X), D = 0) \\ &= E(U_0|P(X), D = 1) - E(U_0|P(X), D = 0). \end{aligned}$$

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 model¹

$$(4) \quad Y_{0i} = X_i\beta + D_iK_1(P_i) + (1 - D_i)K_0(P_i) + \varepsilon_i.$$

We form an adjusted version of (4) by subtracting from it its conditional expectation with respect to P_i and D_i . We obtain

$$(5) \quad Y_{0i} - E(Y_{0i}|P_i, D_i) = [X_i - E(X_i|P_i, D_i)]\beta + \varepsilon_i.$$

We then estimate β from equation (4) 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 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.² 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

¹A 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.

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

$$(6) \quad \min_{\gamma_{1d}, \gamma_{2d}} \sum_{i \in \{D=d\}} \left[c_i - \gamma_{1d} - \gamma_{2d}(\hat{P}_i - P_0) \right]^2 G \left(\frac{\hat{P}_i - P_0}{a_N} \right), \quad d \in \{0, 1\},$$

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;³ \hat{P}_i is the estimated value of P for unit i .⁴

3. 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 (2). Indeed, Heckman and al. (1998) propose to compute those in the following manner

$$(7) \quad \begin{aligned} \hat{B} &= \hat{E}(Y_0|D = 1) - \hat{E}(Y_0|D = 0) = \hat{B}_1 + \hat{B}_2 + \hat{B}_3 \\ \text{where} \\ \hat{B}_1 &= \frac{1}{N_1} \sum_{\substack{i \in \{D=1\} \\ P_i \in S_{1P} \setminus S_P}} Y_0(P_i) - \frac{1}{N_0} \sum_{\substack{i \in \{D=0\} \\ P_i \in S_{0P} \setminus S_P}} Y_0(P_i), \\ \hat{B}_2 &= \frac{1}{N_1} \sum_{\substack{i \in \{D=1\} \\ P_i \in S_P}} \hat{E}(Y_{0i}|P_i, D_i = 0) - \frac{1}{N_0} \sum_{\substack{i \in \{D=0\} \\ P_i \in S_P}} Y_0(P_i), \\ \hat{B}_3 &= \frac{1}{N_1} \sum_{\substack{i \in \{D=1\} \\ P_i \in S_P}} \left[Y_0(P_i) - \hat{E}(Y_{0i}|P_i, D_i = 0) \right], \end{aligned}$$

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 , $S_{1P} \setminus S_P$, $S_{0P} \setminus S_P$ are defined in similar manner as supports S_X , $S_{1X} \setminus S_X$, $S_{0X} \setminus S_X$ in (2).

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.

We will estimate $\hat{E}(Y_{0i}|P_i, D_i = 0)$ by a local linear regression model of Y_{0i} on P_i for the observations $D = 0$, i.e. one have to solve the problem

$$(8) \quad \min_{\gamma_1, \gamma_2} \sum_{i \leq N_0} \left[Y_{0i} - \gamma_1 - \gamma_2(\hat{P}_i - P_0) \right]^2 G \left(\frac{\hat{P}_i - P_0}{a_N} \right),$$

which has the same characteristics as (6).

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

³Heckman 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 a_N , the "bandwidth", is more delicate to determine. We follow the following rule: $a_N = 2.7768(\hat{H} - 1.34)N^{-1/5}$ where \hat{H} is the interquartile range of \hat{P}_i .

⁴One can remark in fact that the problem to solve is those of a weighted least squares regression, weights being given by the kernel.

$$(9) \quad MSB = \hat{B}_{S_p} = \frac{1}{N_1} \sum_{\substack{i \in \{D=1\} \\ P_i \in S_P}} (\hat{K}_{1i}(P_i) - \hat{K}_{0i}(P_i)).$$

4. Data description

In 1996, the KOF ETH Zurich investigated, in the framework of its business innovation survey in the industry sector, the adoption by the Swiss firms of advanced manufacturing technologies (AMT). This survey provides us the data of our study.

The final data base with which 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. 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.

5. Empirical results and sensitivity analysis

In Donzé (2002), besides the measure of the impact of supporting the adoption of advanced manufacturing technologies by Swiss firms, we estimate the selection bias of different matching methods. The different components of the bias may be important although generally not significant. Furthermore, our empirical results show that the definition, and then the estimation of the density support, may have relatively high influence on these biases.

Let us take the case of the Kernel matching. In order to measure the different components of the bias and the mean selection bias, we have to estimate the supports of the densities of the probabilities of participation. More precisely, we have to estimate the common support of both densities (participation vs not participation) and the complements of this common support. But, we encounter the problem that the densities we estimate are discrete. Consequently, it is not so easily, empirically, to combine these two densities in order to find the common support. We have resolved this problem in generating classes with same probability and compute the resulting densities. But the results depends on the number of bins chosen. In fact, the common support of the densities is depending on this number of bins and thus are the bias too.

The results can be summarize as follows. The Figures 1 to 4 show the densities estimated according to different number of bins (25, 50, 75 and 100); the Table 1 gives the number of observations in and out the common support; finally, the Table 2 resumes the decomposition of bias. It is not surprising that the different biases $B1$, $B2$ and $B3$ can vary considerably with the support though the overall bias B , i.e the traditional measure of the bias, stays at the same level. More interesting is the mean selection bias MSB which seems more or less sensitive according to the definition of the support. In conclusion, one can say that a careful attention to the definition of the support is necessary if we want control the bias of the matching method.

REFERENCES

1. **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.
2. **Donzé, L.** (2002): Matched-pair analysis based on business survey data to evaluate the policy of supporting the adoption of advanced manufacturing technologies by Swiss firms, KOF ETH Zurich, Arbeitspapiere / Working Papers, no 65, July 2002.
3. **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.

RÉSUMÉ

Les méthodes d'appariement ont été intensivement utilisées dans l'évaluation de politiques économiques. Cependant, elles ne sont pas sans défaut. En particulier, elles peuvent introduire un biais de sélection. En 1998, Heckman et al. (1998) ont de manière rigoureuse caractérisé ce biais en le décomposant en trois parties qui peuvent être estimées non paramétriquement. Ces estimations dépendent entre autres du support de la fonction de densité conditionnelle des covariables X utilisées dans la procédure d'appariement.

Dans une étude de l'impact de la politique suisse de soutien à la propagation des technologies de production assistée par ordinateur, nous avons calculé les différentes composantes du biais. Nous avons mis en évidence l'extrême sensibilité des calculs à la mesure du support de la densité.

Table 1: Number of observations in and out the common support

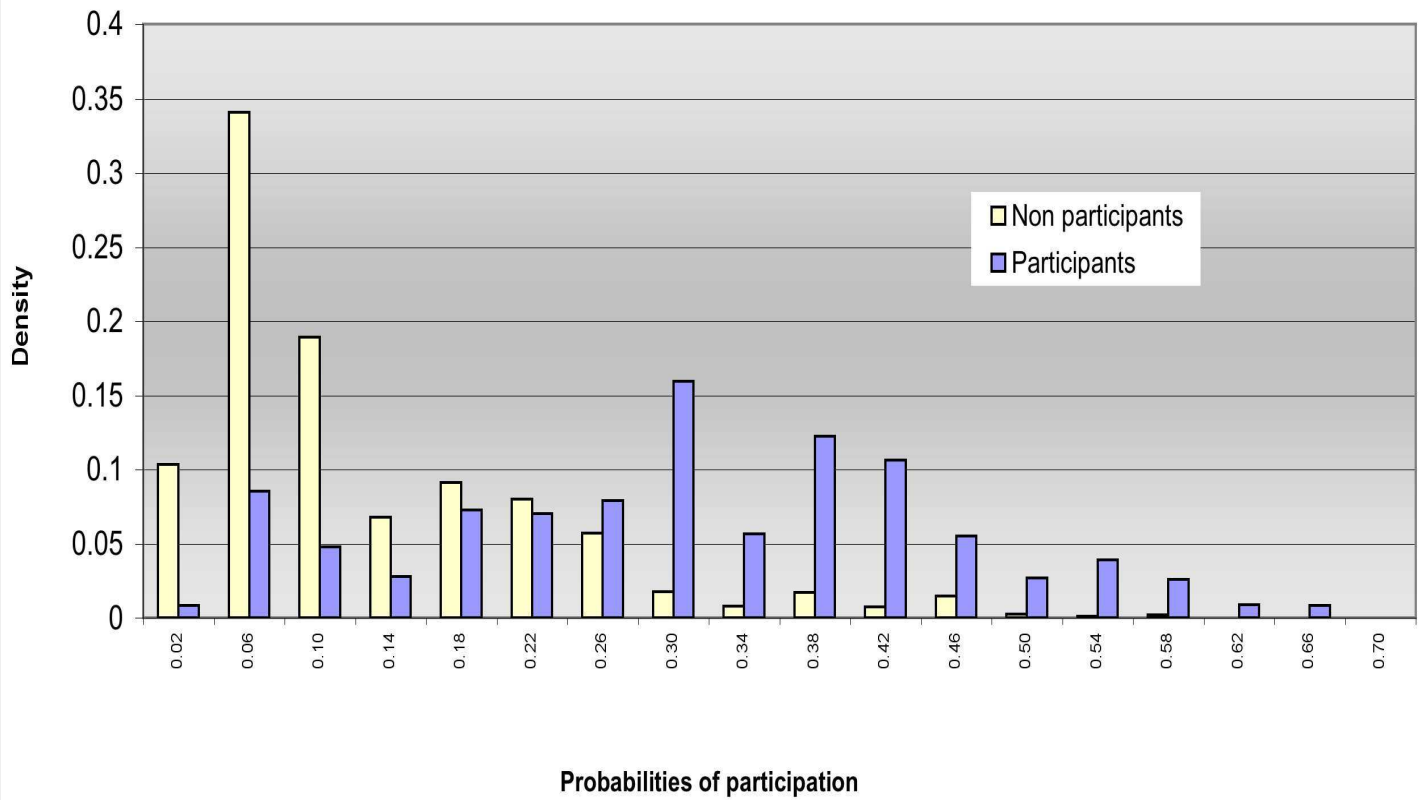
		Data before trim- ming		Common support estimated		Non common sup- port estimated	
N. of bins	CIMPOL	N	In %	N	In %	N	In %
25	0	367	79.96	359	80.67	0	0
	1	92	20.04	86	19.33	4	100
	Total	459	100.00	445	100.00	4	100.00
50	0	367	79.96	343	81.09	16	61.54
	1	92	20.04	80	18.91	10	38.46
	Total	459	100.00	423	100.00	26	100.00
75	0	367	79.96	339	81.49	20	60.61
	1	92	20.04	77	18.51	13	39.39
	Total	459	100.00	416	100.00	33	100.00
100	0	367	79.96	285	78.95	74	84.09
	1	92	20.04	76	21.05	14	15.91
	Total	459	100.00	361	100.00	88	100.00

Table 2: Impact and bias decomposition: Kernel matching method; different number of bins (Output variable : AMTDIF)

Number of bins	Impact M	B ₁	B ₂	B ₃	B	MSB
25	-0.0918 (0.2512)	0.1728 [71.73] (0.0773)	0.0688 [28.55] (0.1728)	-0.00068 [-0.28] (0.2147)	0.2409 [- 262.41] (0.1457)	0.0130 [- 14.16] (0.2016)
50	-0.0918 (0.2512)	0.2016 [83.68] (0.1067)	0.0116 [0.0481] (0.1760)	0.0276 [11.45] (0.1984)	0.2409 [- 262.41] (0.1504)	-0.0170 [18.51] (0.2209)
75	-0.0918 (0.2512)	0.3008 [124.86] (0.1268)	-0.0751 [- 31.17] (0.1733)	0.0153 [6.35] (0.1883)	0.2409 [- 262.41] (0.1504)	-0.01229 [13.38] (0.2315)
100	-0.0918 (0.2512)	-0.09917 [-41.16] (0.1593)	0.3080 [127.85] (0.2026)	0.0321 [13.32] (0.1870)	0.2409 [- 262.41] (0.1504)	-0.0304 [33.11] (0.2358)

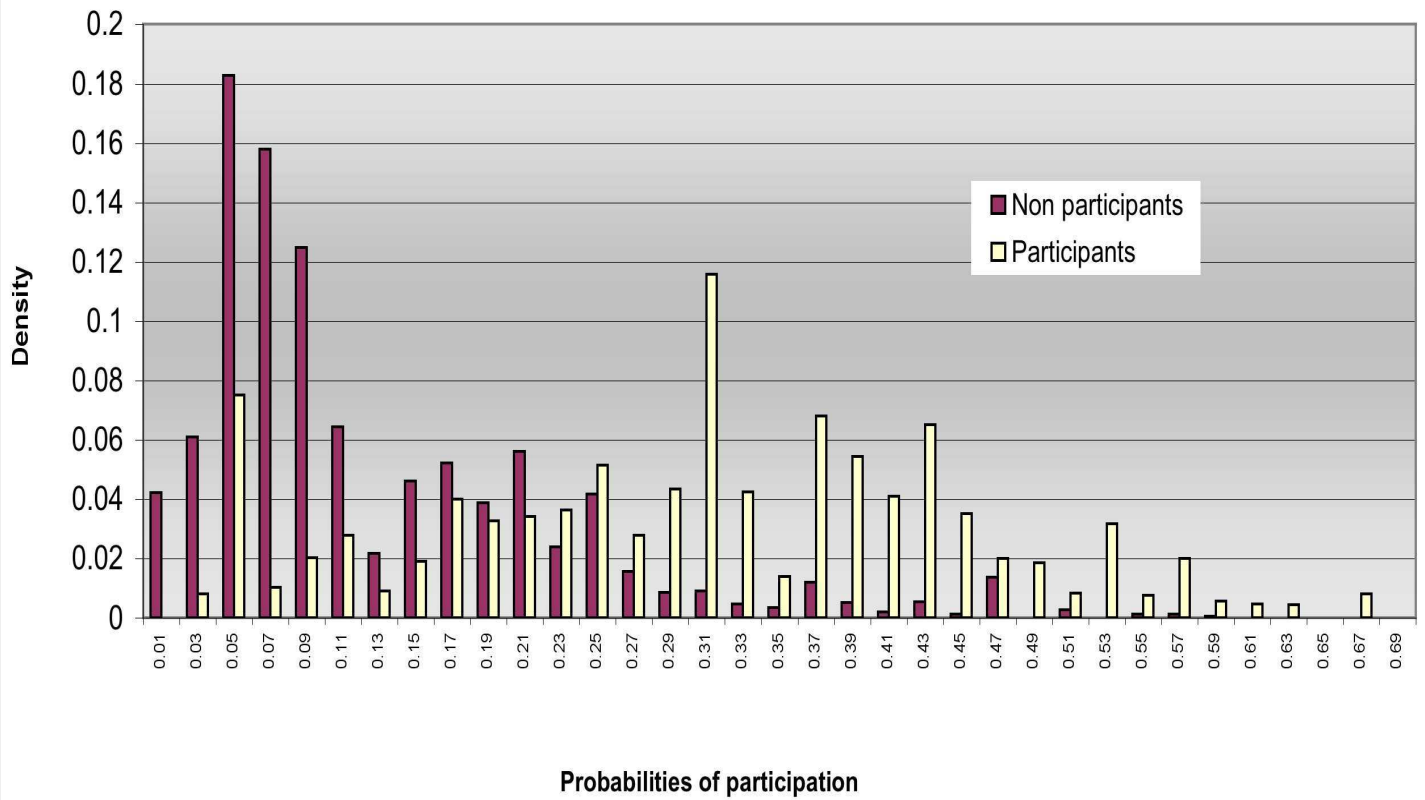
Notes : *AMTDIF* is the number of *AMT* components introduced in the period 1990–1996; we give in parentheses the bootstrap standard values and in brackets for B_1 , B_2 and B_3 the value in % of the bias with respect to the total bias B and for B and MSB , the value in % with respect to the mean impact M .

Figure 1: Estimated density of probabilities of participation



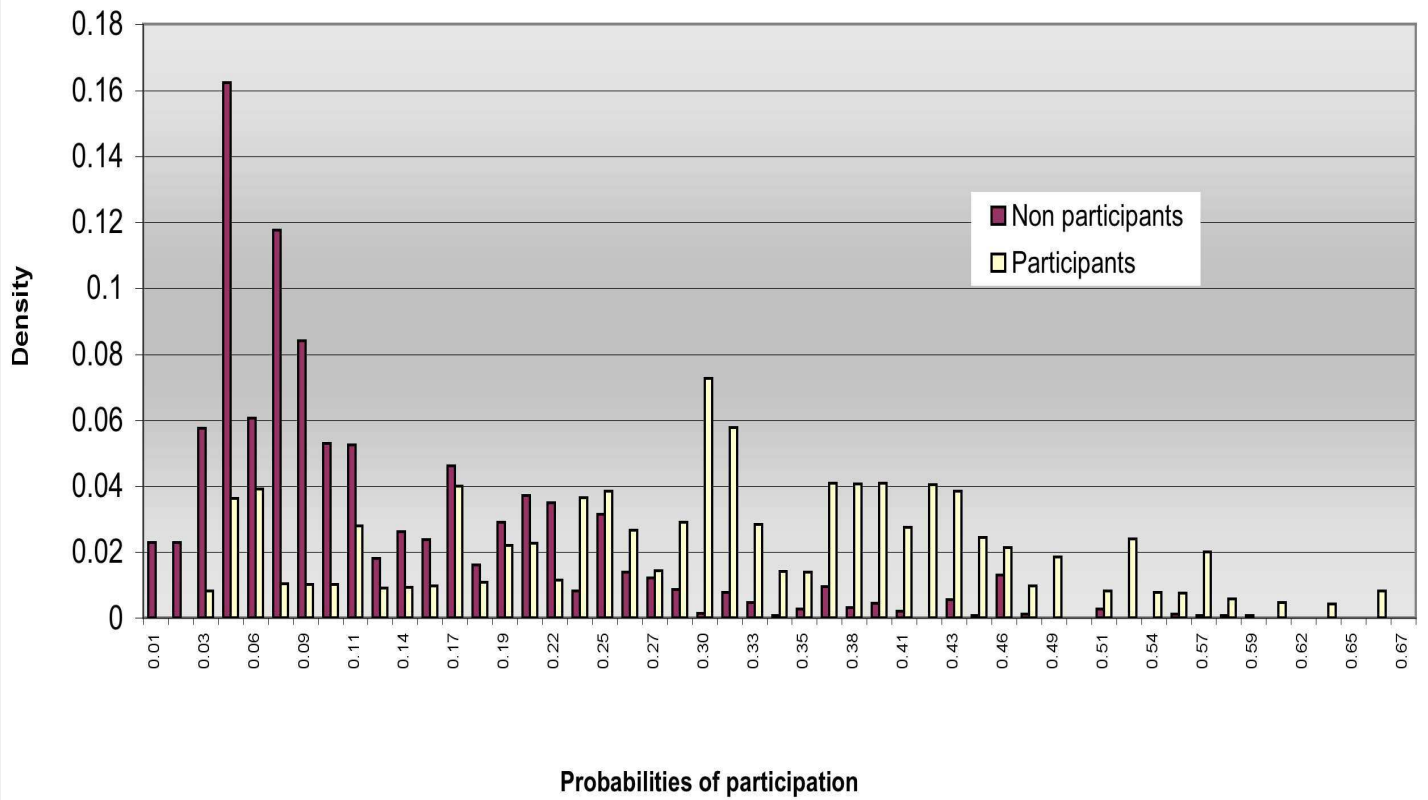
Note : Densities with number of bins equal to 25

Figure 2: Estimated density of probabilities of participation



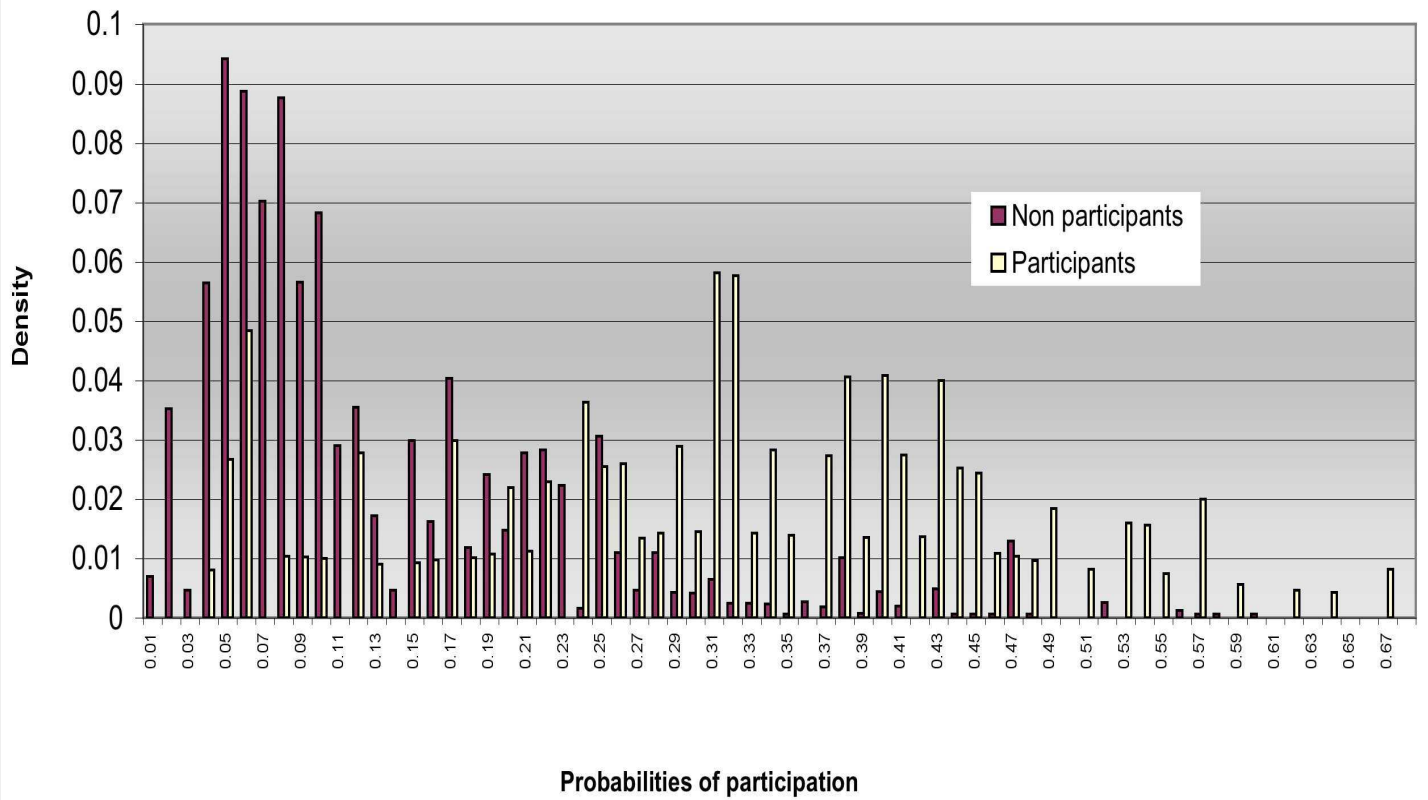
Note : Densities with number of bins equal to 50

Figure 3: Estimated density of probabilities of participation



Note : Densities with number of bins equal to 75

Figure 4: Estimated density of probabilities of participation



Note : Densities with number of bins equal to 100