

Donzé, Laurent

**Working Paper**

## Methodology to correct the unit non-response bias in the case of the KOF ETH Survey 2000 on organization and information technologies

KOF Working Papers, No. 67

**Provided in Cooperation with:**

KOF Swiss Economic Institute, ETH Zurich

*Suggested Citation:* Donzé, Laurent (2002) : Methodology to correct the unit non-response bias in the case of the KOF ETH Survey 2000 on organization and information technologies, KOF Working Papers, No. 67, ETH Zurich, KOF Swiss Economic Institute, Zurich

This Version is available at:

<https://hdl.handle.net/10419/50839>

**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

Methodology to Correct the Unit Non-  
response Bias in the Case of the  
KOF ETH Survey 2000 on Organization  
and Information Technologies



# Methodology to correct the unit non-response bias in the case of the KOF ETH Zurich's survey 2000 on Organization and Information Technologies

by

**Laurent Donzé**

Zurich, August 2002

## **Abstract**

In order to correct the bias due to unit non-response for the KOF ETH Zurich's business (mail) surveys, we usually use the results of a second (phone) survey by the non-respondents. Taking the case of the survey 2000 on 'Organization and Information Technologies' in the Swiss economy, we describe how to build the sample of this second survey and how to use the collected data. Actually, we show how to generate new correcting weights to correct the non-response bias.

## **Keywords**

Unit non-response bias, weighting, calibration, Information Technologies

## **JEL codes**

C42, C81, L2

Prepared for The International Conference on Improving Surveys (ICIS 2002), Copenhagen, 25<sup>th</sup> – 28<sup>th</sup> August 2002. Downloadable at: <http://www.dplanet.ch/users/ldonze>.

## 1. Introduction

The aim of the present paper is to introduce shortly the methodology we used to correct the unit non-response of the KOF ETH Zurich's survey 2000 on "Organization and Innovation Technologies". Usually, this type of non-response is corrected by a weighting factor. As our survey is complex in many aspects we have to develop rather several weighting factors. Hereafter, the principal steps of the construction of these weighting factors are explained. In particular, we show how to use a second survey, addressed specially to a sample of non-respondents, in order to provide the necessary information to correct the bias. In a first section, we present the characteristic of the survey. The topic of the section 2 is devoted to the conception of the second survey. Finally, we describe the different weighting factors.

## 2. Characteristic of the survey

The "Organization and Innovation Technologies" survey was based on the KOF ETH Zurich's business panel. This latter, built from the Swiss business census and regularly updated, included during the time of survey about 7'000 firms of the manufacturing, construction and services sectors. Each sector is stratified by economic activities and three firm sizes. The units was drawn to provide a "representative" sample of the Swiss economy.<sup>1</sup>

The complexity of the survey results in two facts. First, the design plan and, secondly, the length and non-homogeneity of the questionnaire. In fact, the questionnaire used for this survey is divided in 6 parts: 1) Data about the firm; 2) Organisation of the firm; 3) Organisation and cooperation forms on the work place; 4) Working time, wage, continuing education; 5) Information and communication technologies; 6) Use of Advanced Manufacturing Technologies (AMT).<sup>2</sup>

The survey was made by mail and after an intensive call-back by phone, 2'648 enterprises answered. The response rate is about 40% which is rather good for this kind of survey. As the Table 1 shows, the response rate varies slightly from stratum to stratum. A non-response bias can obviously be suspected. The second survey by the non-respondents has to elucidate this question.

## 3. Survey addressed to the non-respondents

This second survey was made by phone. The first problem that we had to face was to select a core of questions. These had to be easy to answer and precise enough to furnish the necessary information in any case. Table 4 shows the questionnaires. A major aspect was that we had to manage three sampling frames. In fact, the first three questions were addressed to all enterprises; questions 4 and 5 only to the enterprises of the industry and construction sectors; questions 6, 7 and 8 only to the enterprises greater than or equal to 20 employees. In this context, how can we drawn a more or less "optimal" sample of the non-respondents?

The first step was to model the response probability in order to construct by the propensity scores method homogenous "adjustment cells".<sup>3</sup> Table 3 gives the parameters of the model we estimated. Four adjustment cells were generated and the respondents and non-respondents of the

---

<sup>1</sup> Cf. **Donzé** (1998).

<sup>2</sup> The questionnaire can be downloaded in French, German and Italian from <http://kof.ethz.ch/>.

<sup>3</sup> Cf. as instance **Rosenbaum and Rubin** (1983), **Donzé** (2001b).

main survey were classified in these cells. According to this new structure, and for a specific variable, the optimal sub-sample size were estimated. In order to do so we proceed as follows.<sup>4</sup>

We suppose an “optimal allocation”. Let  $V$  the desired variance in the estimate of a proportion  $P$  for the whole population – in this case the set of non-respondents -;  $p_h$  the estimated proportion in the adjustment cell  $h$ ,  $h = 1, \dots, 4$ ;  $N_h$  the number of units of the initial sample in the adjustment cell  $h$ . We have  $N = N_1 + \dots + N_4$ ,  $W_h = N_h/N$ ,  $q_h = 1 - p_h$ , and

$$n_0 = \frac{1}{V} \left[ \sum_{h=1}^4 W_h \sqrt{p_h q_h} \right]^2 \quad (1.1)$$

$$n = n_0 \left[ 1 + \frac{1}{NV} \sum_{h=1}^4 W_h p_h q_h \right]^{-1} \quad (1.2)$$

$$n_h = n \frac{N_h \sqrt{p_h q_h}}{\sum_{h=1}^4 N_h \sqrt{p_h q_h}}. \quad (1.3)$$

When we have estimated  $n_h$ , the size  $\bar{r}$  of the non-respondents sub-sample can easily be estimated. Indeed, we have  $\bar{r}_h = n_h - r_h$ .  $\bar{r} = \sum_{h=1}^4 \bar{r}_h$ . The estimated proportions  $p_h$  and  $q_h$  are given by the first results of the initial survey. These are for instance the proportion of firms answering “yes” or “no” to a specific question. We show in Table 5a the estimated optimal size of the non-respondents sub-samples for some questions. We decided to take as reference the variables F51D\_3, F61A\_3 and F34B, which resume perfectly our core questions.<sup>5</sup> On this base, we draw per each adjustment cell at random and without replacement three sub-samples of non-respondent firms considering firstly all the enterprises, secondly, the enterprises of the construction and service sectors and thirdly, the enterprises with 20 or more employees. Combining the three sub-samples, deleting the multiple occurrences and reducing it finally to 650 observations for cost and time questions (16% of the non-respondents), we got the final global sub-sample. Table 5b shows the distribution of this sample according to the adjustment cells. From the 650 enterprises, 392 are issued from the industry and construction sectors and 564 have 20 or more employees. For this second survey, we get as desired and fortunately a high response rate around 94%. The response rates pro adjustment cell are given in Tables 6.

#### 4. Bias correction through weighting factors

The estimated proportions of our core variables, on the one hand with the data of the initial survey and, on the other hand, with the data of the non-respondents survey, show significantly and great differences. These results are shown in Table 7. A bias due to the non-response is therefore certainly present in the first survey and it becomes really important to adjust its data. In order to do this, we have generated appropriated weighting factors. Actually, we have constructed 4 new factors on the base of the variables F45N, F51DE, F61AD and F31,. The procedure is the following.<sup>6</sup> First, compute the marginal totals, i.e. the estimated number of enterprises answering “yes” and those. “no” to a specific question per economic activities or per enterprise sizes. Calibrate then the initial weight with these marginal totals.<sup>7</sup> Eventually, adjust the resulting weight by another factor.

<sup>4</sup> Cf. as instance Cochran (1977), pp. 107-111, Donzé (1999).

<sup>5</sup> Cf. Table 2 for the description of the variables.

<sup>6</sup> Cf. as instance Flottes (1997), Donzé (2001a).

<sup>7</sup> Cf. as instance Deville and al. (1993).

Let  $v$  a variable (relative to a question), e.g. F45N;  $t_{1s}^{(v)}$  and  $t_{1s}^{(\bar{v})}$  the number of firms by stratum  $s$  answering “yes” and those “no” to the question  $v$  in the first survey;  $\hat{p}_{2s}^{(v)}$  the estimated proportion of firms answering “yes” to the question  $v$  by stratum  $s$  in the second survey;  $r_s$  the number of non-respondent firms in the first survey;  $\omega_s = N_s/n_s$  the sampling fraction where  $N_s$  is the number of firms by stratum in the population (business census) and  $n_s$  is the number of firms by stratum in the panel. The estimated totals by stratum can be calculated as follow:

$$\hat{t}_s^{(v)} = \omega_s \left[ t_{1s}^{(v)} + \hat{p}_{2s}^{(v)} r_s \right], \quad (1.4)$$

$$\hat{t}_s^{(\bar{v})} = \omega_s \left[ t_{1s}^{(\bar{v})} + (1 - \hat{p}_{2s}^{(v)}) r_s \right]. \quad (1.5)$$

The marginal totals are easily computed by summing the right strata.

We use then the calibration SAS procedure CALMAR, developed by Deville and Sautory from INSEE,<sup>8</sup> to calculate with the information given by the marginal totals and the initial weight  $\omega_s$ , the new weight  $\omega_s^{(c)}$ . This latter can be still adjusted by, as instance, the relative fraction (measured in terms of number of employees, or sum of turnovers, etc.) of the stratum in the whole economic sector.

Tables 8 and 9 present a sensitivity analysis of the weighting factors on the variable F51D (Introduction of e-mail) and F61AD (Introduction of AMT). According to the sector, the economic activity or the size of the firm, the influence of the weighting factor can be important.

## 5. Conclusion

It is well known that a non-response bias can be important when the non-response rate is high. Our experience with the KOF ETH Zurich’s survey 2000 on ‘Organization and Information Technologies’ shows, as it appears in a second survey by non-respondents, that the non-response has effectively a great influence on the estimation of certain variables. Nevertheless, one can easily develop weighting factors for specific variables in order to correct this bias.

---

<sup>8</sup> Cf. Sautory (1993)

**Table 1: Response rate by sector and firm size for the KOF ETH Zurich' survey on Organization and Information Technologies (in %)**

	Firm size			
	Small	Medium	Large	Total
<b>Manufacturing</b>	42.54	36.81	37.05	39.83
<b>Construction</b>	37.80	46.70	43.42	42.34
<b>Services</b>	41.58	35.69	26.77	38.12
<b>Total</b>	41.76	37.23	34.57	39.32

**Table 2: Description of the variables**

Variable	Description
REP	Dummy variable for the response (1 answer; 0 no answer)
IND_2	Dummy variable for Manufacturing (Paper, Publishing, Petroleum and Chemicals, Rubber and Plastic products. Other non-metallic mineral products)
DL_1	Dummy variable for Services (Wholesale and retail trade, Personal services)
GR_M	Dummy variable for firm size (Medium)
GR_G	Dummy variable for firm size (Large)
SP_F	Dummy variable for language (French)
AC1	Adjusting cell (according to the response probability)
F45	Part of employees who participate to internal or external training courses in 1999
F45N	If F45>0 then F45N=1 ; else F45N=0
F51D_1	Introduction of e-mail before 1995
F51D_2	Introduction of e-mail during 1995-1997
F51D_3	Introduction of e-mail during 1998-2000
F51D	If F51D_1=1 or F51D_2=1 or F51D_3=1 then F51D=1; else F51D=0
F51E_1	Introduction of internet before 1995
F51E_2	Introduction of internet during 1995-1997
F51E_3	Introduction of internet during 1998-2000
F51E	If F51E_1=1 or F51E_2=1 or F51E_3=1 then F51E=1; else F51E=0
F51DE	If F51D=1 or F51E=1 then F51DE=1; else F51DE=0
F61A_1	Introduction of CAD (CAD/CAE, CAD/CAM, etc.) before 1995
F61A_2	Introduction of CAD (CAD/CAE, CAD/CAM, etc.) during 1995-1997
F61A_3	Introduction of CAD (CAD/CAE, CAD/CAM, etc.) during 1998-2000
F61A	If F61A_1=1 or F61A_2=1 or F61A_3=1 then F61A=1; else F61A=0
F61C_1	Introduction of CNC/DNC before 1995
F61C_2	Introduction of CNC/DNC during 1995-1997
F61C_3	Introduction of CNC/DNC during 1998-2000
F61C	If F61C_1=1 or F61C_2=1 or F61C_3=1 then F61C=1; else F61C=0



**Table 2: continued**

Variable	Description
F61D_1	Introduction of other AMT such FMC/FMS (Flexible Manufacturing Cells or Systems), robots, lasers, etc., before 1995
F61D_2	Introduction of other AMT such FMC/FMS (Flexible Manufacturing Cells or Systems), robots, lasers, etc., during 1995-1997
F61D_3	Introduction of other AMT such FMC/FMS (Flexible Manufacturing Cells or Systems), robots, lasers, etc., during 1997-1998
F61D	If F61D_1=1 or F61D_2=1 or F61D_3=1 then F61D=1; else F61D=0
F61AD	If F61A=1 or F61D=1 then F61AD=1; else F61AD=0
F34A	Change in the distribution of competence at the work place since 1995
F34B	Change in the distribution of competence at the work place since 1995 in the direction of the collaborator
F34C	Change in the distribution of competence at the work place since 1995 in the direction of the superior
F31A_1	Permanent work teams
F32A_1	Rotation work places program
F31	If F31A_1=1 or F32A_1 then F31=1; else F31=0

**Table 3: Logit Modelling of Response Probability (Dependent Variable: REP)**

Model Variables	Estimated Parameters	Standard Values
Constant	-0.2128**	0.0404
IND_2	-0.1705*	0.0807
DL_1	-0.2598**	0.0608
GR_M	-0.1800**	0.0540
GR_G	-0.2987**	0.0854
SP_F	-0.2262**	0.0640
N Obs.	6735.0	
-2 Log L	64.055**	

- Notes: 1) "N Obs" is the number of observations; "-2 Log L" is the likelihood ratio statistic to test the global dependency.  
2) "\*\*\*" significant at 1%; "\*" significant at 5%.

**Table 4: Questionnaire of the non-respondent survey**

Question	Variable	Frame (non-respondent enterprises)
A. Qualification of collaborators 1. Have the collaborators of your enterprise participated to internal or external training course in 1999?	F45N	All enterprises
B. Introduction of Information and Communication technologies 2. Has your enterprise introduced e-mail? 3. Has your enterprise introduced internet?	F51D F51E F51DE	All enterprises
C. Use of Advanced Manufacturing Technologies (AMT) 4. In the domain of conception and design, do your enterprise use AMT such CAD, CAE or simulation? 5. Concerning manufacturing: a) Do your enterprise use CNC? b) Are other AMT used such as FMC/FMS (Flexible Manufacturing Cells or Systems), robots, lasers, etc.	F61A F61C F61D F61AD	Industry and construction sectors
D. Work organization 6. Distribution of competence a) Has the distribution of competence at the work place changed since 1995? b) If yes: in direction of the superiors? c) If yes: in direction of the collaborator? 7. Are there in your enterprise permanent work teams (project groups, quality circle, work groups partially free, etc.)? 8. Is there in your enterprise a rotation work places program, i.e. a systematic and planed of tasks?	F34A F34B F34C F31A_1 F32A_1 F31	Enterprises greater than or equal to 20 (full time equivalent employees)

**Table 5a: Optimal size of the sample of the non-respondents according to a specific question of the ICT 2000 survey for a specified global variance level (V is specified for a proportion P given in %)**

AC1	All enterprises			Industry and construction sectors			Enterprises greater than or equal to 20 (full time equivalent employees)				
	F45	F51D_3	F51E_3	F61A_3	F61C_3	F61D_3	F34A	F34B	F34C	F31A_1	F32A_1
	V=0.2	V=0.4	V=0.4	V=0.2	V=0.1	V=0.08	V=0.7	V=0.6	V=0.07	V=0.5	V=0.5
1	238	227	199	61	95	136	471	220	66	257	125
2	88	113	103	78	37	66	116	66	71	69	46
3	140	115	98	122	158	172	139	109	123	142	39
4	112	52	67	28	65	0	0	57	26	101	50
Total		507		289				452			

**Table 5b: Sample size of the non-respondents per adjustment cell (global size reduced to 650 observations (=16% of the non-respondents))**

Adjustment cell	All enterprises	Industry and construction sectors	Enterprises greater than or equal to 20 (full time equivalent employees)
1	256	88	246
2	134	90	105
3	186	151	166
4	74	63	47
Total	650	392	564

**Table 6: Response rates of the non-respondents survey (%)**

Adjustment cell	Response rate
1	92.19
2	94.78
3	93.55
4	98.65
Total	93.85

**Table 7: Comparison between the initial survey and the non-respondents survey**

	<b>F45N</b>	<b>F51D</b>	<b>F51E</b>	<b>F51DE</b>	<b>F61A</b>	<b>F61C</b>	<b>F61D</b>
<b>Initial survey</b>	89.60	80.78	73.34	82.43	56.38	38.55	18.36
<b>Non-respondents survey</b>	81.69	90.66	89.09	91.67	69.69	61.27	28.65
	<b>F61AD</b>	<b>F34A</b>	<b>F34B</b>	<b>F34C</b>	<b>F31A_1</b>	<b>F32A_1</b>	<b>F31</b>
<b>Initial survey</b>	60.18	48.58	48.60	2.82	57.69	16.60	61.83
<b>Non-respondents survey</b>	73.79	67.32	57.12	30.85	64.28	18.95	68.46

- Notes: 1) The table gives the part of “yes” for the different variables;  
 2) The variables are weighted according to the design plan and the non-response.

**Table 8: The influence of weighting factors on the estimation of the variable F51D  
(Introduction of e-mail)**

Economic sector	Weighting factor	Small firms		Medium firms		Gross firms		Total	
		E-mail “yes”	E-mail “no”	E-mail “yes”	E-mail “no”	E-mail “yes”	E-mail “no”	E-mail “yes”	E-mail “no”
Industry	None	86.71	13.29	97.51	2.49	97.44	2.56	91.52	8.48
	GEW1	84.19	15.81	96.82	3.18	97.11	2.89	86.78	13.22
	GEW2_F51DE	87.76	12.24	92.69	7.31	96.63	3.37	88.79	11.21
	GEW3_B_F51DE	87.85	12.15	91.08	8.92	94.97	5.03	88.53	11.47
Construction	None	64.58	35.42	82.08	17.92	100.00	0.00	80.22	19.78
	GEW1	64.68	35.32	82.24	17.76	100.00	0.00	69.02	30.98
	GEW2_F51DE	79.11	20.89	86.02	13.98	100.00	0.00	80.54	19.46
	GEW3_B_F51DE	79.09	20.91	86.02	13.98	100.00	0.00	81.07	18.93
Services	None	85.32	14.68	95.14	4.86	95.89	4.11	89.23	10.77
	GEW1	81.00	19.00	92.00	8.00	95.48	4.52	82.12	17.88
	GEW2_F51DE	86.42	13.58	92.25	7.75	92.55	7.45	86.98	13.02
	GEW3_B_F51DE	83.60	16.40	81.94	18.06	81.32	18.68	83.33	16.67
Total	None	84.62	15.38	94.77	5.23	97.66	2.34	89.40	10.60
	GEW1	79.03	20.97	90.62	9.38	97.29	2.71	80.78	19.22
	GEW2_F51DE	85.37	14.63	90.81	9.19	96.46	3.54	86.14	13.86
	GEW3_B_F51DE	85.33	14.67	85.81	14.19	90.51	9.49	85.51	14.49

Notes: 1) GEW1 takes into account the design plan and the non-response; GEW2\_F51DE is the calibrated weight according to the variable F51 and the design plan as initial weight; GEW3\_B\_F51DE is GEW2\_F51DE adjusted by the fraction of employees per stratum.

**Table 9: The influence of weighting factors on the estimation of the variable F61  
(Introduction of AMT)**

Economic sector	Weighting factor	Small firms		Medium firms		Gross firms		Total	
		AMT “yes”	AMT “no”	AMT “yes”	AMT “no”	AMT “yes”	AMT “no”	AMT “yes”	AMT “no”
Industry	None	75.78	24.22	92.14	7.86	97.27	2.73	84.27	15.73
	GEW1	74.27	25.73	91.61	8.39	97.10	2.90	78.44	21.56
	GEW2_F61AD	77.97	22.03	90.21	9.79	95.10	4.90	80.42	19.58
	GEW3_B_F61AD	75.39	24.61	87.56	12.44	91.20	8.80	77.53	22.47

Notes: 1) GEW1 takes into account the design plan and the non-response; GEW2\_F61AD is the calibrated weight according to the variable F61AD and the design plan as initial weight; GEW3\_B\_F61AD is GEW2\_F61AD adjusted by the fraction of employees per stratum.

## References

- Cochran, W. G.** (1977): "*Sampling Techniques* ", Wiley series in probability and mathematical statistics, John Wiley & Sons, 3, New York Chichester Brisbane Toronto Singapore, 428.
- Deville, J. C.; Särndal, C.-E. and Sautory, O.** (1993): "Generalized Raking Procedures in Survey Sampling", *Journal of the American Statistical Association*, 88(423), Theory and Methods, 1013-1020.
- Donzé, L.** (1998): "*Développement et entretien du "Panel d'entreprises" du KOF/ETHZ. Une étude méthodologique. Programme prioritaire "Demain la Suisse" "*, *Projet 5004-47794*, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Septembre 1998, 18.
- Donzé, L.** (1999): "*Enquête auprès des non-répondants de l'enquête KOF/ETHZ sur l'internationalisation de l'économie suisse. Une étude méthodologique.* " Rapport intermédiaire à l'attention du Fonds national suisse de la recherche scientifique. Programme prioritaire "Demain la Suisse", *Projet 5004-47794*, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Janvier 1999, 8.
- Donzé, L.** (2001a): "*Echantillon et non-réponse. Note sur la construction des facteurs de pondération de l'enquête 1999 sur l'innovation* ", Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Avril 2001, 4.
- Donzé, L.** (2001b): "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), 301-317.
- Flottes, A.** (1997): "*Proposal of CAMIRE to EUROSTAT concerning the non-response analysis and the correction of weights for the Second Community Innovation Survey* ", 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.
- Sautory, O.** (1993): "*La macro CALMAR. Redressement d'un échantillon par calage sur marges..* " *Série des documents de travail de la Direction des Statistiques Démographiques et Sociales, INSEE*, Paris, 55.