

Determinants of Consent in the German SOEP Establishment Survey 2012

Paper prepared for the 2013 European Establishment Statistics Workshop, Nuremberg, Germany

*M. Weinhardt**, *A. Meyermann***, *S. Liebig***, *J. Schupp**

* DIW Berlin, Free University Berlin; ** Bielefeld University

1 Introduction

Record linkage to administrative data is more and more common in present day survey research and offers several advantages (Calderwood & Lessof, 2009): it enriches the information provided by the informant, it allows for the cross-checking of the survey answers in order to control for measurement error and, in an extreme form, might even render the asking of whole questionnaire sections unnecessary, thereby decreasing the response burden to respondents.

In order to be able to link survey data to data collected by public authorities, in most countries the informed consent is required of the individual for which the linkage is envisaged (Schnell, 2012). In the interest of keeping the sample size up for the linked sample, consent from organizations to record linkage should be as high as possible. Low consent rates will result in reduced sample size of the linked sample and decreased statistical power. Furthermore, bias will be introduced into the estimates if the factors which influence consent are at the same correlated to the variables which researcher would like to investigate using the linked data. Hence, the investigation of the refusal to consent is necessary for two reasons: First, to identify causes of non-consent in order to address these in future surveys, trying to achieve higher consent rates. Second, to see whether consent is systematically related to variables which are likely to be of interest for the researchers using the data, resulting in bias of their estimates.

2 Data

The data for this study stem from a project to produce a linked employer employee (LEE) data set for the Socio-economic Panel (SOEP), a longitudinal study of German households representative of the German population with about 20,000 individuals and 10,000 households (Wagner *et al.*, 2007)¹. The

¹ The project is conducted in cooperation between the Socio-economic Panel Study, DIW Berlin and Bielefeld University. The project runs from January 2012 until December 2013;

survey has been implemented by asking all employees in the SOEP to provide local contact information for the employer for which they had worked in 2011. This contact information formed the basis for a separate standardized employer survey conducted from August 2012 to March 2013. This employer information can then be linked to the individual and household data from the SOEP study.

There is very high heterogeneity among the employers in the sample, as all lines of business and sectors are surveyed. The questionnaire was administered face-to-face and comprised 61 questions, 160 items, with drop-off possibility if necessary; mean duration of interviews was 42 minutes (median = 40 minutes). The adjusted response rate at the establishment level is 30.2%, resulting in 1708 interviews. Linking survey data to an establishment data is possible for 1817 individuals (109 establishments with more than one SOEP-employee).

In order to enrich the data of the survey with administrative records, linkage to the Establishment History Panel (BHP, Gruhl *et al.*, 2012) of the German Institute for Employment Research (IAB) is planned for 2013. The BHP consists of data on employees subject to social insurance contributions, such as their education and income, that employers report to the federal employment agency. Of all establishments which participated, for 1667 (97.6 %) an answer to the consent question was recorded and 587 establishments (35.2 %) gave their consent to record linkage.²

An unique feature of this study is the analysis of the employer survey data quality through the measurement of meta- and paradata over the course of the data collection process. Meta- and paradata stem from a detailed contact- and interviewer protocol and from a short additional questionnaire for interviewers. All this data allow for a detailed look at the determinants of consent in establishment surveys, a field of study that has been neglected so far.

3 Quantitative Analysis of the Response Process

We employ multivariate logistic regression models to investigate the effect of different sets of variables on the establishments likelihood to respond. The analysis proceeded as follows. The variables of different levels are entered in a stepwise fashion as four different sets of covariates. The pseudo R² of the different models give some indication as to which level of characteristics has the strongest impact on consent.³The results of the logistic regression models are presented in the table in the appendix.

² Due to the sample design, bigger establishment have a greater chance of being selected as their share is much bigger in the sample than in the actual population. To account for this fact, the inverse of the establishment size is used as the design weight. The weighted percentages of the SOEP-LEE sample come very close to the actual population percentages of establishment size, industry and regional distribution, which shows the validity of the sampling procedure when unequal selection probabilities are taken into account.

³ It has to be kept in mind however, that the pseudo R² depends on the number and nature of the included variables. The pseudo-R² is a standardised value of the reduction in the likelihood resulting from the inclusion of the variables in the model when compared to the baseline model without any covariates. It can be interpreted as the increase in model fit the added variables provide.

3.1 Partial Models

Looking at the results of the inclusion of the first set of covariates, those relating to the establishment level, it is striking that the value of the pseudo-R² is relatively low, even though a large number of establishment variables is included. This is an indication that the level of the establishment as such does not play a prominent role in explaining consent in establishment surveys.

The individual establishment variables do not show much impact either. There are only two variables which show a significant effect. When an establishment is autonomous with regard to hiring their own staff (an indicator for high authority) it is also much more likely to give consent (OR = 1.571; $p < .05$). When an establishment emphasises the importance of keeping their staff informed is also more likely to give consent (OR = 0.728; $p < .001$)⁴. This indicates an impact of “soft” factors such as general organisational policy.

Of the variables relating to the target person within the organisation which answered the interview, there is a couple of variables which show a significant effect. First, this is the age of the respondent where respondent aged 50 and older have a much higher probability of giving consent than their younger counterparts. Second, the regression results suggest that position within the company also play a role: If the person holds a position within controlling she is 1.4 times more likely to give consent compared to those in other positions. Third, the response persons motivation seems to play a pivotal role in providing consent. Asked whether they would like to receive a report on the findings of the survey, respondents who answered “yes” were 2.7 times more likely to agree to the consent question than those who were not. There is two further variables which are just not significant (p-value between .1 and .05) which are the gender of the respondent (OR = 0.734) and position in HR (OR = 1.421).

At the level of the interview situation, there are two variables which have a significant impact on the consent rate. The first one is the variable whether the interviewer was present during the whole interview or whether the questionnaire was left (at least for some questions) with the respondent. If the questionnaire was filled in with the interviewer present only, consent was twice as likely compared to if not (OR = 2.04). Also, the level of difficulty to identify a suitable response person in the first place is significantly related to consent: if there was no great the difficulty, the consent rate was higher (OR = 1.149).

Finally, there are two interviewer characteristics which significantly affect the likelihood of consent, the first one being an interviewer’s education Here it shows that interviewers with a higher secondary school degree have a much higher chance of obtaining consent (OR = 3.273) than interviewers with a lower secondary degree. The education of interviewers does not seem to be related in a linear manner with consent, however: While a interviewers with an university degree also have a much higher likelihood (OR = 2.110), this is lower when compared to those with higher secondary degree .

A second interview variable which is significant is the mean level of item non-response per interviewer (transformed to the natural logarithm). Here it

⁴ The item is coded reversely

shows that interviewer who have higher mean item non-response on average also have lower levels of consent. This suggest that consent is related to general interviewer performance also in other areas and it suggests that the giving of consent relies on similar mechanisms than providing answers to other items as well.

3.1.1 Full Model

In the full model, in which the effects of all variables are estimated simultaneously, the effect sizes and significance levels of the variables discussed so far change only slightly. At the establishment level, the odds ratio for the company policy of keeping employees informed increases slightly and is still significant at the 1 % level. The effect of being autonomous in hiring staff however decreases slightly and the p-value of significance is now above .05 (but still below .1).

At the response person level, while the effect of age and topic interest are still similar in amount and the same in the level in significance, there is more change to the estimated effects. Two variables indicating the respondent's position in the company are affected: the previously significant effect of being involved in controlling and bookkeeping ceases to be significant, while belonging to the overall management of the establishment now is significant at the 5 % level. The odds ratio decreases from .791 to .610, indicating that respondents in managerial positions refuse more often than those who are in other positions.

From the variables describing the interview situation, the effect of the difficulty to identify the response person is virtually unchanged and still significant at the 5 % level, the influence of the interviewer's presence ceases to be so. While the odds ration is still 1.83, it is not significant any more after the inclusion of other covariates.

At the interviewer level, the same two variables are significant: education and mean item non-response. The effect size increases slightly for higher secondary education and mean item non-response, while it decreases slightly for university degree, through which the U-shape of the relation between education and consent is pronounced further.

To summarize, overall, there are five variables whose effect is robust across model specifications: the establishment policy of keeping employees informed, respondents age, respondent's topic interest, the difficulty of identifying a respondent within the organisation, interviewer's education and their mean level of item non-response. Two further variables are significant in the final model: respondent's gender and their managerial position.

4 Discussion

Overall, only a few variables seem to be related to consent. Most importantly, no influence of structural establishment characteristics was found during the analysis. This is good news for researchers wishing to analyse the linked dataset as bias is less likely if structural variables are of interest to research question. However, evidence was found that some „soft“ company characteristics have an

impact on the likelihood to consent, as the variable on an establishment's policy to keep employees informed has a significant and strong affect on the likelihood to consent.

The analysis suggests that there are a few robust influences on the level of the response person. First of all, females seem much less likely to provide consent. In the same vein, younger cohorts are much less likely to consent. In addition, there is a strong effect for respondents who show an interest in the topic of the survey, indicated by them requesting a report on the findings which was promised as incentive for establishments. The consent rate for response person who showed this kind of interest is more than twice the consent rate of those who weren't interested (estimated consent probabilities are 22 % and 44 % respectively).

The analysis also yielded significant influences of the interview situation. As expected, the consent rate was higher if the interviewer was present and lower if the questionnaire was mailed or dropped off respectively. Consent was also lower if the response person was difficult to identify in the organisation which could be due to several reasons: maybe because no one in establishment felt knowledgeable enough to answer the survey or felt to have the legitimisation to do so or maybe no one had the time or motivation to participate.

Finally, some interviewer effects could be identified. Especially interviewers' education was relevant: interviewers with a higher secondary degree showed clearly higher success rates in receiving when compared to interviewers with an lower secondary degree. The relationship seems to suggest a curvilinear effect for high school degree was not as pronounced as the one for higher secondary, but still significant, at least in the model including interviewer level covariates only. As was to be expected, mean item non-response per interviewer was strongly and significantly related to consent, suggesting that interviewer features which elicit item non-response work similarly on consent questions.

Literatur

- Calderwood, Lisa, & Lessof, Carli. 2009. Enhancing longitudinal surveys by linking to administrative data. *Methodology of Longitudinal Surveys*. Chichester: John Wiley & Sons, 55–72.
- Gruhl, Anja, Schmucker, Alexandra, & Seth, Stefan. 2012. *Das Betriebs-Historik-Panel 1975-2010: Handbuch Version 2.1. 1 (Establishment-History-Panel 1975-2010)*. Tech. rept. Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].
- Schnell, Rainer. 2012. *Survey-Interviews: Methoden standardisierter Befragungen*. VS Verlag für Sozialwissenschaften.
- Wagner, Gert G., Frick, Joachim R., & Schupp, Jürgen. 2007. The German Socio-Economic Panel Study (SOEP) : Scope, Evolution and Enhancements. *Schmollers Jahrbuch*, **127**(1), 139–169.

Tab. 1: Multivariate Logit-Modelle

	M1		M2		M3		M4	
	OR	t	OR	t	OR	t	OR	t
Consent to Record Linkage	1.072	(0.681)					1.063	(0.550)
Establishment: Income Inequality	0.661	(-0.531)					0.512	(-0.818)
Type: Head office	0.426	(-1.078)					0.365	(-1.213)
Type: Sub-Company	0.476	(-0.957)					0.499	(-0.849)
Type: Local Branch	0.910	(-0.119)					0.771	(-0.313)
Type: Company (or Franchise)	0.613	(-0.600)					0.465	(-0.893)
Type: School/University								
Type: Public Agency	1.264	(1.306)					1.201	(0.932)
Economic Sector: Manufacturing								
Economic Sector: Services	0.797	(-0.985)					0.902	(-0.409)
Type of activity: Private	0.463	(-0.995)					0.396	(-1.133)
Type of activity: Public								
Type of activity: Nonprofit								
Active: Local order regional	0.454	(-0.687)					1.094	(0.075)
Active: Within GER	1.217	(0.783)					1.235	(0.764)
Active: Within Europe	0.786	(-0.937)					0.846	(-0.593)
Active: Outside Europe	0.963	(-0.122)					0.993	(-0.020)
Establishment: No. departments (cat.)	0.957	(-0.469)					0.851	(-1.547)
Establishment: No. hierarchies (cat.)	1.081	(1.120)					1.094	(1.155)
Establishment: Formalization	1.016	(0.508)					1.016	(0.443)
Establishment: Success	0.927	(-0.094)					1.332	(0.323)
Organizational success (Squared)	1.032	(0.224)					0.962	(-0.254)
Existence of wage agreement	0.968	(-0.190)					0.907	(-0.518)
Management: Owners	0.920	(-0.406)					0.943	(-0.258)
Autonomy: Hiring decisions	1.571 *	(2.256)					1.486+	(1.829)
Company policy: informing employees	0.728***	(-3.327)					0.754**	(-2.615)
Transparency: Income is public	0.752+	(-1.944)					0.744+	(-1.850)
TP Gender: Female							0.684*	(-2.215)
TP-Age: 16-30			0.734+	(-1.955)			0.625	(-1.025)
TP-Age: 31-49			0.512	(-1.553)			0.607**	(-2.941)
TP-Age: 50+			0.651**	(-2.806)				
TP-Education: Haupt								
TP-Education: Real			0.636	(-0.792)			0.481	(-1.178)
TP-Education: (Fach-)Abitur			0.794	(-0.520)			0.632	(-0.955)
TP-Education: University			0.726	(-0.715)			0.658	(-0.858)
TP-Education: Other			0.715	(-0.787)			0.697	(-0.775)
TP-Division: Management								
TP-Division: Human Resources			0.791	(-1.164)			0.610*	(-2.179)
			0.931	(-0.392)			0.980	(-0.098)

TP-Division: Public Relations	1.421+	(1.825)	1.229	(0.974)
TP-Division: Controlling	1.416*	(2.065)	1.285	(1.365)
TP-Division: Other	0.894	(-0.526)	0.750	(-1.240)
TP-Division: Leading Position	1.207	(0.745)	1.041	(0.146)
TP: Time with Employer	1.000	(-0.020)	1.002	(0.265)
TP Interest: Asked for Report	2.687***	(6.084)	2.696***	(5.570)
Mode: PAPI			1.836	(1.626)
Mode: Self			1.212	(0.490)
Mode: PAPI/Self				
Interview: Difficulty identifying TP			1.155*	(2.076)
Visits: One			0.693	(-0.969)
Visits: More than one			0.588	(-1.470)
Visits: None				
Colleagues involved: Yes			0.586+	(-1.664)
Colleagues involved: No			0.651	(-1.424)
Colleagues involved: Unclear				
TP thorough: Very			1.141	(0.844)
TP knowledgeable: Very			1.297	(1.467)
TP knowledgeable: Real			0.789	(-1.423)
TP knowledgeable: Other			1.006	(0.681)
Interviewer: Age			0.995	(-0.481)
Interviewer: Age Squared			1.740+	(1.936)
Int.-Education: Real			3.273***	(3.635)
Int.-Education: (Fach-)Abitur			2.110**	(2.649)
Int.-Education: University			2.168+	(1.752)
Int.-Education: Other				
Int.-Education: Haupt				
Interviewer: Experience establishment survey			0.733	(-0.955)
Interviewer: N Interviews / Year			1.000	(0.532)
Interviewer: Mean INR (log)			0.743***	(-3.387)
Interviewer: Difficulty establishment surveys			0.987	(-0.173)
Interviewer: Data sensitivity			0.901	(-1.250)
Establishment: Number of Employees			1.000+	(1.950)
Region: North	1.000	(1.384)	1.000+	(1.808)
Region: West	0.713	(-0.351)	0.635	(-0.468)
Region: South	1.136	(0.566)	1.296	(1.101)
Region: East	0.867	(-0.745)	0.929	(-0.361)
Constant	1.999	(-0.004)	1.090	(0.423)
	1.536	(0.288)	0.400	(-0.966)
N	943,000	943,000	943,000	943,000
Pseudo-R ²	0.035	0.059	0.056	0.124