Wage data collected by telephone interviews: an empirical analysis of the item nonresponse problem and its implications for the estimation of wage functions

ALFONSO SOUSA-POZA and FRED HENNEBERGER*

1. INTRODUCTION

Perhaps the most attractive properties of telephone interviews are that they are much cheaper than most other survey methods and that hard-to-reach respondents are more likely to participate in a telephone than in a face-to-face interview. Thus, telephone surveys have become very popular in economic and social science research. This popularity also applies to the collection of wage data. There are, however, two problems with wage data collected by telephone interviews. A first problem is that the *validity* of the reported figures could be questioned. The mere fact that the level of concentration on specific (rounded) values is quite high underlines this problem.[1]Although,s on average, the wage data need not be biased, high levels of concentration do indicate that a substantial amount of guessing is taking place. The accuracy of these guesses, however, remains largely unanalysed. In general, CURTIN *et al.* (1989) suggests that wealth figures obtained from surveys may be reported with errors. In an older study, BIELBY *et al.* (1977), however, show that individuals' recall of their own yearly earnings is quite reliable (if the recall period is short enough).

A second problem is associated with the relative numerous item *nonresponses* encountered in income data collected by telephone surveys. There are, in general, three reasons for not reporting wages (see also JUSTER and SMITH, 1997): first, the respondent may not be able to recollect his earnings. Second, the respondent may have a rough idea of the amount but assumes that the interviewer wants a very precise figure. Third, the respon-

* Research Institute for Labour Economics and Labour Law (FAA-HSG), University of St. Gallen, Guisanstrasse 92, 9010 St. Gallen, Switzerland
The authors would like to thank ALOIS FÄSSLER (Swiss Federal Statistical Office), THOMAS KLEIN (University of Heidelberg), HEDWIG PREY (University of St. Gallen), ALEXANDRE ZIEGLER (Stanford University), and an anonymous referee for valuable comments made on an earlier version of this paper. We would also like to thank BRIGITTE BUHMANN, ALOIS FÄSSLER and RAPHAEL GEHRET from the Swiss Federal Statistical Office for their assistance in providing the data. Any remaining errors are the sole responsibility of the authors.

1. The wage data in the 1998 Swiss Labour Force Survey (SLFS), for example, are considerably "heaped" with approximately 13% of all wage figures reported corresponding to one of 10 rounded values (see also SOUSA-POZA, 1999, pp. 110–111). The SLFS wage data is also collected by stylised telephone interviews.

dent may not be willing to publicise his earnings. Although the last point seems the most likely, the first two may play an important role when questioning part-time workers or self-employed individuals, whose earnings within a certain time period are often highly variable. The main problem with such a high nonresponse rate is that if the inclination to report wages is not purely random within a population, then the observed wage data may not be representative. This is especially problematic when such wage data are used to estimate wage functions, since the estimated coefficients may be biased. More specifically, this is the case when the response inclination is correlated with the wage rate, i. e., with the human-capital endowment (see also BEHRMAN and WOLFE, 1984; BEHRMAN *et al.*, 1981). Given the large item nonresponse rates in telephone interviews, this problem could be quite troublesome. However, very few formal and comprehensive analyses exist which explicitly look at income data (see, however, BELL, 1984; GROVES, 1979; ROGERS, 1976; ZWEIMÜLLER, 1992).

In this paper, we analyse this item nonresponse problem using data from the 1998 Swiss Labour Force Survey (SLFS).[2] In HENNEBERGER and SOUSA-POZA (1998), an initial attempt is made to analyse this item nonresponse problem using data from the 1995 SLFS. It is shown that, although the response decision does appear to depend on certain characteristics of the respondent, a large degree of unexplained variation remains. However, in HENNEBERGER and SOUSA-POZA (1998), only a limited number of explanatory variables on the respondents' characteristics are used. Furthermore, the analysis excludes the self-employed since one of the paper's main aims is to analyse the implications of item nonresponse for gender wage discrimination studies. Our aim in this paper is to extend the analysis conducted in HENNEBERGER and SOUSA-POZA (1998) by including several additional explanatory variables, not only on respondents' characteristics, but also on interviewers' and interview characteristics that were collected by the Swiss Federal Statistical Office (SFSO) for the first time in 1998. As far as we can ascertain, no such comprehensive analysis of the item nonresponse problem associated with wage data collected by telephone interviews has been conducted to date.[3] One central aim is therefore to analyse the response inclination. A second aim is to analyse whether or not the estimated coefficients of wage functions based on telephone wage data (with a relative high nonresponse rate) are biased. Finally, it is also worthwhile noting that the SLFS wage data have been widely used in the past to estimate wage functions, primarily for the determination of gender wage discrimination (see, for example, DIEKMANN and ENGELHARDT, 1995; FLÜCKIGER and AHMAD, 1996; BONJOUR, 1997; HENNEBERGER and SOUSA-POZA, 1999).

2. We do not, however, consider the *unit* nonresponse problem here, which could also lead to biased results. Due to the scarcity of data on individuals not participating in such surveys it is often difficult to estimate their participation inclination.
3. In BELL (1984) and ZWEIMÜLLER (1992) as in HENNEBERGER and SOUSA-POZA (1998) only the respondents' characteristics are considered. We are not aware of papers that also consider the interviewers' and interview characteristics in a multivariate analysis.

The paper proceeds as follows: in section 2, the data and empirical specification are discussed. The results are presented in section 3, and the paper concludes in section 4.

2. DATA AND EMPIRICAL SPECIFICATION

The data for this study are taken from the 1998 Swiss Labour Force Survey (SLFS). The SLFS is a nation-wide and representative survey conducted annually by two survey institutes for the SFSO. During computer-aided telephone interviews lasting approximately 20 minutes, individuals are questioned on a number of labour-market related topics (see also SFSO, 1996). In the 1998 survey, the sample size is 16,256 individuals. Furthermore, in 1998, the SFSO also collected data for the first time on the interviewers' characteristics. More specifically, a total of 118 interviewers participated themselves in the SLFS. Thus, numerous variables on interviewers' characteristics are available.

Our aim in this paper is to model the response inclination in the following way:

$$Y_i^* = \beta' X_i + U_i \tag{1}$$

where $Y_i^*$is an unobservable continuous random variable depicting the response inclination, X_i is a vector of regressors, β is the vector of unknown parameters to be estimated, and U_i is a stochastic disturbance term. Furthermore, let

$$Y = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0 \end{cases} \tag{2}$$

Equation (2) thus defines the observable dichotomous variable, which indicates whether an individual reported earnings or not. In this study, individuals who usually work at least 1 hour per week, but who did not report their monthly or annual wages, qualify as nonrespondents. The wording of the question is:[4] *"How much did you earn from paid employment in the past month?"* Respondents were also allowed to state their annual, monthly, or hourly wage, and including or excluding social security payments made by the employee or employer ("AHV- und Pensionskassenbeiträge"). Furthermore, respondents were asked to exclude certain social security payments made by the state ("Kinder- und Familienzulagen"). If one assumes that

$$P(y = 1) = F(\beta' X_i) \tag{3}$$

and that F is the standard normal distribution function, then one is confronted with the standard probit model.

In order to explain the response decision, we have included a number of explanatory variables that can be grouped into four categories: (i) characteristics of the interview, (ii) characteristics of the interviewer, (iii) characteristics of the respondent, (iv) mis-matches between certain characteristics of the interviewer and the respondent. For most of the

4. Interviews were conducted in (Swiss) German, French and Italian.

variables included in the regression analysis, it is difficult to make predictions with regard to the signs of the corresponding coefficients. First, most of the variables can have opposing effects, and, second, there is only a limited amount of empirical research that can be used to develop some stylised facts. In the following paragraphs we do, however, attempt to motivate the choice of some of the explanatory variables.[5]

One can expect that the *interview situation* plays a role in explaining the response inclination. In this paper, we have included three variables that characterise certain aspects of the interview. A variable indicating the number of attempts that were undertaken in order to contact the household was included. It is possible that individuals who are difficult to contact also have a different propensity to report earnings than individuals who are easy to contact. A prediction of the sign of this coefficient is, however, difficult. We have run an OLS regression with the number of attempts to contact an individual as the dependent variable. Our results indicate that primarily young, highly educated, and single individuals who don't own a home, are self-employed, and have irregular working hours are the most difficult to contact. The explanatory power of this regression was, however, not very high (adjusted R^2 equal to 0.04). Interviews taking place after 8pm may not be appreciated by the average respondent. Thus, interviews taking place late in the evening could have a negative effect on a respondent's inclination to report earnings. The same reasoning could apply to interviews taking place on the weekend. We have therefore included two dummy variables that capture such situations.

Substantial research exists that attempts to explain interviewer effects (see, for example, O'MUIRCHEARTAIGH and CAMPANELLI, 1998; SINGER *et al.*, 1989; STOKES and YEH, 1988).[6] We have included four socio-demographic variables that capture an *interviewer's characteristics*: age, high education, low education[7], children[8]. Perhaps one could argue that older interviewers and/or interviewers with a high education can communicate better and therefore have a better response rate. A dummy variable indicating whether an interviewer has a child or not has been included. The motivation for including this variable is that it may identify interviewers (especially among the female interviewers) who need to work out of financial necessity. It is possible that such constrained individuals have different response rates than less constrained individuals. In addition,

5. It is beyond the scope of this paper to analyse the theoretical issues associated with an individual's response inclination. The interested reader is referred to REINECKE (1991) and SCHNELL (1997).
6. It is also interesting to note that sample attrition in panel data also appears to be influenced by the characteristics of the interview and the interviewer (see ZABEL, 1998). Needless to say, the underlying mechanisms at work are similar to those of item nonresponse.
7. Degrees from the following institutions were considered to be "high education": university, technical college ("höhere Fachschule", "Technikon"), and high school ("Matura", "Diplommittelschule"). The following categories were considered to be "low education": no degree (i. e., still in compulsory schooling), only compulsory schooling, and lower apprenticeship schemes ("Anlehre", "Haushaltslehrjahr"). The reference category was primarily made up of apprenticeships ("Berufslehre") and similar qualifications ("Vollzeitberufsschule", "höhere Berufsausbildung").
8. A variable on the gender of the interviewer has also been included in the category "mis-match between the characteristics of the interviewer and the characteristics of the respondent". See below.

and in order to control for the interviewers' experience, we include a variable which captures the number of interviews conducted by the interviewer in the 1998 survey. One possible prediction would be that experienced interviewers have lower nonresponse rates. This, however, need not be the case if it is assumed that an interviewer's motivation is inversely related to the number of interviews that he or she conducts.[9]

Perhaps the most important determinant of the response inclination is the *respondent's characteristics*. We have included nine variables depicting a respondent's socio-economic and demographic characteristics: age, marital status, high education, low education[10], managerial position, home ownership, self-employment, working time, working time squared. One could argue that older individuals (due to different attitudes) are less inclined to report their earnings than young individuals. Furthermore, recall is often assumed to diminish with age (although not necessarily; see DEX, 1991). Married people could hesitate to report their income as they may not always be sure if their partner would approve. Respondents with a high educational level may identify themselves with such surveys and would therefore be less secretive about their earnings. Home ownership is often used as a proxy for the level of total income. It can be speculated that individuals with high household incomes prefer not to report their earnings. Individuals with irregular working times (which often implies shorter working hours) as well as self-employed individuals may, within a certain time period, have irregular earnings, and thus find it more difficult to report. It is also conceivable that self-employed individuals are more secretive about their earnings than other workers.

One potential source of nonresponse is the existence of a *mis-match between the characteristics of the interviewer and the characteristics of the respondent.* We have thus included two variables that measure the absolute difference in age and the difference in education between the interviewer and the respondent.[11] Furthermore, we have included a dummy variable indicating the gender of the interviewer. KOCH (1991) assumes that female interviewers have better success rates than males. This conclusion is supported by REINECKE (1991), although the effect is marginal. Furthermore, in certain survey topics (such as surveys on sexual behaviour), female respondents appear to be more influenced by interviewer-gender differences than men (see CATANIA *et al.*, 1996).

9. For one of the survey institutes participating in the SLFS we also have exact data on the experience of the interviewers. More specifically, we have data on the number of hours that an interviewer spent on Swiss Labour Force Surveys and the number of years that the interviewer has worked for the institute. We have, however, not included these two variables in our analysis, since (at the time of writing) we only had such information for one institute and we did not want to reduce our sample size unnecessarily. However, we have analysed these variables, and an interesting result is that the item nonresponse rate is positively correlated with (job-specific) interviewer experience. It may well be that interviewers who have worked for several years with the same survey institute (i. e., are more experienced) need not necessarily be the most effective interviewers.
10. See footnote 7.
11. The difference in education is a dummy variable that has a value equal to 1 if a mis-match between the educational levels of the interviewer and respondent exists. The level of education can either be "high education", "low education", or "medium education". See footnote 7.

We have restricted the analysis to individuals between the ages of 18 and 62. Such a restriction makes sense since most studies that use the SLFS wage data restrict likewise in order to estimate wage functions. We have also partitioned the sample into a male and female sample, as this allows us to analyse specific gender differences. This is interesting and necessary since the overwhelming majority of papers that use the SLFS wage data do so in order to estimate gender-specific wage functions.

As stated above, one aim in this paper is to determine whether ignoring the item nonresponses leads to a selectivity bias when estimating wage functions. In order to do this, we estimate the following sample selection model:

$$\left.\begin{aligned} Y_i^* &= \beta_i' X_{1i} + U_{1i} \\ Y_i &= 0 \ \text{ if } Y_i^* \leq 0 \\ Y_i &= 1 \ \text{ if } Y_i^* > 0 \\ W_i^* &= \beta_2' X_{2i} + U_{2i} \\ W_i &= W_i^* \ \text{ if } Y_i = 1 \\ W_i & \quad \text{not observed if } Y_i = 0 \end{aligned}\right\} \tag{4}$$

where Y^* is an unobserved continuous variable depicting the inclination to report earnings. This variable is assumed to be normally distributed, with independent error, U_1, which has mean zero and constant variance σ_{U1}^2. X_1 is a vector of regressors (described above). Y is the corresponding dummy variable which is equal to zero if Y^* is less than or equal to zero, and equal to one otherwise. W, the logarithm of the hourly wage rate, is the observed realisation of the latent variable W^*, which is also assumed to be normally distributed, with independent error, U_2, which has mean zero and constant variance σ_{U2}^2. Hourly wage rates were calculated on the basis of the usual hours worked per year. The two error terms are assumed to have correlation ρ, implying that the joint distribution of U_1 and U_2 is bivariate normal. X_2 is the corresponding vector of regressors. More specifically, we estimate MINCER (1974) earnings functions with the following regressors: years schooling, experience[12], and experience squared. Furthermore, all individuals who worked at least 1 hour per week were defined as being employed. This model was estimated by HECKMAN'S (1979) two-step approach, and also by the more efficient ML approach (see, for example, BREEN, 1996, pp. 34–42).

3. RESULTS

3.1. Nonresponse decision

Taking a look at the results for the *female* sample first (see table 1), we note that a total of 648 of the 4072 questioned females did not report their earnings, i. e., one is confronted

12. Experience has been approximated by age – years schooling – 6.

with an item nonresponse rate of approximately 16%. The results for the probit models show that several variables have a significant effect on the response inclination.[13] The number of attempts to contact a household has a significant and negative effect on the response inclination. This is the only significant situational variable in the female sample. Four interviewer characteristics have significant effects. Perhaps contrary to expectations, the interviewer's age variable has a negative coefficient. This implies that older interviewers do not have a better response rate than corresponding younger interviewers when questioning women. In fact, one could instead argue that younger interviewers are more motivated or persuasive. The number of interviews within a survey year has a negative coefficient implying that the monotony encountered in telephone interviews could have a negative effect on the response rate. An interesting observation is that interviewers with a high and a low education have better response rates than individuals in the reference category. The positive coefficient of the high education variable can be expected. The positive coefficient of the low education variable can be explained by the fact that students (i. e., individuals still in compulsory schooling) are the largest group within this category. It is conceivable that this group is very motivated and persuasive. Older women are less likely to report their earnings. As in BELL (1984), married women are also less willing to report their earnings. Women with a high education have a higher response inclination than women with a lower education. Women who own a home are less likely to report earnings. As stated above, this could imply that women with high household incomes are less willing to publicise their earnings.[14] Individuals who are self-employed are more likely not to report earnings. One explanation for this phenomenon is that a number of self-employed individuals have irregular earnings which makes reporting them more difficult. A further reason is that self-employed individuals are more cautious in publicising their earnings. The probability of not responding increases with lower working times. This may suggest that women working few hours within a specific time period may be confronted with somewhat irregular earning patterns.

13. We tested the homoskedasticity and normality assumptions of the probit models with conditional moment tests described in PAGAN and VELLA (1989). Heteroskedasticity appeared through a number of variables in both the male and the female models. In order to take account of this heteroskedasticity we estimated a probit model with multiplicative heteroskedasticity that uses the general formulation analysed by HARVEY (1976) (see GREENE, 1995, pp. 422–425; GREENE, 1997, pp. 889–891). The results in table 1 are based on such models. The normality assumption only posed a problem in the female sample. In order to check the robustness of our results in the female sample against the normality assumption, we also estimated a logit model. A large departure from normality in the tails of the distribution should generate quite different results in a probit and in a logit model (see KAHN and LANG, 1992). Since the results of the logit model corresponded quite closely to the results of the probit model we conclude that the departure from normality in the female sample does not appear to be a major problem. These results can be obtained from the authors upon request.
14. Similar results for the respondent's characteristics were obtained by BELL (1984), HENNEBERGER and SOUSA-POZA (1998), and (to a lesser extent) by ZWEIMÜLLER (1992) – the only other multivariate studies that we are aware of which analyse the item nonresponse problem of wage data. As mentioned above, however, these studies do not include interviewer and interview characteristics.

Table 1: response decision

	females		males	
	mean	probit model[c]	mean	probit model[c]
constant		1.885***		1.676***
		(0.152)		(0.210)
interview situation				
no. of contacts	6.235	–0.014***	6.633	–0.016***
	(6.823)	(0.003)	(7.164)	(0.003)
after 8pm[a]	0.221	–0.063	0.277	–0.118***
	(0.415)	(0.050)	(0.448)	(0.049)
weekend[a]	0.564	–0.021	0.574	0.005
	(0.496)	(0.044)	(0.495)	(0.049)
characteristics of interviewer				
age	33.672	–0.004***	33.230	0.003*
	(15.094)	(0.001)	(15.122)	(0.002)
no. of interviews $\times 10^{-3}$	221.656	–0.353***	226.503	–0.673***
	(178.607)	(0.127)	(188.663)	(0.067)
high education[a]	0.237	0.114*	0.225	–0.063
	(0.426)	(0.057)	(0.417)	(0.060)
low education[a]	0.276	0.407***	0.290	0.206***
	(0.447)	(0.141)	(0.454)	(0.070)
children[a]	0.546	0.024	0.532	–0.048
	(0.498)	(0.049)	(0.499)	(0.055)
characteristics of respondent				
age	40.085	–0.011***	40.045	–0.009***
	(11.040)	(0.002)	(10.701)	(0.003)
married[a]	0.499	–0.364***	0.587	0.015
	(0.500)	(0.099)	(0.492)	(0.050)
high education[a]	0.217	0.321**	0.283	0.117**
	(0.412)	(0.157)	(0.450)	(0.053)
low education[a]	0.188	–0.138*	0.099	–0.024
	(0.391)	(0.081)	(0.298)	(0.076)
managerial position[a]	0.247	0.054	0.400	0.066
	(0.432)	(0.055)	(0.490)	(0.056)

	females		males	
	mean	probit model[c]	mean	probit model[c]
home owner[a]	0.335 (0.472)	−0.268*** (0.080)	0.374 (0.484)	−0.090* (0.050)
self-employed[a]	0.097 (0.296)	−0.645*** (0.093)	0.137 (0.344)	−0.705*** (0.096)
working time	29.782 (14.623)	0.009** (0.004)	42.512 (9.523)	−0.001 (0.007)
working time squared $\times 10^{-3}$	1100.799 (900.079)	−0.181*** (0.060)	1897.907 (857.404)	0.046 (0.073)
mis-match in characteristics of interviewer and respondent				
age difference compared to interviewer $\times 10^{-2}$	16.267 (11.027)	−0.498** (0.213)	16.240 (10.958)	−0.007 (0.230)
education difference compared to interviewer[a]	0.602 (0.490)	−0.141 (0.051)	0.614 (0.487)	0.076 (0.053)
male interviewer[a]	0.227 (0.421)	0.109* (0.056)	0.219 (0.413)	0.134** (0.062)
no. of observations		4072		4660
$Y = 0$		648		593
log-likelihood		−1593		−1660
pseudo-R^2		0.107[b]		0.065[b]

Note: standard errors in parenthesis
[a] dummy variables
[b] the pseudo-R^2 measure is that of McFadden (1973)
[c] model with multiplicative heteroskedasticity
*/**/*** significant at the 10%/5%/1% levels, respectively

The results in table 1 show that the working time/response inclination curve is a parabolic one with its maximum at about 26 hours. An interesting result is that there is a negative correlation between the absolute age difference of the interviewer and respondent and the inclination to report earnings, whereas a difference in the level of education between the interviewer and the respondent does not have an influence. A further interesting result is that male interviewers are more effective than female interviewers when questioning women respondents.

Turning now to the *male* sample we note that a total of 593 of the 4660 questioned males did not report their earnings, i. e., one is confronted with an item nonresponse rate of approximately 13%. Thus, male respondents have a slightly lower nonresponse rate than females. A similar result was obtained by BELL (1984). In the male sample, not only the number of attempts to contact a respondent, but also the fact that an interview takes place after 8pm has a negative effect on the response inclination. An interesting result is that older interviewers have a positive effect on the response inclination. This result contrasts with that of the female sample. As in the female sample, male interviewers have lower nonresponse rates than women interviewers, the number of interviews has a negative coefficient, and interviewers with a low formal education have higher response rates than interviewers with a medium education. Contrary to the female sample, interviewers with a high education do not have more success in inducing males to respond. Five coefficients for the respondent's characteristics are significant, and the signs correspond to those of the female sample. The coefficients are: age, high education, home ownership, self-employed. The interpretations are similar to those made for the female sample.

Compared to the results in HENNEBERGER and SOUSA-POZA (1998), the goodness-of-fit is considerably better with pseudo-R^2 values of 0.107 and 0.065 for the female and male models, respectively.[15] However, tables A1 and A2 in the appendix show that, although over 85% of the observations are correctly predicted, only a very small percentage of the nonrespondents could be identified with this model (about 14% in the female and only 2% in the male samples).[16] In other words, our models are not able to capture the factors affecting an individual's nonresponse decision very well. Considering the numerous variables used in this study, one can conclude that the response inclination is, to a large degree, random.

In table 2, we present the pseudo-R^2 values for regressions that only included a subset of the explanatory variables. As can be seen, the respondent's characteristics explain the largest part of the variation. This is especially the case in the female sample. The interviewer's characteristics, the interview situation, and mis-matches between certain characteristics of the respondent and the interviewer do not explain very much.[17]

15. In HENNEBERGER and SOUSA-POZA (1998) the pseudo-R^2 values were equal to 0.026 for the female and 0.017 for the male regressions.
16. Predictions are based on the rule that $\hat{D} = 1$ if $F(\hat{\beta}'x_i) > 0.5$ and equal to 0 otherwise. One must, however, be aware that this measure of goodness-of-fit should be interpreted with caution, especially if the sample (as in our case) is unbalanced, i. e., there are considerably more responses than nonresponses (see also VEALL and ZIMMERMANN, 1996). Nevertheless, taken together with the pseudo-R^2 measure, it does support our conclusion that the goodness-of-fit is not very large.
17. This is an interesting result for the SFSO and for the survey institutes since it reveals that item nonresponse cannot primarily be attributed to the characteristics of the interviewers.

Table 2: explanatory power of individual explanatory variable categories (pseudo-R^2 values)[a]

	situation	interviewer	respondent	mis-match
females	0.005	0.016	0.074	0.008
males	0.013	0.024	0.030	0.011

Note: results are based on probit models for different subsets of explanatory variables
[a] the pseudo-R^2 measure is that of McFadden (1973)

3.2. *Wage functions*

Considering the relatively poor predictive power of the models estimated here, and if one assumes that this is primarily due to the randomness of the response decision, then one would not expect the wage functions estimated with observed wage data (i. e., ignoring the item nonresponses) to be seriously biased. This conclusion is supported by the results in table 3. As can be seen, the estimated coefficients do not vary substantially between the OLS estimates of the regression that refrains from making a correction for the potential selectivity bias and the two models which do. The results in table 3 show that there is a negative correlation between the response inclination and the hourly wage rate (λ and ρ are both significantly negative), which would imply that individuals with high hourly wage rates are underrepresented in the observed wage data. A similar result was obtained by Behrman *et al.* (1981). What is also evident, however, is that the estimated selectivity-correction term is quite large. As pointed out in Henneberger and Sousa-Poza (1998) the size of this term is primarily due to the low variance of the estimated hazard rate instrument, which, in turn, is a result of the poor predictive power of the underlying selection model, i. e., the model depicting the response inclination. This is especially the case in the male regressions. Berk and Ray (1982) argue that "if one is unable to explain much of the variation in the selection process (via the selection equation), the hazard rate instrument will have very little variance. One important consequence is high collinearity with the intercept in the substantive equation; in effect, a second constant term has been added" (Berk and Ray, 1982, p. 386). As can be seen in table 3, the main effect of including a selectivity-correction term is a change in the constant term. This change is offset by the large value of the selectivity-correction coefficient.[18] Thus, the negative correlation between the hourly wage rate and the response inclination should be treated with caution.

18. A similar result was obtained by Zweimüller (1992) using data from the Austrian "Mikrozensus".

Table 3: wage functions for females and males

	females			males		
	OLS model without correction for selectivity bias	Heckman's two-step approach[a]	ML estimates of a sample selection model[b]	OLS model without correction for selectivity bias	Heckman's two-step approach[a]	ML estimates of a sample selection model[b]
constant	2.284*** (0.055)	2.344*** (0.056)	2.309*** (0.061)	2.170*** (0.044)	2.304*** (0.047)	2.220*** (0.046)
years schooling	0.066*** (0.003)	0.066*** (0.004)	0.066*** (0.003)	0.069*** (0.002)	0.067*** (0.002)	0.068*** (0.003)
experience[c]	0.024*** (0.003)	0.026*** (0.003)	0.025*** (0.004)	0.041*** (0.003)	0.044*** (0.003)	0.043*** (0.003)
experience squared $\times 10^{-3}$	–0.407*** (0.007)	–0.403*** (0.007)	–0.406*** (0.007)	–0.611*** (0.005)	–0.597*** (0.006)	–0.609*** (0.005)
λ		–0.437*** (0.096)			–0.794*** (0.082)	
ρ			–0.384*** (0.053)			–0.734*** (0.014)
no. of observations	3424	3424	3424	4067	4067	4067
log-likelihood			–3996			–4027
adjusted R^2	0.106	0.118		0.200	0.227	

Note: standard errors in parenthesis
[a] results corrected for heteroskedasticity
[b] only wage function coefficients presented here
[c] approximated by age – years schooling – 6
*/**/*** significant at the 10%/5%/1% levels, respectively

In tables 4 and 5, we present the 95% confidence interval for the estimated coefficients for the male and the female samples. Furthermore, for the two models that take the potential selectivity bias into consideration we show (i) the extent to which their confidence intervals correspond to that of the model that does not make a correction for the potential selectivity bias, and (ii) the t-statistics that test whether or not the estimated coefficients are significantly different to the corresponding coefficient values obtained from the model that makes no correction. What we note is that the confidence intervals for the models that make a selectivity-bias correction and the model that does not are very similar. This is especially the case with the ML estimates that have a coverage of at least 79% for all coefficients in both the male and female samples. In tables 4 and 5, we also show the t-statistics that test the null hypotheses that the coefficients in the models that make a correction have the same values as the coefficients in the model that does not make a correction. In no case can the null hypothesis be rejected at conventional significance levels. This result comes as no surprise since the confidence intervals are very similar.[19]

Table 4: confidence intervals – females

	no correction	Heckman's two-step approach			ML estimates of a sample selection model		
	conf. int.[a]	conf. int.[a]	coverage[b]	t-statistic[c]	conf. int.[a]	coverage[b]	t-statistic[c]
years schooling	0.074;0.058	0.075;0.057	95%	0.135	0.074;0.059	97.5%	0.056
experience	0.031;0.016	0.033;0.019	72%	–0.737	0.033;0.016	86.8%	–0.338
experience squared $\times 10^{-3}$	–0.256;–0.558	–0.247;–0.558	97%	–0.063	–0.254;–0.558	99.4%	–0.015

[a] 95% confidence interval
[b] percentage coverage with the confidence interval for the model with no correction
[c] tests the hypothesis that the coefficient is equal to the coefficient value of the model with no correction

Table 5: confidence intervals – males

	no correction	Heckman's two-step approach			ML estimates of a sample selection model		
	conf. int.[a]	conf. int.[a]	coverage[b]	t-statistic[c]	conf. int.[a]	coverage[b]	t-statistic[c]
years schooling	0.074;0.063	0.073;0.061	71%	0.800	0.074;0.062	89.6%	0.259
experience	0.048;0.035	0.051;0.037	68%	–0.917	0.048; 0.037	79.3%	–0.514
experience squared $\times 10^{-3}$	–0.584;–0.738	–0.459;–0.734	90%	–0.256	–0.493;–0.725	98.6%	–0.032

[a] 95% confidence interval
[b] percentage coverage with the confidence interval for the model with no correction
[c] tests the hypothesis that the coefficient is equal to the coefficient value of the model with no correction

Our analysis has concentrated on *gender-specific* nonresponse behaviour and wage functions. As mentioned above, the main motivation for this is that, in the past, the SLFS wage data have nearly exclusively been used for estimating gender-specific wage functions. The results presented here show that no serious selectivity bias arises by simply ignoring the item nonresponses. It is, however, conceivable that wage functions estimated for other subsets of the population are more severely affected by a sample-selectivity bias. This could especially be the case for self-employed individuals and part-time workers (who work less than 20 hours per week, i. e., have approximately a 50% employ-

19. The impression that one gets is that the sample selectivity bias is most noted by the experience coefficient. However, even for this coefficient, no significant differences between the models exist.

ment status or less[20]) since these groups have above-average nonresponse rates.[21] In table 6, we present (in an analogous manner as above) the wage functions for these two groups. An interesting result is that there is a positive correlation between the response inclination and earnings. Thus, one could conclude that self-employed and part-time workers with high hourly wage rates are overrepresented in the observed wage data. However, one also notes in these regressions that the primary effect of correcting for the possible selectivity bias is to reduce the constant term, which, once again, is a result of the poor predictive power of the underlying models depicting the response inclination.[22]

Table 6: wage functions for the self-employed and part-timers

	self-employed			part-timers		
	OLS model without correction for selectivity bias	Heckman's two-step approach[a]	ML estimates of a sample selection model[b]	OLS model without correction for selectivity bias	Heckman's two-step approach[a]	ML estimates of a sample selection model[b]
constant	1.923*** (0.204)	1.756*** (0.212)	1.789*** (0.245)	2.268*** (0.161)	2.215*** (0.163)	2.208*** (0.174)
years schooling	0.074*** (0.009)	0.077*** (0.010)	0.076*** (0.011)	0.077*** (0.010)	0.078*** (0.011)	0.077*** (0.010)
experience[c]	0.033** (0.013)	0.033** (0.014)	0.033** (0.015)	0.020*** (0.008)	0.019** (0.008)	0.018* (0.018)
experience squared 10^{-3}	–0.533** (0.244)	–0.539** (0.247)	–0.534** (0.266)	–0.363*** (0.002)	–0.354** (0.016)	–0.352** (0.016)
λ		0.296** (0.140)			0.278 (0.203)	-
ρ			0.308** (0.153)			0.584*** (0.073)
no. of observations	784	784	784	966	966	966
log-likelihood			–1392			–1505
adjusted R^2	0.078	0.078		0.059	0.061	

Note: standard errors in parenthesis
[a] results corrected for heteroskedasticity
[b] only wage function coefficients presented here
[c] approximated by age – years schooling – 6
*/**/*** significant at the 10%/5%/1% levels, respectively

20. In Switzerland, the usual (most reported) weekly work time is 42 hours per week.
21. Part-time workers and self-employed individuals have a nonresponse rate of 18% and 31%, respectively.
22. We have refrained from presenting the results of the probit models for these two groups. The pseudo-R^2 values for the probit models of the self-employed and part-time workers are 0.086 and 0.091, respectively. Since most part-time workers are women, we also specifically analysed part-time *female* workers, yet the results did not differ very much from those of all part-time workers.

In tables 7 and 8, we present the confidence intervals for the three models, the degree of coverage of the confidence intervals, and the t-statistics that test the null hypotheses whether or not the estimated coefficients are significantly different to the corresponding coefficient values obtained from the model that makes no correction. Once again, there is no evidence that ignoring the item nonresponses leads to a significant selectivity bias in these two groups.

Table 7: confidence intervals – self-employed

	no correction	Heckman's two-step approach			ML estimates of a sample selection model		
	conf. int.[a]	conf. int.[a]	coverage[b]	t-statistic[c]	conf. int.[a]	coverage[b]	t-statistic[c]
years schooling	0.095;0.052	0.100;0.054	86.9%	–0.337	0.100;0.051	90.9%	–0.236
experience	0.063;0.002	0.064;0.001	99.8%	0.005	0.066;–0.001	99.9%	–0.001
experience squared $\times 10^{-3}$	0.031;–1.096	0.030;–1.108	98.9%	0.025	0.079;–1.147	99.8%	0.006

[a] 95% confidence interval
[b] percentage coverage with the confidence interval for the model with no correction
[c] tests the hypothesis that the coefficient is equal to the coefficient value of the model with no correction

Table 8: confidence intervals – part-timers

	no correction	Heckman's two-step approach			ML estimates of a sample selection model		
	conf. int.[a]	conf. int.[a]	coverage[b]	t-statistic[c]	conf. int.[a]	coverage[b]	t-statistic[c]
years schooling	0.100;0.054	0.104;0.052	95.9%	–0.101	0.099;0.054	98.0%	0.046
experience	0.040;0.001	0.037;0.000	91.0%	0.215	0.039;–0.003	87.4%	0.323
experience squared $\times 10^{-3}$	0.000;–0.726	0.006;–0.714	97.5%	–0.057	0.027;–0.732	97.1%	–0.069

[a] 95% confidence interval
[b] percentage coverage with the confidence interval for the model with no correction
[c] tests the hypothesis that the coefficient is equal to the coefficient value of the model with no correction

4. CONCLUSIONS

A first aim of this paper is to analyse whether or not the item nonresponse associated with wage data collected by telephone interviews can be modelled. A second aim is to analyse the potential selectivity bias which could result from ignoring these item nonresponses. Using data from the 1998 Swiss Labour Force Survey (SLFS), data on interviewers, and data on the interview situation, we show that the response inclination depends on a number of factors. In the *female* sample, the number of contact attempts, the age of the interviewer, the number of interviews conducted by an interviewer, the age of the respondent, being married, home ownership, self-employment, and the age difference between the interviewer and the respondent have a negative effect on the response inclination. Interviewers with a low education, respondents with a high education, the working time of the respondent, and male interviewers have a positive effect on the response inclination. In the *male* sample, the number of contact attempts, interviews taking place after 8pm, the number of interviews conducted by an interviewer, the age of the respondent, home ownership, and self-employment have a negative effect on the response inclination. The age of the interviewer, interviewers with a low education, and respondents with a high education have a positive effect on the response inclination. Further results to note are: (i) males are more likely to respond than females; (ii) the respondent's socio-economic and demographic characteristics are the most important nonresponse determinants; (iii) the goodness-of-fit of the female regressions is slightly better than that of the male regression.

However, a large degree of unexplained variation remains. This leads us to conclude that the response inclination is, to a large extent, randomly distributed amongst the underlying population. Since we have used numerous explanatory variables, not only on the characteristics of the respondent but also on the characteristics of the interviewer and the interview situation, this conclusion appears to be justified. Furthermore, this conclusion also applies to self-employed individuals and part-time workers, i. e., individuals that have a relatively high nonresponse rate.

It must, however, be mentioned that we do not have any data on vocal characteristics of the interviewer. Some authors argue that this could be an important determinant of nonresponses (see, for example, OKSENBERG and CANNELL, 1988). Furthermore, we also lack data on the subjective attitudes of interviewers on any particular topic (for example, an interviewer's opinion of setting earnings-related questions), which could also affect the response inclination. However, such subjective attitudes are less of a problem in telephone surveys than in face-to-face interviews.

If we accept that the item nonresponse decision is primarily considered to be randomly distributed, then explicitly ignoring these nonrespondents when estimating wage functions (as most studies do) will not lead to a sample-selection bias. This is an interesting result if one considers the fact that the wage item nonresponse in the SLFS of approximately 14% is by no means trivial. Furthermore, it provides support for those studies that have ignored these nonresponses in the past and also for all those which may do so in the future.

Finally, it must once again be stressed that we have only analysed one specific aspect of the quality of wage data collected by telephone surveys, namely the item nonresponse problem. What remains largely unanalysed (especially with regard to the wage data of the SLFS) is the accuracy of wage data collected by telephone surveys. As pointed out in CURTIN *et al.* (1989), wealth data collected by surveys are problematic (this, however, does not apply only to telephone surveys). Furthermore, we have ignored the unit nonresponse problem, which could also influence the validity of such wage data. If one considers the number of studies which use (and will use) the wage data of the SLFS, then future research on these two topics is also necessary.

APPENDIX: PREDICTIONS MADE BY THE PROBIT MODELS

Table A1: predictions of the probit model – females

		Predicted	
		0	1
Actual	0	91	557
	1	47	3377

Note: Predictions are based on the rule that $\hat{D} = 1$ if $F(\hat{\beta}'x_i) > 0.5$ and equal to 0 otherwise.

Table A2: predictions of the probit model – males

		Predicted	
		0	1
Actual	0	14	579
	1	13	4054

Note: Predictions are based on the rule that $\hat{D} = 1$ if $F(\hat{\beta}'x_i) > 0.5$ and equal to 0 otherwise.

REFERENCES

BEHRMAN, J.R., WOLFE, B.L., TUNALI, I. (1981), "Determinants of Women's Earnings in a Developing Country: A Double Selectivity, Extended Human Capital Approach", Discussion Paper No. 596–80, University of Wisconsin, Madison.

BEHRMAN, J.R. ,WOLFE, B.L. (1984), "Labor Force Participation and Earnings Determinants for Women in the Special Conditions of Developing Countries", *Journal of Development Economics,* 15:259–288.

BELL, R. (1984), "Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income", *Social Science Quarterly,* 65:207–215.

BERK, R.A., RAY, S.C. (1982), "Selection Biases in Sociological Data", *Social Science Research,* 11:352–398.

BIELBY, W.T., HAUSER, R.M., FEATHERMAN, D.L. (1977), "Response Errors on Non-Black Males in Models of the Stratification Process", in: Goldberger, A.S., Aigner, D.J. (eds.), "Latent Variables in Socioeconomic Models", 227–251, North-Holland, Amsterdam.

BONJOUR, D. (1997), "Lohndiskriminierung in der Schweiz: Eine ökonometrische Untersuchung", Paul Haupt, Berne.

BREEN, R. (1996), "Regression Models: Censored, Sample Selected, or Truncated Data", Series: Quantitative Applications in the Social Sciences, No. 111, SAGE Publications, London.

CATANIA, J.A., BINSON, D., CANCHOLA, J., POLLACK, L.M., HAUCK, W., COATES, T.J. (1996), "Effects of Interviewer Gender, Interviewer Choice, and Item Wording on Responses to Questions Concerning Sexual Behavior", *Public Opinion Quarterly*, 60:345–375.

CURTIN, R., JUSTER, F.T., MORGAN, J. (1989), "Survey Estimates of Wealth: An Assessment of Quality", in: Lipsey, R.E., Tice, H.S. (eds.), "The Measurement of Saving, Investment, and Wealth", University of Chicago Press, Chicago.

DEX, S. (1991), "The Reliability of Recall Data: A Literature Review", Working Paper No. 11, ESRC Research Center on Micro-Social Change, University of Essex, Colchester.

DIEKMANN, A., ENGELHARDT, H. (1995), "Einkommensungleichheit zwischen Frauen und Männern. Eine ökonomische Analyse der Schweizer Arbeitskräfteerhebung", *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 131:57–83.

FLÜCKIGER, Y., AHMAD, Z. (1996), "Analyse économique des causes des inégalités salariales entre hommes et femmes en Suisse", Rapport scientifique final, PNR 35 "Femmes, Droit et Societé", Berne.

GREENE, W.H. (1995), "LIMDEP Version 7.0 User's Manual", Econometric Software, Castle Hill.

GREENE, W.H. (1997), "Econometric Analysis", 3rd ed., Prentice Hall, Upper Saddle River, New Jersey.

GROVES, R.M. (1979), "Actors and Questions in Telephone and Personal Interview Surveys", *Public Opinion Quarterly*, 43:190–205.

HARVEY, A. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity", *Econometrica*, 44:461–465.

HECKMAN, J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47:153–161.

HENNEBERGER, F., SOUSA-POZA, A. (1998), "Estimating Wage Functions and Wage Discrimination using Data from the 1995 Swiss Labour Force Survey: A Double-Selectivity Approach", *International Journal of Manpower*, 19:486–506.

HENNEBERGER, F., SOUSA-POZA, A. (1999), "Geschlechtsspezifische Lohndiskriminierung. Neueste Evidenz von den Mikrodaten aus der Schweizerischen Arbeitskräfteerhebung", *Schweizerische Zeitschrift für Soziologie*, 25:259–279.

JUSTER, F.T., SMITH, J.P. (1997), "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD", *Journal of the American Statistical Association*, 92:1268–1278.

KAHN, S., LANG, K. (1992), "Constraints on the Choice of Work Hours: Agency vs. Specific-Capital", *Journal of Human Resources*, 27:661–678.

KOCH, A. (1991), "Zum Zusammenhang von Interviewermerkmalen und Ausschöpfungsquoten", in: ZUMA-Nachrichten, No. 28, 41–54.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior", in: Zarembka, P. (ed.), "Frontiers in Econometrics", Academic Press, New York, 105–142.

Mincer, J. (1974), "Schooling, Experience and Earnings", National Bureau of Economic Research, New York.

Oksenberg, L., Cannell, C. (1988), "Effects of Interviewer Vocal Characteristics on Nonresponse", in: Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L., Waksberg, J. (eds.), "Telephone Survey Methodology", John Wiley & Sons, New York, 257–269.

O'Muircheartaigh, C., Campanelli, P. (1998), "The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision", *Journal of the Royal Statistical Society,* Series A, 161:63–77.

Pagan, A., Vella, F. (1989), "Diagnostic Tests for Models Based on Individual Data: A Survey", *Journal of Applied Econometrics,* 4:29–59.

Reinecke, J. (1991), "Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte", Studien zur Sozialwissenschaft, Bd. 106, Westdeutscher Verlag, Opladen.

Rogers, T.F. (1976), "Interviews by Telephone and in Person: Quality of Responses and Field Performances", *Public Opinion Quarterly,* 40:51–65.

Schnell, R. (1997), "Nonresponse in Bevölkerungsumfragen. Ausmass, Entwicklung und Ursachen", Leske und Budrich, Opladen.

Singer, E., Frankel, M.R., Glassmann, M.B. (1989), "The Effect of Interviewer Characteristics and Expectations on Response", in: Singer, E., Presser, S. (eds.), "Survey Research Methods: A Reader", University of Chicago Press, Chicago, 272–287.

Sousa-Poza, A. (1999), "The Allocation and Value of Time Assigned to Housework and Child-Care. An Analysis for Switzerland using Data from the 1997 Swiss Labour Force Survey", PhD Thesis, University of St. Gallen, Paul Haupt, Berne.

Stockes, L., Yeh, M-Y. (1988), "Searching for Causes of Interviewer Effects in Telephone Surveys", in: Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L., Waksberg, J. (eds.), "Telephone Survey Methodology", John Wiley & Sons, New York, 357–373.

Swiss Federal Statistical Office (1996), "Die Schweizerische Arbeitskräfteerhebung (SAKE): Konzepte, methodische Grundlagen, praktische Ausführung", SFSO, Berne.

Veall, M.R., Zimmermann, K.F. (1996), "Pseudo-R^2 Measures for some Common Limited Dependent Variable Models", *Journal of Economic Surveys,* 10:241–259.

Zabel, J.E. (1998), "An Analysis of Attrition in the PSID and the Survey of Income and Program Participation", *Journal of Human Resources,* 33:479–506.

Zweimüller, J. (1992), "Survey Non-response and Biases in Wage Regressions", *Economics Letters,* 39:105–109.

SUMMARY

In this paper, the item nonresponse problem associated with wage data collected by telephone surveys is analysed. Using data from the 1998 Swiss Labour Force Survey (SLFS) and data on interviewers, it is shown that the response decision can only partially be explained by the chracteristics of the interview situation, the respondent, and the interviewer. This suggests that the response inclination is, to a large extent, randomly distributed amongst the underlying population. It is therefore argued that wage functions estimated using only the observed wage data are not biased by the large (wage) item nonresponse encountered in telephone interviews.

ZUSAMMENFASSUNG

Der Beitrag analysiert Antwortausfälle bei Lohndaten, die sich bei telephonischen Befragungen ergeben. Wir verwenden Daten aus der Schweizerischen Arbeitskräfteerhebung (SAKE) des Jahres 1998 sowie Daten über die Merkmale der Interviewer. Es wird gezeigt, dass das Antwortverhalten nur zum Teil mit der Interviewsituation sowie den Eigenschaften der Interviewer und der Befragten erklärt werden kann. Dieses Ergebnis legt nahe, dass die Antwortbereitschaft in hohem Masse zufällig innerhalb der zugrunde liegenden Grundgesamtheit verteilt ist. Es wird deshalb argumentiert, dass Schätzungen von Lohnfunktionen, die nur auf den tatsächlich beobachtbaren Lohndaten basieren, nicht durch die hohe Antwortausfallquote verzerrt sind.

RESUME

Cet article analyse le problème des non-réponses dans les données salariales collectées au moyen d'enquêtes téléphoniques. A l'aide de données de l'Enquête Suisse sur la Population Active (ESPA) 1998 et de données sur les enquêteurs, il est montré que la décision de réprondre ou non ne peut que partiellement être expliquée par les caractéristiques de l'interview, de la personne interrogée, et de l'enquêteur. Ceci suggère que la propension à répondre est, dans une large mesure, distribuée au hasard dans la population. Par conséquent, les fonctions salariales estimées sur la base des données salariales observées ne sont pas biaisées par le taux élevé de non-réponses rencontré dans les interviews téléphoniques.