

Johansson, Fredrik

Working Paper

How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML?

Working Paper, No. 2007:22

Provided in Cooperation with:

Department of Economics, Uppsala University

Suggested Citation: Johansson, Fredrik (2007) : How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML?, Working Paper, No. 2007:22, Uppsala University, Department of Economics, Uppsala, <https://nbn-resolving.de/urn:nbn:se:uu:diva-11341>

This Version is available at:

<https://hdl.handle.net/10419/82742>

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.



UPPSALA
UNIVERSITET

Working Paper 2007:22

Department of Economics

How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML?

Fredrik Johansson

Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2007:22
August 2007
ISSN 1653-6975

HOW TO ADJUST FOR NONIGNORABLE
NONRESPONSE: CALIBRATION, HECKIT OR FIML?

FREDRIK JOHANSSON

Papers in the Working Paper Series are published
on internet in PDF formats.
Download from <http://www.nek.uu.se>
or from S-WoPEC <http://swopec.hhs.se/uunewp/>

HOW TO ADJUST FOR NONIGNORABLE NONRESPONSE: CALIBRATION, HECKIT OR FIML?*

FREDRIK JOHANSSON[†]

August 22, 2007

Abstract

When a survey response mechanism depends on the variable of interest measured within the same survey and observed for only part of the sample, the situation is one of nonignorable nonresponse. Ignoring the nonresponse is likely to generate significant bias in the estimates. To solve this, one option is the joint modelling of the response mechanism and the variable of interest. Another option is to calibrate each observation with weights constructed from auxiliary data. In an application where earnings equations are estimated these approaches are compared to reference estimates based on large a Swedish register based data set without nonresponse.

JEL-classification: C15; C24; C34; C42; J31.

Keywords: Earning equations; Nonignorable response mechanism; Calibration; Selection; Full-information maximum likelihood.

*Valuable comments from Anders Klevmarken, Per Johansson, Markus Jäntti, Kurt Brännäs, Lars Lindvall, Matias Eklöf, Patrik Hesselius and participants in the 2006 SUDSWec conference in Uppsala and the twelfth annual SOLE meeting in Chicago are appreciated. This paper is part of an NIA (R03AG21780) and FAS (2001-2830) funded project (Comparison of Survey and Register Data: The Swedish Case) in collaboration with Arie Kapteyn (RAND) and Susann Rohwedder (RAND).

[†]Department of Economics, Uppsala University, Box 513, SE-751 20, Uppsala, Sweden. Phone: 46-18-471 16 36. E-mail: Fredrik.Johansson@nek.uu.se

1 Introduction

Depending on the relationship between nonresponse and the true responding data set nonresponse can take different forms. In this paper we study the case often referred to as missing not at random (MNAR) or having a nonignorable response mechanism. It is then the case that the response probabilities depend both on the observed data and on the nonresponse mechanism, i.e., on the unobserved data, see Little and Rubin (1987). The consequence of nonignorable nonresponse is both biased and less efficient estimates. Other less serious types of nonresponse are; missing completely at random (MCAR) and missing at random (MAR) for which the consequence is less efficient estimates.

Since the seminal work by Heckman (see Heckman (1979)), the conventional method among empirical economists to model samples with nonresponse (or selectivity) has been to use one equation that specifies the selection rule and one regression equation for the variable of interest. This approach is sometimes called the heckit model by analogy. Sampling statisticians have viewed this approach with scepticism, mainly because their primary interest is in inferences to finite populations and not to an infinite population. Another critique mentioned is that one has to make assumptions about the distributional form of the error term.¹ Statisticians propose numerous solutions to handle nonresponse bias. One class of solutions deals with changing the weights given to each respondent – one of these approaches is calibration. An objection against calibration is that it could be viewed as a black box since there is no explicit underlying model. Calibration is designed for the adjustment of nonresponse in inferences to a finite population whereas heckit is used for inferences to a superpopulation. In this paper the calibration weights are used to obtain nonresponse adjusted estimates for inferences to a superpopulation. Hence all inference will be to the parameter estimates generated from the model.

The purpose of this paper is to estimate earnings equations with OLS, using both the calibration and the heckit approach to adjust the earnings estimates for nonresponse. Also estimated is a model using the Full Information Maximum Likelihood (FIML) approach. These estimates will then be used to determine which approach results in the best earnings estimates. The best earnings estimates are the ones where the estimates are as near as possible to an earnings equation based on a sample that is much larger and without any nonresponse – i.e., no selection. The type of earnings equations that are estimated in this paper is frequent in applied labour economics. It has its theoretical basis in human capital theory and is of the sort estimated by Mincer (1974). Since it is probably the most widely used equation estimated in labour economics it is of great importance that these estimates are correct – a statement usually not necessarily true, but controllable in this paper.

The sample frame used is a Swedish register; hence we have access to extensive register information for the whole target population. This provides great opportunities for measuring nonresponse in survey data.

The disposition of this paper is as follows; the next section presents the relevant literature.

¹Since the Heckman (1979) paper both semi-parametric (see Powell (1994)) and non-parametric (see Das, Newey, and Vella (2003)) sample selection models has been developed.

Section three starts with a presentation of the general model to be estimated, then a discussion about the data collection process follows. In section three descriptive statistics is also presented. Section four starts with a description of the different procedures estimated; beginning with the reference model and proceeding with the methodology behind calibration, heckit and the FIML procedures. This section also includes a discussion about the differences between design-based and model-based inference. In section five the results are presented and section six concludes.

2 Literature review

In survey studies it is often assumed that nonresponders and responders have similar distributions of the variables of interest, however this implies that nonresponse is ignorable. In the event the nonresponse is said to be ignorable – imputation can provide unbiased estimates of population parameters, see Copas and Farewell (1998). Propensity scores can also be used to minimize the effects of selectivity when nonresponse is ignorable, see Rosenbaum and Rubin (1984). The major problem with this assumption of ignorable nonresponse is that it is unfeasible to test whether data is ignorable without the aid of additional data. Unfortunately, all these approaches, assuming an ignorable response mechanism, will be systematically in error if the probability of response varies with the variable of interest. In general, an erroneous assumption of ignorable nonresponse results in a nonresponse bias in the estimates.

A rich literature has emerged with different approaches to correct for nonignorable nonresponse using models, see e.g. Vella (1998) for a recent overview of sample selection models. The accepted practice among empirical economists is usually to consider two components of a joint model. The first component regards the relation (or relations) of interest, for which one wants to estimate some parameters of interest. The other component describes the sample selection mechanism. The difficult part in this type of models is usually to specify the process of selection correctly. Given that, several estimation approaches of the relation of interest exist, see e.g. Greenless, Reece, and Zieschang (1982) and Little (1982).

Within the sampling literature emphasis is on model assisted and design consistent estimates. Without specifying an explicit response model auxiliary data are used to compensate for nonresponse. For a review of classical methods like post-stratification, re-weighting and imputations, see for instance Kalton and Kasprzyk (1986) or Bethlehem (2002). A more recent and general approach is calibration; it is general because other re-weighting methods turn out to be special cases of calibration. A calibration estimator uses calibrated weights. These weights are as close as possible, given a distance measure, to the original design weights (the inverse of the inclusion probability) while also respecting a set of constraints.² These constraints are the calibration equations; they ensure that the sample sum values of the weighted auxiliary variable must be equal to the known population totals for that auxiliary variable, see e.g. Deville and Särndal (1992) or Lundström and Särndal (2002).

²The interpretation of a design weight is that an element with a design weight equal to ten will represent itself and nine other population elements.

3 Model structure and data discussion

Estimating earnings equations using survey data implies, in practice, that there is a double selectivity problem. First, and as mentioned in the introduction, the choice of participating in the survey is not random, second, the choice of participating in the labor market is not random either; hence the earnings equations must be corrected for these two types of nonrandomness (Tunali (1986)).

Brose and Klevmarken (1993) and more recently Hamermesh and Donald (2004) formalized the issues of double selectivity with the following three equations

$$y_1^* = x_1' \delta_1 + \epsilon_1 \quad \text{-- responded} \quad (1)$$

$$y_2^* = x_2' \delta_2 + \epsilon_2 \quad \text{-- employed | response} \quad (2)$$

$$y_3^* = x_3' \delta_3 + \epsilon_3 \quad \text{-- earnings} \quad (3)$$

$$y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0 \end{cases} \quad (4)$$

$$y_2 = \begin{cases} 1 & \text{if } y_2^* > 0 \text{ and } y_1^* > 0 \\ 0 & \text{if } y_2^* \leq 0 \text{ and } y_1^* > 0 \\ \text{not observed} & \text{if } y_1^* \leq 0 \end{cases} \quad (5)$$

$$y_3 = \begin{cases} y_3^* & \text{if } y_2^* > 0 \text{ and } y_1^* > 0 \\ 0 & \text{if } y_2^* \leq 0 \text{ and } y_1^* > 0 \\ \text{not observed} & \text{if } y_1^* \leq 0, \end{cases} \quad (6)$$

where y_1^* , y_2^* and y_3^* are latent variables that determine response, labor force participation and the logarithmic of earnings respectively. The x_j -vectors ($j = 1, 2, 3$) are vectors of exogenous explanatory variables assumed to be uncorrelated with the ϵ_j . The objective with this paper is to estimate δ_3 in eq. (3) consistently.³

Since the sample frame was the Swedish longitudinal register data set LINDA we have access to high quality register data. This has three implications; (i) in all regressions register information will be used as independent variables. The earnings measure used is conditioned on positive responses from two survey earnings questions, but instead of using the survey measure for these respondents we used register data. The reasons for this are to reduce the impact of measurement error and other nonsampling errors. Also, the number of observations would be even smaller because of item nonresponse if we were to use survey data. (ii) register information will be used as auxiliary variables to construct the calibration weights. And (iii) the quality, i.e., precision, of the procedures for the correction of nonresponse bias and sample selection can be assessed by comparing with the corresponding relation for the whole population. The comparison of procedures and estimates is facilitated if the same earnings measure is used.

Several different earnings equations will be estimated using the following procedures to

³It should be noted that in eqs. (1) to (3) the order of sequence is of course that the labour market decision comes before the survey decision. It has no practical meaning but for pedagogical reason I follow the order given in Brose and Klevmarken (1993) and Hamermesh and Donald (2004).

adjust for nonignorability:

- (i) The first procedure is the reference model and denoted *Heckit-ref*. The sample used in this specification is that of all individuals in LINDA in the age range 50-65 resulting in a total sample size equal to 117,334 individuals. The sample is so large that for practical purposes this is equivalent to a population analysis. Also, for this population there is full response therefore Heckit-ref only adjust for the decision to enter the labor market, this is done by using the heckit approach. The next procedure is denoted *Heckit-res* and is specified as Heckit-ref but using the restrictive sample instead, i.e., the data originated from the survey. The inverse mills ratios for these procedures are estimated using a probit model.

All these procedures that are described next use the same restrictive sample as heckit-res when estimating the earnings equation.

- (ii) Procedure *Cali* uses calibrated weights to compensate for nonresponse but no adjustments for the labor market status. Whereas procedure *Cali-heck* also use weights to adjust for the selection of being employed or not – this is done by using the inverse mills ratio from estimating a calibrated weighted probit explaining the labor market status.
- (iii) Procedure *Two-probit* uses the heckit approach to adjust for selection where the selection rules are being a respondent or not and being employed or not, given that you are a respondent. The inverse mills ratios are estimated by two separate probit models. Procedure *Bi-probit* deal with the same type of selectivity issues, with the exception that a censored bivariate probit model is used to estimate the mills ratios.
- (iiii) Procedure *FIML* uses the Full Information Maximum Likelihood approach with the same selection rules as for procedure Two-probit and Bi-probit. FIML is an approach assumed to be the most efficient within the heckit family.

Comparability exists between procedures Heckit-ref and Heckit-res – the differences in these estimates is due to sampling; Cali and Cali-heckit – differs since Cali-heckit also models the labor market decision. Cali-heckit could also be compared with Heckit-ref and Heckit-res that uses heckit instead of calibration. Then finally Cali-heckit, Two-probit, Bi-probit and FIML could be compared – the first uses a combination of calibration and heckit, the next two use different specifications of heckit and procedure FIML estimate all three equations simultaneously.

Finally all of the estimated earnings equations will have the same variable specification except for the procedures that also includes the inverse mills ratios. All earnings equations are estimated using ordinary least squares (OLS) except for the last procedure that uses maximum likelihood.

3.1 Data and descriptive statistics

The data originates from a survey conducted in Sweden in 2003 by Statistics Sweden (SCB). The year of primary focus for this survey was 2002. The survey included questions originated from the U.S. HRS (Health and Retirement Study) questionnaire and the European SHARE (Survey of Health, Aging, and Retirement in Europe) questionnaire. A random sample was drawn from the 2001 wave of LINDA (Longitudinal INDividual DAta) and the target population was individuals in the age range 50-84. Since the target population includes retirees, and we are only interested in the working population, all analysis in this paper is conditioned on being below 65 years old, i.e., below the customary retirement age in Sweden.

Using LINDA as a sample frame was appropriate since it provided access to the necessary register information for the whole target population. LINDA is a Swedish register-based data set that is representative for the population from 1960 and onwards. In LINDA the household members of every sampled individual are also included. A cross-section of LINDA is a random sample of the Swedish population containing a few hundred thousand individuals. For a more detailed discussion about the data used see Johansson and Klevmarken (2006).⁴

In Table 1 the data used for the correction procedures are divided into sub-samples according to eqs. (4) to (6). Respondents have positive earnings if their answers were positive on the following survey questions: "*Did you have any income from employment in 2002?*" and "*How much did you earn per month in 2002, before taxes?*". Procedures Two-probit and Bi-probit will both use the full sample, i.e., all 819 respondents when estimating eq. (1), whereas when estimating eq. (2) procedure Heckit-res and Two-probit will use the restrictive sample of 504 respondents and Bi-probit uses the full sample which is possible because the use of a censored model. The discrepancy of 315 individuals will be zeros in the first equation (eq. (1)) and missing values in the second equation (eq. (2)). The consequence of this is that the censored bivariate model contains more information than the two separate probit models and should accordingly result in estimates with higher efficiency. Then, finally, conditioning both on being a respondent and having positive earnings results in a sub-sample of 317 respondents. These are the observations used to estimate eq. (3) for all procedures except for Heckit-ref.

Table 1: Response by earnings status

Status	Earnings	No earnings	Not observed	Total
Respondents	317	187	.0	504 (61.7)
Nonrespondents	.0	.0	315	315 (38.3)
Sample size	317	187	315	819 (100.0)

Note: Column percent within parenthesis.

Totally there are 315 individuals who are not observed; undoubtedly some of these would have been classified as employed given that they had responded. Estimating earnings equations with these data could be problematic due to double-selection. Since the majority of the

⁴For a comprehensive review of LINDA, see Edin and Fredriksson (2000).

respondents are employed the problem of selection into the labor market is most likely not as important as selection into the response group.

To estimate earnings equations we need the earnings per hour. LINDA contains two measures of hours worked (denoted TJOMF and ARBTID in LINDA). The first variable is a variable between 0 and 100, which measures the share of full-time work. This variable has missing values for 30 respondents, for those we used the second variable. This variable is a categorical variable with five brackets. Bracket (i) 0 hours, (ii) between 1 and 15 hours, (iii) between 16 and 19, (iv) between 20 and 34 hours, (v) more than 35 hours of work per week. We used the mean value within each bracket.⁵ As an example, take an individual in bracket (ii): Then the constructed TJOMF value for this individual would be: $\left(\left(\frac{1+15}{2}\right) \times 4.3\right) / 165 = 20.8$, where 4.3 is the normal number of weeks worked per month and 165 is the normal number of hours worked per month in Sweden. The individual works 20.8 percent of full-time. These values are then imputed as actual TJOMF values for the respondents with missing values. Finally all hourly earnings are calculated with the following formula: $\text{Hourly earnings} = \left(\frac{\text{Monthly earnings}}{(\text{TJOMF}/100)}\right) / 165$.

For all respondents with positive earnings we used a LINDA earnings measure (denoted CARBL and it is the annual labor earnings) instead of the survey alternative, this is because of eliminating the impact of measurement errors, it also facilitates the comparison with the reference model. All observations, regardless of sample, with an hourly wage below 20 and above SEK 600 were treated as outliers – and omitted.⁶ The earnings measure for Heckit-ref used the same type of data but includes all individuals in LINDA, this resulted in a sample of 35,046 individuals that had positive earnings and an additional 82,288 that had zero earnings. This sample of 117,334 observations is used to estimate the mills ratio included in procedure Heckit-ref.⁷

The independent variables used in the estimations are; if living in a big city (the reference category), smaller city or in an urban area, number of years in school (Schooling), labor market experience (Experience) and squared experience ((Experience)²). Schooling is a discrete variable that takes on the values 6, 9-18 or 20 as a measure for the total number of years in school. Whereas experience is defined as age minus 7 (the mandatory school attendance age) minus schooling.

Defining a valid measure of experience for females is complicated – primarily because females are out of the labor market due to childbirth. How to measure this is not straightforward, especially since LINDA contains no information about the total number of child births. There is only information about the number of children living in each household in 2002. Since the frame population is elderly people the majority of all children have most likely already moved. Consequently, due to data limitations, a reliable measure of experience for females can not be

⁵None of the respondents worked zero hours according to LINDA. Respondents belonging to bracket (v) were treated as full-time workers.

⁶This resulted in that three observations were omitted from the restricted sample, two from the left tail of the earnings distribution and one from the right tail, that observations had an hourly earnings equal to SEK 930. From the large sample a total of 883 observations were omitted.

⁷Available if requested is the distributions of the monthly hourly earnings both for the population and for the restricted sample. The two distributions have the same shape – the right tail is large and significant.

constructed. A solution to this problem is to define experience likewise for males and females and include a dummy variable equal to one if female and zero otherwise.⁸

In Table 2 summary statistics are presented both for the population used to estimate the reference model, 35,046 individuals and for the restrictive sample of 317 individuals. The full sample has mean hourly earnings of approximately SEK 8.0 more than the small sample. From comparing the independent variables we see that due to the elderly sample the mean experience is unusual high. Individuals belonging to the large sample are also less likely to be females, more likely to live in a larger city, has a lower schooling attainment and lower experience. Differences in the standard deviations are lower than the differences in mean values, although they are low too.

In the Appendix summary statistics for the sample used to estimate eqs. (1) and (2) are presented. Comparing Table 2 with Table 4 gives an idea of the differences between the relevant samples. From this it is clear that the sample in Table 4 is less likely to be females, but the geographical distribution is similar to the samples in Table 2.

Table 2: Descriptive statistics for the samples used to estimate the wage equations

Variable	Mean	S.D.	N
<u>Procedure Heckit-ref</u>			
Hourly earnings	152.19	69.40	35,046
If female	.504	.499	35,046
Demographic area			
Major city	.337	.473	35,046
Urban	.548	.497	35,046
Rural	.113	.317	35,046
Schooling	10.87	2.360	35,046
Experience	37.90	4.673	35,046
(Experience) ²	1,465	357.8	35,046
<u>All other procedures</u>			
Hourly earnings	147.57	75.61	317
If female	.529	.499	317
Demographic area			
Major city	.319	.466	317
Urban	.539	.499	317
Rural	.142	.349	317
Schooling	10.98	2.301	317
Experience	37.96	4.622	317
(Experience) ²	1,462	351.9	317

⁸Dividing the 35,046 observations into males and females and testing the following hypothesis for all δ_3 :s in eq. (3)

$$\begin{aligned}
H_0 &: \delta_3^{Males} - \delta_3^{Females} = 0 \\
H_A &: \delta_3^{Males} - \delta_3^{Females} \neq 0,
\end{aligned}$$

resulted in that we could not reject the null hypotheses for the intercept and the experience coefficients. Implicating that the male and female samples has similarities, a result that provides support for our approach of including a gender variable.

4 Approaches for handling a nonignorable response mechanism

After the rather short and nontechnical literature review above, the different procedures used in this paper will be presented in a more technical setting. There will also be a discussion about which types of inference that can be drawn from the calibration approach.

4.1 The reference estimates

The first model to be estimated is the reference model and it is denoted Heckit-ref. It is considered to be the reference since the sample is large enough to be viewed as a population. And since there is no nonresponse in this data source, and hence no nonresponse bias, these are the correct estimates served as reference to the other procedures estimated.

The population used is all individuals between 50 and 65 in LINDA. Heckit-ref has a sample size equal to 117,334 observations when estimating the mills ratio that adjusts for the decision to enter the labor market, and where a subset of 35,046 observations are used for the earnings equation. The mills ratio is estimated using a probit model. I.e., 117,334 observation are used when estimating eq. (2) from where a mills ratio is constructed and included in eq. (3) that is estimated using 35,046 observations. This population includes the respondents who will be used in the other earnings regressions using the correction procedures, i.e., the individuals who are used in the correction procedures are also included in this procedure with exactly the same earnings measure.

4.2 The calibration approach

The access to high quality auxiliary information is essential to be able to successfully adjust for nonresponse bias when using the calibration approach. Performed correctly such access will reduce both the nonresponse bias and the variance of the calibration estimator.

In Lundström and Särndal (2002) the criteria on an auxiliary vector is stated, an auxiliary vector is the vector containing the auxiliary information used for the construction of the calibration weights. The auxiliary vector should as far as possible fulfil one or both of the following principles:

- (i) Explains the variation of the response probabilities.
- (ii) Explains the variation of the main study variables.

Fulfilling the first principle reduces the nonresponse bias in the estimates for all study variables. Fulfilling principle (ii) reduces the bias primarily in the estimates for the main variables. The minimum requirement to qualify as an auxiliary variable is that the information is available for every sampled element.

The remaining part of this section will give a technical description of the calibration approach, it is based on the work of Deville and Särndal (1992) and Lundström and Särndal

(2002). Start with a finite population of N elements, namely $U = \{u_1, \dots, u_k, \dots, u_N\}$. Where y_k is the value on the target variable for the k :th element. Assume that s is a probability sample with n observations drawn from the target population U , with a probability $p(s)$. The inclusion probabilities are then $\pi_k = \sum_{s \ni k} p(s)$ and they are known for all k belonging to the population U . Also assume that the design is such that $\pi_k > 0$ for all elements k . The sample design weight of element k is d_k and equal to $1/\pi_k$. Figure 1 below displays the response set. Since we still have not introduced nonresponse the inner circle can be ignored. But it is easy to see the relationship between the population U and sample s .

Then assume that there is not full response, i.e., we are assuming that $r \neq s$.

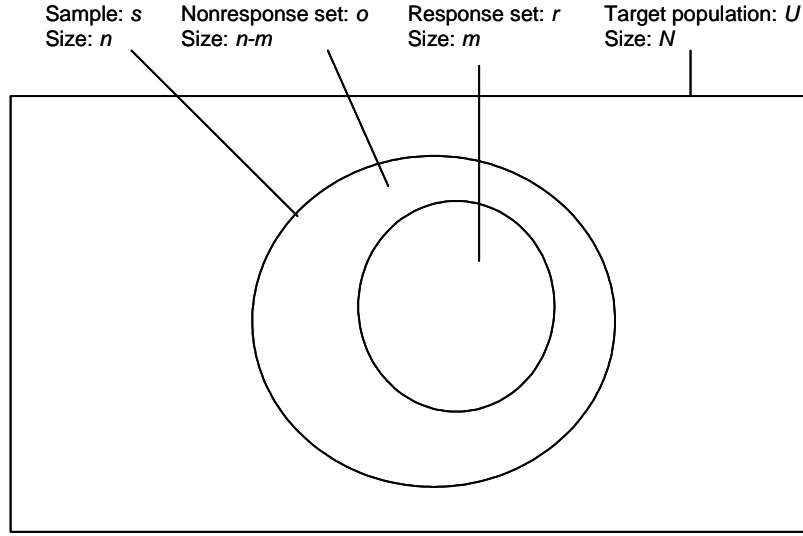


Fig. 1: Illustration of the response set. Source: Lundström and Särndal (2002).

In Figure 1 the response set for the case with nonresponse is illustrated, i.e., now the relationship between population U and samples r , s and o is illustrated.⁹

When data suffer from nonresponse there will be some nonresponse bias. In the presence of nonresponse the desirable properties of the chosen estimator are; (i) a nonresponse bias as small as possible, (ii) a small total variance, where the total variance is the sum of sampling variance and nonresponse variance, and (iii) when there is full response the estimator should agree with the estimator used as if it were full response.

The calibration estimator when nonresponse is present, could be formed as a linearly weighted sum of the observed y_k -values. Thus

$$\hat{Y}_W = \sum_r w_k y_k, \quad (7)$$

where $w_k = d_k v_k$ with

$$v_k = 1 + (\sum_U \mathbf{x}_k - \sum_r d_k \mathbf{x}_k)' \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \text{ for } k \in r, \quad (8)$$

⁹A well-known result in the survey literature is that there are fundamental differences between nonresponders that refuses to participate and nonresponders that are not located. A result also found in this survey, see Johansson and Klevmarken (2006). It has been mentioned that calibration should be applied restrictively when the majority of the nonresponders are not located, see Thorburn (2005). In this survey this is not an issue since 15.8 percent is not located and 22.6 percent are refusals.

where \mathbf{x} is the auxiliary vector, its value for element k is denoted by the column vector $\mathbf{x}_k = (x_{1k}, \dots, x_{jk}, \dots, x_{Jk})'$. Where x_{jk} is the value for element k of j :th auxiliary variable. The difference between eq. (8) and the equation used if $r = s$, when there is full response, is that in eq. (8) the summation is over the response set r whereas in the full response case summation is over the whole sample s .

The principle behind eqs. (7) and (8) is to minimize a function measuring the distance between the d_k and w_k weights, given the calibration equation

$$\sum_r d_k v_k \mathbf{x}_k = \sum_U \mathbf{x}_k.$$

In summary, the idea behind the calibration approach is to find a new set up of weights that is as close as possible to the old sample design weights d_k , but give estimates that coincide with known population totals. For example, imagine that being married or not explain some of the variation in the response probabilities and is therefore included in the auxiliary vector \mathbf{x}_k . Then the estimate of the number of married in sample r will be identical with the number of married in the population U .¹⁰ How the calibration estimator succeeds in fulfilling the desired properties (i) and (ii) depends on the quality of the auxiliary vector \mathbf{x}_k . Some \mathbf{x}_k -vectors perform better than others.

4.2.1 How to decide the auxiliary vector?

In the paragraph above it was stated that the quality of the calibrated estimates depend on the information included in the auxiliary vector \mathbf{x}_k . Since our survey data has a large and rich sample frame we could include numerous variables in \mathbf{x}_k as long as they fulfil the requirements set on an auxiliary vector. The negative consequences of including abundant information in \mathbf{x}_k are that some of the weights v_k could become negative, a result, given its definition above, viewed to be non-intuitive. And also that including too many variables could result in empty cells (categories without any observations) or very small cells because of the rather small sample.

Above it was stated that the auxiliary vector ideally should fulfil one or two of the two criteria's. The construction of a powerful auxiliary vector that fulfils these criteria's is built somewhat on the results in Johansson and Klevmarken (2006). These results tell us that calibration should be executed against population totals for immigrants, marital status and degree of urbanization; while for instance, gender does not contribute much too explaining response. But these results are based on a partly different sample.

The formal approach is to start by analysing if the response rates for the working population is different conditioned on various potentially important auxiliary variables. The primary interest is on those 317 individuals that have positive earnings, i.e., following the notation from above $r = 317$. The more these response rates differ the more likely they are of explaining the

¹⁰Deville and Särndal (1992) introduces a distance measure that measures the difference in the calibrated weights, v_k , and the original design weights d_k . After minimization of this distance measure, using the method of Lagrange, it can be shown that we have the desirable property of unbiased estimates.

variation in the response behaviour – and should be included into the auxiliary vector.¹¹ The final auxiliary vector to estimate procedure Cali consists of; yearly earnings, sorting individuals into quartiles,¹² living in a big city, a smaller city or in a rural area and finally being married or not. This results in 24 different outcomes to group our auxiliary data into.¹³ The weights are then used in an OLS earnings regression. For procedure Cali-heckit the same auxiliary vector is used but here the primary interest is on those 504 ($r = 504$) individuals that have responded regardless if they reported earnings, i.e., a different response set than for Cali. These weights are then used in a probit model (eq. (2)) estimating the probability of having earnings, from where a mills ratio is constructed and included in the earnings regression eq. (3)).

4.2.2 Differences between design-based and model-based inference

With probability statements the analyst makes inference from a sample to a larger population using, for instance confidence intervals or hypothesis tests. Inference to finite populations is based on design-based, model-assisted or model-based approaches. The model-based approach can also be used for inference to an "infinite" population or a "superpopulation" defined by a model. According to Särndal, Swensson, and Wretman (1992) the design-based inference is focused on the present state of the finite population, whereas model-based inference is more focused on the process that underlies the finite population. The design-based approach uses sample weights, e.g., the inverse of the selection probabilities to achieve consistent estimates of finite population statistics. These weights have not necessarily a role to play in model-based inference. The primary interest of the model builder is the causal system, e.g., relating y_3^* to x_3 in eq. (3), and not the finite population. In this kind of inference sampling weights are needed if inclusion probabilities are a function of endogenous variables. The sampling design in this case has been called nonignorable.

In addition to the difference in focus the main difference between these two kinds of inferences is how randomness is introduced in order to give stochastic structure to the inference. In design-based inference the source of randomness is by sampling units from the finite population. In inference to a superpopulation, the values of response associated with the units of the finite population are treated as realizations of random variables. The study variables are random variables and hence any finite population values are outcomes of these random variables. The function of the model is to describe the process by which the population values are generated.

If we were to categorize the procedures used in this paper according to kind of inference Heckit and FIML are model-based and used for inferences to a superpopulation. The calibration approach has been designed to compensate for nonresponse in inferences to a finite population. It will in general not give unbiased and consistent estimates of finite population statistics, but the idea is to use auxiliary information to reduce the bias as much as possible. Särndal and Lundström (2005) discuss the bias of the calibration estimator and introduces the concept of

¹¹The response rates for different variables can be provided if requested.

¹²Observations with a yearly labor earnings below SEK 58,000 and above SEK 30 000,000 were deleted.

¹³There are 24 different outcomes since the four earnings variables, that measure each quartile, can only be true once for each individual, the same is true for the geographical variables.

"near unbiased".

When using the calibration approach in this paper we are, however, not interested in finite populations, but rather in a superpopulation defined by a model. Using the analogy of a non-ignorable sampling design the question is if one can use the calibration weights to compensate for selective nonresponse when estimating the model parameters. If we had a method which estimated finite population moments consistently and an estimator of the model parameters which uses finite population moments to produce consistent estimates of the parameters, then the same estimator using the consistent estimates of the finite population moments, rather than the true moments, is consistent. Unfortunately the consistency of the calibrated estimates is not guaranteed and using the calibrated weights to estimate model parameters thus not in general give consistent estimates. It depends on the auxiliary information used in the calibration. But, with a good choice of auxiliary variables one can hope to get parameter estimates with a small (asymptotic) bias. As a first shot towards the usefulness of the calibration approach in estimating model parameters we will in this paper compare the calibrated estimates with estimates obtained from data without nonresponse.

4.3 The heckit approach

The sample selection model consists of two equations in its simplest form, a participation equation and a regression equation. A typical situation is when the outcome variable is observed for only part of the sample. The reduced form estimates obtained by conventional linear regression techniques from the sub-sample with complete observations will be biased for the population parameters, unless the sub-sample happens to be a random subset of the original sample. The only requirement put on data, enabling the estimation of consistent population parameters, is that data originates from random sampling.

Hamermesh and Donald (2004) states that eqs. (1) to (3) can solve the problem of nonrandomness in response and employment. The regression function with only respondents included in the sample could be written as

$$E(y_3 | x_3, \varpi) = x_3' \delta_3 + \sigma_3 E(\epsilon_3 | x_3, \varpi), \quad (9)$$

where ϖ denotes the joint outcome of the two selection rules and σ_3 is the standard deviation of ϵ_3 , an unknown scale parameter. Estimating eq. (9) for the selected sample with the conditional expectation $E(\epsilon_3 | \delta_3, \varpi) = 0$ would result in the same estimated parameters as if we used a nonselected sample or that the selection process was ignorable. The only cost is a loss in efficiency. Whereas if $E(\epsilon_3 | \delta_3, \varpi) \neq 0$ there is a nonignorable response and the consequence would be a selection bias. To correct for this bias we need information about the conditional distribution $E(\epsilon_3 | x_3, \varpi)$ of the error term.¹⁴

¹⁴In Heckman (1979), who deals with only one selection rule, the solution to this selection bias is to first estimate a probit model of the selection rule. I.e., estimate the probability of being in the labor market to a set of determinants; use these estimates to compute the inverse Mills ratio. Finally include that computed variable as a covariate in the earnings equation.

Since we know that we only observe y_3 if $y_1 > 0$ and $y_2 > 0$ the conditional expectation of eq. (9) could be written as

$$\begin{aligned} E\left(\epsilon_3 \mid x, \epsilon_1 > -x'_1\delta_1, \epsilon_2 > -x'_2\delta_2\right) &= \sigma_3\rho_{13}\frac{\phi(x'_1\delta_1)}{\Phi(-x'_1\delta_1)} + \sigma_3\rho_{23}\frac{\phi(x'_2\delta_2)}{\Phi(-x'_2\delta_2)} \\ &= \sigma_3\rho_{13}\lambda_1 + \sigma_3\rho_{23}\lambda_2, \end{aligned} \quad (10)$$

where the ρ_{13} is the correlation between eqs. (1) and (3) and ρ_{23} is the correlation between eqs. (2) and (3). The λ_i ($i = 1, 2$) terms are the inverse mill's ratios. The mills ratio is a monotone decreasing function of the probability that an individual is selected into the sample $\Phi(-x'_i\delta_i)$ or $1 - \Phi(x'_i\delta_i)$. $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal pdf and cdf respectively. Inserting eq. (10) into eq. (9) results in a regression equation that takes into account that y_3 is observed for only part of the sample. Then

$$y_3 = x'_3\delta_3 + \sigma_3\rho_{13}\lambda_1 + \sigma_3\rho_{23}\lambda_2 + \sigma_3V_3, \quad (11)$$

where $V_3 = \epsilon_3 - \rho_{13}\lambda_1 - \rho_{23}\lambda_2$ with $E(V_3 \mid y_1 > 0, y_2 > 0) = 0$. Inserting the estimated correction terms, $\hat{\lambda}_i$, into eq. (11) results in

$$\begin{aligned} y_3 &= x'_3\delta_3 + \sigma_3\rho_{13}\hat{\lambda}_1 + \sigma_3\rho_{23}\hat{\lambda}_2 + \sigma_3\tilde{V}_3 \\ &= x'_3\delta_3 + \gamma_1\hat{\lambda}_1 + \gamma_2\hat{\lambda}_2 + \sigma_3\tilde{V}_3, \end{aligned} \quad (12)$$

where $\tilde{V}_3 = V_3 + \rho_{13}(\lambda_1 - \hat{\lambda}_1) + \rho_{23}(\lambda_2 - \hat{\lambda}_2)$. Eq. (12) is then estimated by OLS of y_3 on x_3 , $\hat{\lambda}_1$ and $\hat{\lambda}_2$ for all individuals who has responded and was employed. Consistency of the coefficients in eq. (12) follows from consistency from estimating the λ_i s.

The λ_i s will be estimated in two different ways. In the first approach (denoted procedure Two-probit above) it is assumed, perhaps somewhat unrealistically, that $\rho_{12} = 0$. Therefore it is straightforward to use two probit models to estimate eqs. (1) and (2) separately. From these we obtain $\hat{\delta}_i$ and $\hat{\lambda}_i$ s. These $\hat{\lambda}_i$ s are then used as covariates in eq. (12). Since we assume that $\rho_{12} = 0$ we also assume that there is no problem of sample selection in estimating the employment status using only responders.

In the second approach (Bi-probit) joint normality of ϵ_1 and ϵ_2 is assumed and $\rho_{12} \neq 0$ is allowed. δ_1 , δ_2 and ρ_{12} are estimated jointly using a censored probit model. Thereafter these estimates are used once again to construct the $\hat{\lambda}_i$ s. According to Hamermesh and Donald (2004) this is a more efficient and general approach compared with Two-probit since it includes more information. The nonrespondents are not treated as missing values as in procedure Two-probit but as censored and included in the formal analysis. The censoring problem is caused by the fact that we only observe whether the individuals will be employed if they are responders.

The conventional standard error estimates from estimating eq. (12) will be inconsistent when selectivity is present, i.e., when $\gamma_i \neq 0$, since the standard errors will be heteroskedastic.¹⁵

¹⁵Another shortcoming with the heckit approach is that the method does not impose the constraint that the estimated correlation coefficient constraint is $|\rho| \leq 1$. It turns out that sometimes $|\rho| > 1$ in empirical

Wooldridge (2002) states that it is not enough to use a robust technique to estimate the standard errors. Hamermesh and Donald (2004) state that under the assumption that the correction terms have zero coefficients it is enough using a robust technique for most of the heckit procedures. One must, as we do in this paper, for Bi-probit also recognize that $\hat{\delta}_i$ are an estimator of δ_i using a joint distribution when obtaining the correct standard errors otherwise the standard errors will be inconsistent.

Also estimated, as mentioned above, are heckit procedures Heckit-ref (the reference) and Heckit-res that only adjust for the selection of entering the labor market, i.e., only including a $\hat{\lambda}_2$ term in eq. (12) and thereby ignoring the response decision. Heckit-res use the restrictive sample of 317 observations for the earnings equation and 819 observations for the labor market decision.

4.3.1 Identification and variable specification

One of the major problems with selectivity models is identification; there are two separate ways these models could be identified, (i) from the nonlinearity in the inverse mills ratios and (ii) from exclusion restrictions in the x -vectors. We will use both approaches since it is well known that when using the mills ratios identification is weak, this is because they are linear over a wide range of its argument, see e.g. Vella (1998) or Wooldridge (2002).

Using the exclusion restriction the first problem is that the identifying variables must be obtainable also for the nonrespondents. Those respondents that select themselves into the labor market are usually viewed as a less serious problem – the usual identifier is the presence of young children. It has been argued that the presence of young children determines labor force participation but does not affect earnings conditional on participation, meaning that the presence of young children should be included in x_2 but not in x_3 (Hamermesh and Donald (2004)). It is hard to see the use of this result in this setting since the population is above 50 years of age and the presence of young children is most likely low. Instead we use the average unemployment rate during 2002 at the municipality level. The intuition behind this is straightforward; its intent is to capture municipality differences in the labor market. These differences are however unlikely to affect neither the probability of response nor the monthly earnings of the working population. Other natural candidates in the labor force participation equation are besides the unemployment rate also age, geographical residence and immigration status.

The other case of identification is that of including a variable or set of variables that identify why a randomly chosen individual decides to respond to the survey, i.e., some measure or measures that is included in x_1 but not in x_2 or x_3 . We believe that variables capturing the presence of any competing activities should work as an identification variable. Hence we include the ownership of a leisure home in x_1 . See Johansson and Klevmarken (2006) for a general discussion of how eq. (1) should be specified. The general finding within the survey literature is that the response rate is usually lower among metropolitan residents, single persons

and older persons. Hence, these are all natural candidates for inclusion into x_1 .

Consequently there are some common features between the auxiliary information included in the auxiliary vector and the vector deciding response for the heckit procedures. Two of the three variables included in the auxiliary vector are also used in the heckit procedures. These are demographic residence, i.e., big city, small city or rural area and being married or not. The third not included is yearly earnings who is stochastically determined by hourly earnings and thereby excluded.

4.4 The full-information maximum likelihood approach

The most efficient approach, within the selection family, is to estimate a FIML model where ϵ_j ($j = 1, 2, 3$) are multivariate normally distributed with zero mean and covariance matrix Σ , the covariance matrix has the following structure

$$\Sigma = \begin{bmatrix} 1 & \tilde{\rho}_{12} & \tilde{\rho}_{13} \\ \tilde{\rho}_{12} & 1 & \tilde{\rho}_{23} \\ \tilde{\rho}_{13} & \tilde{\rho}_{23} & 1 \end{bmatrix}.$$

A FIML approach is more general in the way that the shortcomings mentioned above could be solved.¹⁶ The only assumption that has to be imposed is that the error terms are normally distributed. And as long as the underlying models are correctly specified FIML estimates are recognized as the most efficient among all estimators. In the FIML function all equations and all parameters are estimated simultaneously; therefore it is more convenient to see eqs. (1) to (3), within the FIML setting, as a simultaneous equation system.

When deriving the FIML function one has to take into account that there are three possible outcomes according to the selection rules in eqs. (4) to (6), these outcomes are; (i) not responding and consequently no information about labor force participation ($y_1^* = 0$). (ii) responding but no labor force participation ($y_1^* = 1$ and $y_2^* = 0$). And (iii), responding and participating in the labor force ($y_1^* = 1$ and $y_2^* = 1$). These three outcomes result in three different probabilities. They are

$$\begin{aligned} P_1 &= P(y_1 = 0) = P(\epsilon_1 \leq -x_1' \delta_1) \\ &= F(-x_1' \delta_1) \end{aligned} \tag{13}$$

$$\begin{aligned} P_3 &= P(y_1 = 1, y_2 = 0) = P(\epsilon_1 > -x_1' \delta_1, \epsilon_2 \leq -x_2' \delta_2) \\ &= G(x_1' \delta_1, -x_2' \delta_2; -\tilde{\rho}_{12}) \end{aligned} \tag{14}$$

$$\begin{aligned} P_4 &= P(y_1 = 1, y_2 = 1) = P(\epsilon_1 > -x_1' \delta_1, \epsilon_2 > -x_2' \delta_2) \\ &= G(x_1' \delta_1, x_2' \delta_2; \tilde{\rho}_{12}), \end{aligned} \tag{15}$$

where $F(\cdot)$ denotes the standard univariate normal distribution function and $G(\cdot, \cdot; \cdot)$ denotes

¹⁶Both the Heckit approach and the FIML approach assumes normality. There is no difference in robustness of estimates from these models but FIML estimates are more efficient.

the bivariate normal distribution function with correlation coefficient $\tilde{\rho}_{12}$. Eqs. (13) to (15) describe all possible outcomes for the data used; hence these sub-samples are all included in the FIML function.

When there is complete coverage of y_3^* the conditional normal distribution could be written as

$$f(y_3^* | y_1^* = 1, y_2^* = 1) = \frac{1}{P_4} \int_{-x_2'\delta_2}^{\infty} \int_{-x_1'\delta_1}^{\infty} \frac{1}{\sigma_3} h\left(\epsilon_1, \epsilon_2, \frac{y_3^* - x_3'\delta_3}{\sigma_3}\right) d\epsilon_1 d\epsilon_2, \quad (16)$$

where σ_3 once again is an unknown scale parameter and $h(., ., .)$ denotes the standard normal trivariate density.¹⁷

Letting \prod indexed by S_j denote the product operator for the observations in sub-sample j , the likelihood function for the first two sub-samples defined in eqs. (13) to (14) above and the trivariate normal distribution in eq. (16) is then simply

$$\begin{aligned} \mathcal{L} = & \prod_{S_1} F(-x_1'\delta_1) \times \prod_{S_2} G(x_1'\delta_1, -x_2'\delta_2; -\tilde{\rho}_{12}) \\ & \times \prod_{S_3} \int_{-x_2'\delta_2}^{\infty} \int_{-x_1'\delta_1}^{\infty} \frac{1}{\sigma_3} h\left(\epsilon_1, \epsilon_2, \frac{y_3^* - x_3'\delta_3}{\sigma_3}\right) d\epsilon_1 d\epsilon_2. \end{aligned} \quad (17)$$

It can be shown, see e.g. Hayashi (2000), that the FIML estimator $(\hat{\delta}_j, \hat{\Sigma})$ which maximizes eq. (17), is consistent and asymptotically normal. This likelihood function could also be found in Tunalı (1986). The corresponding log-likelihood function ℓ is then

$$\begin{aligned} \ell = & \sum_{S_1} \ln F(-x_1'\delta_1) + \sum_{S_2} \ln G(x_1'\delta_1, -x_2'\delta_2; -\tilde{\rho}_{12}) \\ & + \sum_{S_3} \ln \int_{-x_2'\delta_2}^{\infty} \int_{-x_1'\delta_1}^{\infty} \frac{1}{\sigma_3} h\left(\epsilon_1, \epsilon_2, \frac{y_3^* - x_3'\delta_3}{\sigma_3}\right) d\epsilon_1 d\epsilon_2, \end{aligned} \quad (18)$$

and this is the expression to be maximized.¹⁸

4.4.1 Variable specification

The variables included when maximizing eq. (18) are the same as when estimating the heckit models, except for the mills ratios. I.e., x_1 contain the same variables as eq. (1) when estimating response. x_2 includes the same variables as eq. (2) when estimating labor force participation status. And x_3 , the earnings equation, uses the same variables as the other earnings equations estimated.

A comment on the data is also in place. From Table 1 it is clear that the first term of eq. (18) includes 315 nonrespondents, the second term includes 187 respondents with zero earnings and the last term includes 317 respondents with positive earnings, resulting in a total of 819

¹⁷The trivariate density function, $h(., ., .)$, is specified as $h(\epsilon_1, \epsilon_2, \frac{y_3 - x_3'\delta_3}{\sigma_3}) = (2\pi)^{-\frac{3}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \epsilon' \Sigma^{-1} \epsilon\right\}$ where ϵ is the joint (1×3) vector of ϵ_j .

¹⁸The CML procedure in GAUSS was used as a maximization routine for the likelihood function in eq. (18). We set the upper integration limit to 99 for both integrals, a figure viewed to be sufficiently large.

observations.

5 Results

In Table 5 in Appendix the parameter estimates and their corresponding standard errors from estimating eq. (2) for procedures Cali-heckit, Heckit-ref and Heckit-res are presented, and in Table 6 the same statistics are presented from estimating eqs. (1) and (2) for the heckit models and the first two terms of eq. (17) for the FIML model.¹⁹ All these coefficients have the expected signs although few are statistically significant – not true for the reference procedure where all coefficients are statistically significant. E.g. being an immigrant is negative both for the response and for the labor force decision, being on welfare benefits reduces the response probability whereas being married increases the same. A negative result is that the identifying variable included in the labor force participation equation (the unemployment rate) and in the response equation (leisure home) is insignificant for all procedures – except for the unemployment rate in procedure Heckit-ref. Hence identification comes mainly from the nonlinearities of the mills ratios. In Table 5 there are large discrepancies between Heckit-ref and procedures Cali-heckit and Heckit-res. Thinking of Heckit-ref:s estimates as a better representation of the correct estimates it is clear that the other two procedures do not perform well in predicting the labor market decision. For procedures Two-probit, Bi-probit and FIML in Table 6 there are minor differences in point estimates, and in line with theory the FIML estimates are more efficient since the standard errors are lower than for the heckit estimates.

The results from the earnings estimations are found in Table 5 below. The first two procedures are the Heckit-ref and the Heckit-res, next are the Cali and Cali-heckit that use the calibration approach and a mixture between calibration and heckit, respectively. The next two procedures use the heckit approach, the only difference is how the λ_i s are estimated. The final procedure is the estimates generated from the last term of the FIML.

The best way of testing for selectivity bias is, according to Vella (1998), to t -test the null hypothesis that $\gamma_i = 0$, this tells us that we can reject the null for the λ_1 estimates but not for the λ_2 estimates. Consequently there is a nonrandom selection regarding who chooses to respond to the survey, but that is not true for the labor market decision.²⁰ These results are in line with what was noted above and with previous studies, this also holds for the signs of the coefficients, see Hamermesh and Donald (2004). The negative sign of λ_1 suggests that, based on unobservable characteristics, individuals who are believed to be more likely to respond earn less than otherwise observed identical individuals.

The correlation ρ_{12} estimate is significant, and negative, on a 5 percent level. The implication of that is that omitted variables have a joint, but opposite, impact on nonresponse and nonparticipation in the labor market (procedure Bi-probit). For the FIML we had to impose

¹⁹The standard errors are robust for all procedures except for the FIML, for the FIML the standard errors are computed from the Hessian.

²⁰When adjusting the whole LINDA sample for the labor market decision the null hypothesis for λ_2 is rejected.

the identification restriction that $\tilde{\rho}_{12} = \tilde{\rho}_{13}$.²¹ Both these correlations are significant at a one percent level and negative – indicating that there, also for this procedure, are omitted variables that influence these equations in opposite directions.

It is clear that the Heckit-ref generate estimates in line with what are to be expected. Females have 19.3 percent lower hourly earnings than males. Earnings are higher for individuals living in larger cities than they are for individuals living in urban and even higher than for individuals living in rural areas. One extra year of schooling increases earnings with 6.3 percent and experience increases earnings at a decreasing rate. All estimates are significant at a one percent significance level giving support to view that this procedure is equivalent to a population analysis. Comparing the point estimates from Heckit-ref and Heckit-res, where the only difference is the number of observations used, tells us that all coefficients are larger for Heckit-res except for the experience coefficients and the constant. The main objection is that Heckit-res is extremely bad in predicting the experience coefficients compared with the other procedures.

The calibration approach (Cali and Cali-heckit) generates estimates with the same sign as the uncorrected model, although the experience coefficients are insignificant. In general the estimates have larger absolute values for the calibration procedures compared with Heckit-ref. The main difference is that Cali-heckit (that includes λ_2) is much better in predicting the experience estimates than Cali since these estimates are closer to the estimates from Heckit-ref. Procedure Cali-heckit is comparable with procedures Heckit-ref and Two-probit, Bi-probit and FIML. Noticing that Bi-probit is better than Two-probit we henceforth focus on comparing Cali-heckit, Bi-probit and FIML with Heckit-ref. The result that Bi-probit is preferred over Two-probit is as expected, since Bi-probit uses more information when estimating λ_2 .²²

Just by comparing the point estimates it is clear that estimates from the FIML procedure is very close to Heckit-ref for all coefficients, except for the intercept. It often differs on the second or third decimal. But, contrary to what is expected, the earnings estimates are very inefficient resulting in only two significant coefficients. An explanation for this results could be the imposed identification restriction that $\tilde{\rho}_{12} = \tilde{\rho}_{13}$ or the small sample size.

By comparing the calibration (Cali-heckit) and heckit (Bi-probit) procedures we find that the first is better in predicting the gender and the geographical coefficients, whereas it is less accurate in predicting the experience coefficients compared with Bi-probit. The explanation for this result is that the geographical variables are included in the auxiliary vector resulting in more accurate estimates. By comparing standard errors one comes to the conclusion that Bi-probit is marginally more efficient than Cali-heckit.

So, which of the three recommended correction procedures is best? It is difficult to single out one approach that performs better than the others in such a way so that the estimates are closer to the true estimates and also efficient. The point estimates from the FIML procedure

²¹Without this restriction the model had problem of converging to a global maximum. $\tilde{\rho}_{12}$ became almost one and all correlations got very high gradient values.

²²These results are in line with Hamermesh and Donald (2004), who also found that there were only minor differences between these two Heckit approaches, although our results suggests that there are larger differences in R^2 values.

are in line with the estimates from the population, but they are very inefficient. The calibration and heckit estimates are more efficient but not as close to the reference. When comparing these correction procedures it is clear that the calibration procedure is better in predicting gender and geographic estimates whereas the heckit procedure is better in predicting the returns to schooling and experience. If one has to recommend one procedure it has to be Cali-heckit, i.e., when calibration is used to adjust for the nonrandom nonresponse and calibration plus heckit is used for the nonrandom labor market decision. Overall it has the best properties – it is efficient, although not most efficient. The point estimates are much more precise for some of the estimates, and when they are less precise the difference to Bi-probit is not that large.

In Table 5 almost all of the estimates are larger in absolute values for the procedures using the restricted sample compared with the reference procedure. For example, the heckit approach – probably the most prevalent procedure used within applied economics in this framework, generates a negative effect of being a female 50 percent larger compared with the population estimate. And the difference for the rural estimate is even larger – instead of a negative effect of 11 percent the heckit estimates are more than three times as large. These results suggest that statistical inference from these procedures is problematic since it does not reflect the true relationship.

Table 3: Earnings estimates from five different procedures

Variables	Heckit-ref	Heckit-res	Cali	Cali-heckit	Two-probit	Bi-probit	FiML
If female	-.193*** (.004)	-.222*** (.051)	-.248*** (.052)	-.237*** (.052)	-.337*** (.056)	-.308*** (.053)	-.206** (.086)
Urban ^A	-.070*** (.004)	-.133** (.055)	-.128** (.053)	-.132** (.053)	-.195*** (.056)	-.189*** (.056)	-.092 (.097)
Rural ^A	-.117*** (.006)	-.214** (.085)	-.190** (.090)	-.221** (.088)	-.348*** (.089)	-.340*** (.090)	-.110 (.136)
Schooling	.063*** (.001)	.074*** (.013)	.072*** (.013)	.077*** (.012)	.058*** (.012)	.060*** (.013)	.061 (.375)
Experience	.027*** (.006)	.001 (.093)	.050 (.077)	.024 (.089)	.036 (.089)	.024 (.089)	.030 (.152)
(Experience) ²	-.037*** (.007)	.006 (.123)	-.066 (.100)	-.024 (.118)	-.058 (.118)	-.040 (.119)	-.033 (.203)
Constant	3.772*** (.119)	4.327** (1.806)	3.372** (1.488)	3.846** (1.692)	4.968*** (1.769)	5.090*** (1.792)	5.422** (2.585)
Calibrated	No	No	Yes	Yes	No	No	No
$\hat{\lambda}$							
λ_1					-1.476*** (.406)	-1.406*** (.400)	
$\hat{\lambda}$							
λ_2	.123*** (.012)	-.237 (.196)		-.241 (.154)	-.018 (.198)	-.126 (.296)	
$\hat{\rho}_{12}$						-.738** (.361)	
$\hat{\rho}_{12}$							-.359*** (.030)
$\hat{\rho}_{13}$							-.359*** (.030)
$\hat{\rho}_{23}$							-.096 (.075)
R^2	.222	.171	.178	.199	.239	.247	
Mean ℓ							-1.596
Observations	35,046	317	317	317	317	317	317

Note. standard errors are presented within parenthesis, * significant at 10%; ** significant at 5%; *** significant at 1%. ^AThe reference category for demographic area is individuals living in major cities.

6 Concluding remarks

This paper focuses on how to estimate earnings equations using survey data that has a nonignorable response mechanism. The procedures used are a calibration approach that re-weights each observation using auxiliary information, different types of the heckit model, and finally, a full-information maximum likelihood (FIML) model. These procedures adjust both for a nonrandom response mechanism and a nonrandom labor force decision. This is enabled since we have access to extraordinary data, population totals for our sample frame, used in the calibration procedure and register information both for the respondents as for the nonrespondents used in the heckit and FIML procedures.

The data used for these procedures is from a Swedish survey conducted in 2002 with a final sample of 819 of which 504 responds and 317 reports earnings. As an earnings measure for these individuals Swedish register data is used. The procedures are then validated against the true parameters obtained from using 117,334 individuals when estimating the nonrandom labor force decision from where 35,046 individuals are used for estimating the earnings equation. These reference regression equations uses the same type of Swedish registers data as when estimating the correction procedures – which enhances comparability.

Although all of the procedures produce estimates with the same, and expected, sign as the estimates from reference model the results from this paper are that there are fundamental differences between the correction procedures. The calibration procedure produce gender and geographic estimates more in line with the reference whereas the heckit estimates are better in estimating the returns to schooling and experience effects. The point estimates from FIML are very close to the estimates from the reference but are very inefficient for some of the estimated coefficients. Therefore it is not obvious that one correction procedure performs better than all others. But overall, considering preciseness and efficiency, a joint procedure using both calibration and heckit is recommended.

This paper further shows that using the correction procedures generate earnings estimates that are larger compared with the reference – true for almost all estimates. For instance, the negative effect of being female is overestimated by almost 50 percent for two of the heckit procedures.

References

- BETHLEHEN, J. G. (2002): “Weighting Nonresponse Adjustments Based on Auxiliary Information,” in *Survey Nonresponse*, ed. by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little. John Wiley & Sons, Inc, New York.
- BROSE, P., AND N. A. KLEVMARKEN (1993): “Modeling Response in a Panel Survey,” Working papers of the European Scientific Network on Household Panel Studies. Paper 81, Colchester: University of Essex. (Presented at the 8th Session of the International Statistical Institute, Cairo).
- COPAS, A. J., AND V. T. FAREWELL (1998): “Dealing with Non-Ignorable Non-Response by Using an ‘Enthusiasm-to-Respond’ Variable,” *Journal of the Royal Statistical Society*, 161(3), 385–396.
- DAS, M., W. K. NEWHEY, AND F. VELLA (2003): “Nonparametric Estimation of Sample Selection Models,” *Review of Economic Studies*, 70(1), 33–58.
- DEVILLE, J. C., AND C.-E. SÄRNDAL (1992): “Calibration Estimators in Survey Sampling,” *Journal of the American Statistical Association*, 87(418), 376–382.
- EDIN, P.-A., AND P. FREDRIKSSON (2000): “LINDA – Longitudinal Individual Data for Sweden,” Working Paper 2000:19, Department of Economics, Uppsala university.
- GREENLESS, J. S., W. S. REECE, AND K. D. ZIESCHANG (1982): “Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed,” *Journal of the American Statistical Association*, 77(378), 251–261.
- HAMERMESH, D. S., AND S. G. DONALD (2004): “The Effect of College Curriculum on Earnings: Accounting for Non-Ignorable Non-Response Bias,” Working Paper 10809, NBER.
- HAYASHI, F. (2000): *Econometrics*. Princeton University Press, Princeton.
- HECKMAN, J. J. (1979): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47(1), 153–161.
- JOHANSSON, F., AND N. A. KLEVMARKEN (2006): “Explaining the Size and Nature of Response in a Survey on Health Status and Economic Standard,” Working Paper 2006:2, Department of Economics, Uppsala University.
- KALTON, G., AND D. KASPRZYK (1986): “The Treatment of Missing Data,” *Survey Methodology*, 12(1), 1–16.
- LITTLE, R. J. A. (1982): “Models of Nonresponse in Sample Surveys,” *Journal of the American Statistical Association*, 77(378), 237–250.
- LITTLE, R. J. A., AND D. B. RUBIN (1987): *Statistical Analysis with Missing Data*. Wiley, New York.
- LUNDSTRÖM, S., AND C.-E. SÄRNDAL (2002): *Estimation in the Presence of Nonresponse and Frame Imperfections*. Statistics Sweden, Örebro.
- MINCER, J. (1974): *Schooling, Experience and Earnings*. Columbia University Press for NBER, New York.

- POWELL, J. L. (1994): “Estimation of Semiparametric Models,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden. Elsevier Science B.V, Amsterdam.
- ROSENBAUM, P. R., AND D. B. RUBIN (1984): “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score,” *Journal of the American Statistical Association*, 79(387), 516–524.
- SÄRNDAL, C.-E., AND S. LUNDSTRÖM (2005): *Estimation in Surveys with Nonresponse*. John Wiley & Sons, Inc, West Sussex, England.
- SÄRNDAL, C.-E., B. SWENSSON, AND J. WRETMAN (1992): *Model Assisted Survey Sampling*. Springer-Verlag, New York, New York.
- THORBURN, D. (2005): “Synpunkter efter läsning av boken: Estimation in Surveys with Nonresponse, Särndal and Lundström,” Department of Statistics, Stockholm university.
- TUNALI, I. (1986): “A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Re-Migration,” in *Research in Labor Economics*, ed. by R. G. Ehrenberg. Jai Press Inc, Greenwich, Connecticut.
- VELLA, F. (1998): “Estimating Models with Sample Selection Bias: A Survey,” *Journal of Human Resources*, 33(1), 127–169.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT press, Cambridge, Massachusetts.

Appendix

Table 4: Descriptive statistics for the sample used to estimate response and labor force status

Variables	Mean	S.E.	<i>N</i>
Age	56.50	4.186	819
If female	.515	.500	819
Education			
Compulsory	.278	.448	819
High school	.471	.499	819
University	.251	.433	819
Demographic area			
Major city	.335	.473	819
Urban	.536	.499	819
Rural	.128	.334	819
If Immigrant	.045	.209	819
If Married	.609	.488	819
Hospital stay	.074	.263	819
Wage	16,007	13,312	819
If Sickness benefit	.193	.394	819
If welfare benefits	.028	.165	819
Leisure home	.170	.375	819
Unemployed	.032	.009	819

Table 5: Probit estimates of labor force participation

Variables	Cali-hekit	Hekit-ref	Hekit-res
Age1 (55-60) ^A	.006 (.140)	.073*** (.008)	.020 (.132)
Age2 (60-65) ^A	-.840*** (.168)	-.248*** (.011)	-.788*** (.160)
Urban ^C	.049 (.142)	.027*** (.008)	.038 (.134)
Rural ^C	.023 (.207)	-.058*** (.014)	.064 (.196)
If immigrant	-.915** (.381)	-.229*** (.023)	-1.093*** (.357)
If married	-.070 (.125)	-.414*** (.008)	.009 (.122)
Sick income	-.196 (.152)	-.030*** (.010)	-.281* (.144)
Unemployment	-.072 (.662)	-.688*** (.037)	-.379 (.628)
Constant	.575** (.248)	-.181*** (.015)	.667*** (.237)
Observations	504	117,334	819

Note. standard errors are presented within parenthesis, * significant at 10%; ** significant at 5%; *** significant at 1%. ^AThe reference for age is individuals between 50 and 55. ^BThe reference category for demographic area is individuals living in major cities.

Table 6: Probit and FIML estimates of response and labor force participation

Variables	Two-probit	Bi-probit	FIML
		Response	
Age1 (55-60) ^A	.045 (.101)	.043 (.101)	.050 (.049)
Age2 (60-65) ^A	.065 (.124)	.067 (.124)	.058 (.070)
If female	.118 (.091)	.088 (.088)	.100** (.044)
Urban ^C	.082 (.100)	.082 (.100)	.083* (.048)
Rural ^C	.205 (.151)	.207 (.152)	.195*** (.072)
Immigrant	-.488** (.218)	-.490** (.216)	-.465*** (.148)
If married	.193** (.093)	.197** (.094)	.242*** (.045)
If hospital stay	-.177 (.175)	-.243 (.163)	-.224** (.100)
If welfare benefits	-1.069*** (.320)	-1.002*** (.341)	-.944*** (.202)
Leisure home	.094 (.123)	.117 (.118)	.094* (.056)
Constant	.066 (.121)	.078 (.118)	.131** (.057)
Observations	819	819	315
		Labor force participation	
Age1 (55-60) ^A	.020 (.132)	.015 (.122)	.043 (.081)
Age2 (60-65) ^A	-.788*** (.160)	-.694*** (.181)	-.688*** (.094)
Urban ^C	.038 (.134)	.007 (.120)	.055 (.094)
Rural ^C	.064 (.196)	-.012 (.177)	.057 (.122)
If immigrant	-1.093*** (.357)	-.654 (.466)	-.959*** (.194)
If married	.009 (.122)	.077 (.111)	.034 (.068)
Sick income	.282* (.144)	-.219 (.147)	-.149 (.073)
Unemployment	-.379 (.628)	-.343 (.533)	-.200 (.536)
Constant	.667*** (.237)	1.076*** (.237)	.835 *** (.168)
Observations	504	819	187

Note. robust standard errors are presented within parenthesis, * significant at 10%; ** significant at 5%; *** significant at 1%. ^AThe reference for age is individuals between 50 and 55. ^BThe reference category for demographic area is individuals living in major cities.

WORKING PAPERS*

Editor: Nils Gottfries

- 2006:23 Martin Ågren, Does Oil Price Uncertainty Transmit to Stock Markets? 29pp.
- 2006:24 Martin Ågren, Prospect Theory and Higher Moments. 31pp.
- 2006:25 Matz Dahlberg, Eva Mörk, Jørn Rattsø and Hanna Ågren, Using a discontinuous grant rule to identify the effect of grants on local taxes and spending. 26pp.
- 2006:26 Jukka Pirttilä and Håkan Selin, How Successful is the Dual Income Tax? Evidence from the Finnish Tax Reform of 1993. 40pp.
- 2006:27 Henrik Jordahl and Che-Yuan Liang, Merged Municipalities, Higher Debt: On Free-riding and the Common Pool Problem in Politics. 34pp.
- 2006:28 Per Engström, Ann-Sofie Kolm and Che-Yuan Liang, Maternal Addiction to Parental Leave. 18pp.
- 2006:29 Jonas Björnerstedt and Andreas Westermarck, Delay in Bargaining with Externalities. 27pp.
- 2006:30 Pär Österholm, Incorporating Judgement in Fan Charts. 36pp.
- 2006:31 Mikael Carlsson and Andreas Westermarck, Monetary Policy and Staggered Wage Bargaining when Prices are Sticky. 26pp.
- 2007:1 Mikael Elinder, Local Economies and General Elections. 26pp.
- 2007:2 Ouarda Merrouche, The Long Term Impact of French Settlement on Education in Algeria. 19pp.
- 2007:3 Ouarda Merrouche, The Long Term Effect of Education Spending Decentralization on Human Capital in Spain. 15pp.
- 2007:4 Erik Post, Macroeconomic imbalances and exchange rate regime shifts. 38pp.
- 2007:5 Christian Andersson, Teacher density and student achievement in Swedish compulsory schools. 31pp.
- 2007:6 Thomas Aronsson, Sören Blomquist and Luca Micheletto, Where Should the Elderly Live and Who Should Pay for their Care? A Study in Demographics and Geographical Economics. 22pp.
- 2007:7 Sören Blomquist and Vidar Christiansen, Public Provision of Private Goods and Nondistortionary Marginal Tax Rates. 17pp.

* A list of papers in this series from earlier years will be sent on request by the department.

- 2007:8 Marcus Eliason and Henry Ohlsson, Living to Save Taxes. 13pp.
- 2007:9 Åsa Ahlin and Eva Mörk, Effects of decentralization on school resources: Sweden 1989-2002. 31pp.
- 2007:10 Henry Ohlsson, The equal division puzzle – empirical evidence on intergenerational transfers in Sweden. 20pp.
- 2007:11 Daniel Hallberg and Mårten Lagergren, Moving in and out of public geriatric care in Sweden. 26pp.
- 2007:12 Per Engström, Wage Formation and Redistribution. 22pp.
- 2007:13 Henry Ohlsson, Tax avoidance – a natural experiment. 21pp.
- 2007:14 David Kjellberg and Erik Post, A Critical Look at Measures of Macroeconomic Uncertainty. 27pp.
- 2007:15 Mikael Carlsson and Andreas Westermarck, Optimal Monetary Policy under Downward Nominal Wage Rigidity. 52pp.
- 2007:16 Robin Douhan and Anders Nordberg, Is the elephant stepping on its trunk? The problem of India's unbalanced growth. 33pp.
- 2007:17 Annika Alexius and Bertil Holmlund, Monetary Policy and Swedish Unemployment Fluctuations. 27pp.
- 2007:18 Meredith Beechey and Pär Österholm, The Rise and Fall of U.S. Inflation Persistence. 23pp.
- 2007:19 Henry Ohlsson and Donald Storrie, Long term effects of public policy for displaced workers in Sweden – shipyard workers in the West and miners in the North. 26pp.
- 2007:20 Niklas Bengtsson, How responsive is body weight to transitory income changes? Evidence from rural Tanzania. 38pp.
- 2007:21 Karin Edmark, Strategic Competition in Swedish Local Spending on Childcare, Schooling and Care for the Elderly. 38pp.
- 2007:22 Fredrik Johansson, How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML? 25pp.

See also working papers published by the Office of Labour Market Policy Evaluation
<http://www.ifau.se/>