Stat Papers (2015) 56:829–847
DOI 10.1007/s00362-014-0610-x

REGULAR ARTICLE

Uncertainty intervals for regression parameters with non-ignorable missingness in the outcome

Minna Genbäck · Elena Stanghellini · Xavier de Luna

Received: 6 September 2013 / Revised: 12 June 2014 / Published online: 17 July 2014

Abstract When estimating regression models with missing outcomes, scientists usually have to rely either on a missing at random assumption (missing mechanism is independent from the outcome given the observed variables) or on exclusion restrictions (some of the covariates affecting the missingness mechanism do not affect the outcome). Both these hypotheses are controversial in applications since they are typically not testable from the data. The alternative, which we pursue here, is to derive identification sets (instead of point identification) for the parameters of interest when allowing for a missing not at random mechanism. The non-ignorability of this mechanism is quantified with a parameter. When the latter can be bounded with a priori information, a bounded identification set follows. Our approach allows the outcome to be continuous and unbounded and relax distributional assumptions. Estimation of the identification sets can be performed via ordinary least squares and sampling variability can be incorporated yielding uncertainty intervals achieving a coverage of at least $(1-\alpha)$ probability. Our work is motivated by a study on predictors of body mass index (BMI) change in middle age men allowing us to identify possible predictors of BMI change even when assuming little on the missing mechanism.

Keywords Heckman model · Informative dropout · Selection models · Sensitivity analysis · Set identification · Two stage least squares

M. Genbäck · X. de Luna
Department of Statistics, Umeå School of Business and Economics, Umeå University, Umeå, Sweden

E. Stanghellini (✉)
Department of Economics, Università di Perugia, Perugia, Italy
e-mail: elena.stanghellini@stat.unipg.it

1 Introduction

In this paper we introduce inferential procedures, based on identification sets, for regression parameters in situations where a continuous outcome (response in a linear regression model) is not observed for all individuals. A vast part of the literature on missing outcome deals with situations where the missingness mechanism is independent of the outcome conditionally (or not) on observed covariates, called missing (completely) at random mechanism, see Little and Rubin (2002). In this case, parameters are identified and inference can be performed with standard inferential procedures. When this assumption does not hold (i.e. the missingness mechanism is said to be non-ignorable), several contributions are concerned with introducing other restrictions to obtain identification, such as monotonicity (Manski 2003, Chap. 8) or conditional independence restrictions (e.g. pattern mixture models, Daniels and Hogan 2008; Little 2009).

An alternative, which we pursue here, is to determine a region on the parameter space, that we call identification set, that contains all parameters identified under plausible missing data mechanisms, and to propose inferential procedures accordingly. In this sense, this contribution is in line with Vansteelandt and Goetghebeur (2001), Manski (2003), Imbens and Manski (2004), Vansteelandt et al. (2006) and Horowitz and Manski (2006). Results available in this literature on set identification with non-ignorable nonresponse require situations where the outcome is bounded.

In this paper we focus on set identification of regression parameters when the unbounded outcome is continuous and the missing mechanism is non-ignorable. We do that in a framework where the outcome (continuous valued) and missingness indicator (binary variable) are regressed parametrically against a set of covariates, yielding an outcome equation and a selection (missingness mechanism) equation respectively. The sets depend on the parameter ρ, the correlation between the residuals of the two equations. We show that the identification set can be bounded when only mild restrictions are imposed on the missing data model.

We avoid making strong distributional assumptions, initially focusing on a probit regression model for the selection equation but later relaxing to a more general class. Notice that assuming a probit selection equation allows for identification of the outcome equation parameters and estimation of the parameters can be performed via either ML or two stage least square (TSLS), see Heckman (1979). The first procedure relies on the assumption of joint normality of the error terms and is very sensitive to misspecification (Olsen 1982; Wooldridge 2003, p. 566), thus TSLS has become widely used. However, this method too suffers of serious finite sample instability due to collinearity. That is usually addressed by using exclusion restriction assumptions whereby some covariates excluded in the outcome equation are assumed to predict the missingness mechanism (Little 1985). However, it is well known (Puhani 2000) that results can be sensitive to the choice of exclusion restrictions since different assumptions lead to different conclusions on the parameters of interest. We provide further illustration of this issue with a follow up study on body mass index (BMI). By allowing for set identification, our approach avoids the use of restriction assumptions in studies where no strong theory is available to justify them. Furthermore, the theory applies also to situations outside normality.

When only sets of possible values are identified, Vansteelandt et al. (2006) have provided an inferential framework, and for instance they propose to combine the estimated sets with sampling variation to yield a $(1-\alpha)$ 100 % uncertainty region, which covers the identification set with a probability of at least $(1-\alpha)$. In this paper, we deduce uncertainty intervals for the parameters of interests in our context. Uncertainty intervals are the counterpart of confidence intervals in the case of point identification.

A related stream of the literature has developed methods to assess the sensitivity of the inference to departures from the missing at random assumption; see, e.g., de Luna and Lundin (2014), Little et al. (2012), Andridge and Little (2011), Rosenbaum (2010), Copas and Eguchi (2005), Imbens (2003) and Scharfstein et al. (1999). The uncertainty intervals that we introduce may be used as a tool for sensitivity analysis as we illustrate in our case study. Our approach is in this respect closely related to the one proposed by Copas and Li (1997), as the selection parameter θ in their paper is a transformation of ρ. Copas and Li (1997) build a profile log likelihood for θ in order to carry out a sensitivity analysis. Similar models and methods are used for sensitivity analysis to publication bias in meta analysis (Copas 2013, Henmi et al. 2007).

In Sect. 2 we present a motivating example, a follow up study on individual BMI increase within a ten year interval. We introduce the model, discuss identification and illustrate the instability of the results to different exclusion restrictions. Section 3.1 contains the results on set identification under the probit assumption for the missingness mechanism. The latter assumption is relaxed in Sect. 3.2. In Sect. 3.3 we deduce the uncertainty intervals taking into account sampling variation. The BMI study is presented in detail in Sect. 4, illustrating the results obtained in the paper. Final sample properties are illustrated in a simulation study in Sect. 5, where the data generating mechanisms are chosen to mimic a text book case study on unobserved women wages due to non-participation in the labour market. The paper is concluded in Sect. 6.

2 Motivating studies

We utilize two different studies to motivate the contribution of this paper. The first study is concerned with finding predictors of BMI increase within a ten year interval, between 40 and 50 years of age, see Sect. 4 for more details. The second study estimates a wage offer function for married women and is used as background in the simulation study of Sect. 5. In both cases we have an outcome that is observed only for a selected subsample; in the BMI study selection is due to drop out, where some individuals have no BMI measure ten years after the first measure; and in the wage offer study the selection consists in that wage is observed only for those women participating in the labour force, i.e. women without employment are assumed to have a latent wage offer. In both cases we may use the following model. Let

$$y = \nu_2 + \mathbf{x}^T \boldsymbol{\beta} + \eta_2 \tag{1}$$

be the outcome equation, where the outcome y (BMI change or wage in the above mentioned examples) is observed only for individuals with $z = 1$ (no drop out and labour force participation in the above examples), where this selection is modelled as $z = I(z^* > 0)$ with

Table 1 Results of probit regression (2) for the BMI change case study.

			Stepwise elimination	
	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
Intercept	0.939	0.021	1.090	0.000
Baseline BMI	−0.019	0.001	−0.019	0.001
log(Earnings)	0.019	0.039	0.023	0.010
log(Spouse earnings/earnings)	0.019	0.005	0.022	0.001
Spouse age - age	0.007	0.248		
Number of children	0.106	0.057		
Hospitalization (days)	−0.112	0.090		
Children aged leq 3	−0.092	0.071		
Education > 9 years	0.074	0.361		
Education difference	0.060	0.370		
Parent leave benefits > 0	0.117	0.006	0.103	0.010
Student benefits > 0	−0.023	0.843		
Sick leave benefit > 0	−0.096	0.036	−0.119	0.007
Unemployment benefits > 0	−0.140	0.019	−0.146	0.013
Tobacco use	−0.062	0.002	−0.064	0.001
Positive self-reported health	−0.017	0.733		
Living in urban area	−0.018	0.667		

$$z^* = \nu_1 + \mathbf{x}^T \boldsymbol{\delta} + \eta_1. \tag{2}$$

Let us further assume that $\eta_1 \sim N(0, 1)$, $\mathrm{E}(\eta_2) = 0$ and $\mathrm{Var}(\eta_2) = \sigma_2^2$. Note that η_1 has variance one without loss of generality. We allow for the errors to be correlated (non-ignorable selection) such that $\eta_2 = \rho\sigma_2\eta_1 + \varepsilon$, where ρ is the correlation between η_1 and η_2. The variable ε is independent from η_1, and has zero mean and variance σ_ε^2; we make no further assumptions about its distribution. The parameter of interest is $\boldsymbol{\beta}$. Consistent estimation of $\boldsymbol{\beta}$ can be obtained with a maximum likelihood estimator or a two stage least squares (TSLS) estimator (Heckman 1979; Wooldridge 2003, Sect. 17.4).

Table 1 presents the results of fitting the selection equation (probit regression) for the sample of 4,648 males for which BMI is observed at 40 years of age, of which 1,324 do not show up at the 50 years of age call (selection by drop out). The table displays the covariates available as well as their δ coefficient and corresponding p-values. We notice that seven out of sixteen variables are significant at the five percent level, see Table 1 (first two columns). A backward elimination procedure was used and the final model is also given in Table 1 (last two columns). The subsequent analyses are made by restricting the set of covariates in the probit regression to those which are significant. The outcome equation is then fitted using different estimators and results are displayed in Table 2. Thus, we use ordinary least squares (i.e. assuming missingness is ignorable, results in the first two columns of the table, denoted with OLS), and TSLS without exclusion restrictions (last two columns, denoted with TSLS no ER). We can note here that OLS and TSLS results differ. In fact, letting ρ free, $\boldsymbol{\beta}$ is not well identified. This can be illustrated by considering

Table 2 Results of TSLS with different exclusion restrictions and OLS, the stars indicate that the variable is included in the first stage.

	OLS		TSLS ER3		TSLS ER2		TSLS ER1		TSLS no ER	
	Estimate	p value	Estimate	p value	Estimate	p value	Estimate	p value	Estimate	p value
∗Intercept	7.20	0.00	7.63	0.00	7.63	0.00	8.70	0.00	26.66	0.64
∗Baseline BMI	−0.16	0.00	−0.13	0.00	−0.13	0.00	−0.10	0.04	0.72	0.76
∗log(Earnings)	0.00	0.97	−0.01	0.53	−0.01	0.53	−0.07	0.27	−1.09	0.72
∗log(Spouse earnings/earnings)	0.01	0.55					−0.06	0.29	−1.02	0.72
Spouse age - age	0.02	0.20	0.02	0.22	0.02	0.22	0.02	0.42	0.02	0.96
Number of children	−0.01	0.94	−0.01	0.96	−0.01	0.96	−0.01	0.97	−0.06	0.99
Hospitalization (days)	0.01	0.97	0.01	0.94	0.01	0.94	0.01	0.96	0.04	0.99
Children aged leq 3	0.08	0.52	0.10	0.40	0.10	0.40	0.08	0.67	0.04	0.99
Education > 9 years	−0.24	0.21	−0.27	0.16	−0.27	0.16	−0.24	0.44	−0.21	0.96
Education difference	−0.07	0.66	−0.10	0.55	−0.09	0.55	−0.06	0.80	−0.08	0.98
∗Parent leave benefits > 0	−0.11	0.24	−0.25	0.05	−0.25	0.07	−0.41	0.13	−4.64	0.71
Student benefits > 0	−0.22	0.42	−0.23	0.40	−0.23	0.40	−0.21	0.64	−0.31	0.96
∗Sick leave benefit > 0	0.06	0.57	0.22	0.13	0.21	0.16	0.40	0.19	5.40	0.71
∗Unemployment benefits > 0	−0.21	0.15			−0.01	0.94			6.57	0.72
∗Tobacco use	0.14	0.00	0.23	0.00	0.23	0.00	0.33	0.03	3.03	0.70
Positive self-reported health	−0.42	0.00	−0.42	0.00	−0.42	0.00	−0.42	0.02	−0.41	0.89
Living in urban area	−0.16	0.09	−0.16	0.09	−0.16	0.09	−0.17	0.28	−0.12	0.96
InvMillsRatio			−2.70	0.04	−2.63	0.10	−5.82	0.12	−90.44	0.71

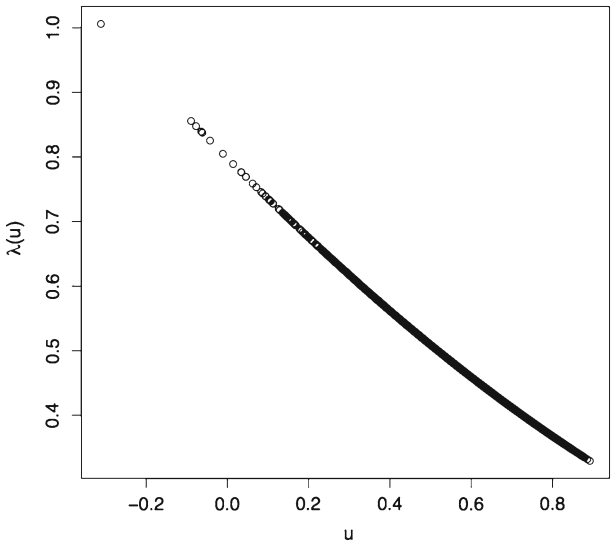

Fig. 1 The inverse Mills' ratio as a function of the linear predictor u

$$E(y \mid \mathbf{x}, z = 1) = \nu_2 + \mathbf{x}^T \boldsymbol{\beta} + \rho\sigma_2\lambda(u), \tag{3}$$

where $u = \mathbf{x}^T \boldsymbol{\delta} + \nu_1$, and $\lambda(u) = \frac{\phi(u)}{\Phi(u)}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are, in order, the standard normal density and cumulative distribution function. The term $\lambda(u)$ is often called inverse Mills' ratio in the literature. It is clear from (3) that OLS will be biased if $\rho \neq 0$. In applications the inverse Mills' ratio is often close to linear in u (Puhani 2000; Jonsson 2012) and this is also the case in our example, see Fig. 1. Since the second stage of TSLS is a regression of y on $\mathbf{x}$ and $\lambda(u)$, this will imply a collinearity problem, generating large standard errors (parameters are non significant), see Table 2 (last two columns). In order to avoid collinearity, TSLS is usually performed with exclusion restrictions on some variables in the outcome equation. Indeed, assuming that some components of $\boldsymbol{\beta}$ are zero while the corresponding components of $\boldsymbol{\delta}$ are not, ensures that the Mills' ratio is not close to be linear in u; see e.g. (Wooldridge 2003, p. 564). However, unless exclusion restrictions are available from scientific theories, such assumptions are controversial.

Table 2 also contains TSLS results based on different exclusion restrictions: in column seven and eight (TSLS ER1) we have excluded one covariate, 'Unemployment benefits', from the outcome equation; in column five and six (TSLS ER 2) we have excluded another 'log(spouse earnings/earnings)'; in column three and four (TSLS ER3) we have excluded both of them. All exclusion restrictions are made on covariates significant in the probit regression but not in the OLS fit. We obtain p values for the coefficient of the inverse Mills' ratio of 71, 12, 10 and 4 %, indicating non-ignorable selection in the latter case only. The results obtained differ most between TSLS without exclusion restrictions (Table 2 last two columns) and the other fits, as expected, due to collinearity. The results also differs between OLS and TSLS with exclusion restrictions, and, most worryingly, between the three fits with different exclusion restriction

assumptions. The most clear example of this is 'Parent leave benefits' which is estimated to -0.11 (p value 24 %) in the OLS fit and to -0.41, -0.25 and -0.25 (p values 13, 7 and 5 %) in the TSLS fits with exclusion restrictions. Another example is 'Sick leave benefits' which is estimated to 0.06 (p value 57 %) in the OLS fit and 0.40, 0.21 and 0.22 (p values 19, 16 and 13 %) in the TSLS fits with exclusion restrictions. A conclusion of this exercise is that unless one has a clear theoretical knowledge on which variables among those affecting selection should be excluded from the outcome equation, results may vary, both in effect size and precision. This can happen irrespective of the inverse Mills' ratio being significant or not and will be even more apparent if we include all variables in the probit regression. Similar findings are in Lennox et al. (2012).

In this paper we avoid the above described problems (collinearity with the inverse Mills' ratio, need of exclusion restrictions, instability of results with respect to exclusion restriction chosen) by proposing identification sets for $\boldsymbol{\beta}$ valid for a certain degree of selection to be specified in advance.

3 Theory

3.1 Model and identification set

We reformulate the model from Sect. 2 in matrix form. Let $\mathbf{y}$ be a N vector with the complete outcome and $\mathbf{X}$ the $(N \times (p+1))$ complete data regression matrix, i.e.

$$\mathbf{X} = \begin{bmatrix} 1 & \mathbf{x}_1^T \\ 1 & \mathbf{x}_2^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_N^T \end{bmatrix}.$$

The model can be written as follows:

$$\mathbf{y} = \mathbf{X} \begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} + \boldsymbol{\eta}_2,$$

the outcome equation, where $\mathbf{y}$ and $\boldsymbol{\eta}_2$ are vectors of dimension N, and

$$\mathbf{z}^* = \mathbf{X} \begin{bmatrix} \nu_1 \\ \boldsymbol{\delta} \end{bmatrix} + \boldsymbol{\eta}_1,$$

the selection equation, where $\mathbf{z}^*$ and $\boldsymbol{\eta}_1$ are vectors of dimension N. As earlier, we assume that all elements of $\boldsymbol{\eta}_1$ are i.i.d. N(0, 1) and all elements of $\boldsymbol{\eta}_2$ are i.i.d. with zero mean and homogenous variance σ_2^2. Also, $\boldsymbol{\eta}_2 = \rho\sigma_2\boldsymbol{\eta}_1 + \boldsymbol{\varepsilon}$, where ρ is the correlation between the corresponding components of $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, and $\boldsymbol{\varepsilon}$ is independent from $\boldsymbol{\eta}_1$

and has elements with zero mean and variance σ_ε^2; we make no further assumptions about its distribution.

Let $\mathbf{y}_s$ be a $n < N$ vector with the observed outcome and $\mathbf{X}_s$ be the corresponding $(n \times (p+1))$ incomplete data regression matrix. Then the OLS estimates of the linear regression coefficients of $\mathbf{y}_s$ on $\mathbf{X}_s$ are:

$$\begin{bmatrix} \hat{\nu_2}_{OLS} \\ \hat{\boldsymbol{\beta}}_{OLS} \end{bmatrix} = (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{y}_s. \tag{4}$$

Note that $\mathrm{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ but $\mathrm{E}(\mathbf{y}_s \mid \mathbf{X}_s) \neq \mathbf{X}_s\boldsymbol{\beta}$ if we have nonignorable missingness. Let $\lambda(u)$ be the inverse Mills' ratio as introduced in Sect. 2. We have:

$$\begin{aligned} \mathrm{E}\left(\begin{bmatrix} \hat{\nu_2}_{OLS} \\ \hat{\boldsymbol{\beta}}_{OLS} \end{bmatrix}\right) &= \mathrm{E}[\mathrm{E}((\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{y}_s \mid \mathbf{X}_s)] = \\ &= \mathrm{E}[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathrm{E}(\mathbf{y}_s \mid \mathbf{X}_s)] = \\ &= \mathrm{E}\left[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \left(\mathbf{X}_s \begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} + \rho\sigma_2 \boldsymbol{\lambda}_u\right)\right] = \\ &= \begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} + \rho\sigma_2 \mathrm{E}[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \boldsymbol{\lambda}_u] \end{aligned} \tag{5}$$

where $\boldsymbol{\lambda}_u^T = [\lambda(u_1), \quad \lambda(u_2), \ldots, \lambda(u_n)]$, i.e. the values of the inverse Mills' ratio for the n observations.

To get an identification set for $\boldsymbol{\beta}$ we use (5). We see that in order to estimate $\boldsymbol{\beta}$ both ρ and σ_2 are needed. Since we know that ρ ranges between -1 and $+1$, the strategy we pursue here is to provide bounds for σ_2, which will depend on ρ, and then let our identification set depend on a restricted subset of reasonable values for ρ.

Let $\sigma_r^2 = \mathrm{E}(\mathrm{Var}(y \mid \mathbf{x}, z = 1))$ and $\tilde{\sigma}_1^2(\mathbf{x}) = \mathrm{Var}(z^* \mid \mathbf{x}, z = 1)$. Since $\sigma_\varepsilon^2 = \sigma_2^2(1 - \rho^2)$ we have:

$$\begin{aligned} \sigma_r^2 &= \mathrm{E}(\mathrm{Var}(\eta_2 \mid \mathbf{x}, z = 1)) = \mathrm{E}\left[\mathrm{Var}\left(\rho\sigma_2\eta_1 + \varepsilon \mid \mathbf{x}, z = 1\right)\right] = \\ &= \mathrm{E}\left[\sigma_\varepsilon^2 + \rho^2\sigma_2^2\tilde{\sigma}_1^2(\mathbf{x})\right] = \mathrm{E}\left[\sigma_2^2 - \rho^2\sigma_2^2 + \rho^2\sigma_2^2\tilde{\sigma}_1^2(\mathbf{x})\right] = \\ &= \sigma_2^2\left(1 - \rho^2\left(1 - \mathrm{E}\left[\tilde{\sigma}_1^2(\mathbf{x})\right]\right)\right) \end{aligned}$$

where $0 \leq \left(1 - \mathrm{E}\left[\tilde{\sigma}_1^2(\mathbf{x})\right]\right) \leq 1$ for all $\mathbf{x}$, since $\tilde{\sigma}_1^2(\mathbf{x}) \leq \mathrm{Var}(z^* \mid \mathbf{x}) = 1$, for all $\mathbf{x}$. Hence, we get the inequality:

$$\sigma_r^2 \leq \sigma_2^2 \leq \frac{\sigma_r^2}{1 - \rho^2}. \tag{6}$$

From (5) and (6) we now can obtain identification sets for all components of $\boldsymbol{\beta}$. Let:

$$
\begin{aligned}
b_{1,j} &= \mathrm{E}\left(\hat{\beta}_j\right) - \rho_{min}\frac{\sigma_r}{\sqrt{1-\rho_{min}^2}}\mathrm{E}[(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_u]\boldsymbol{e}_j,\\
b_{2,j} &= \mathrm{E}\left(\hat{\beta}_j\right) - \rho_{min}\sigma_r\mathrm{E}[(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_u]\boldsymbol{e}_j,\\
b_{3,j} &= \mathrm{E}\left(\hat{\beta}_j\right) - \rho_{max}\frac{\sigma_r}{\sqrt{1-\rho_{max}^2}}\mathrm{E}[(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_u]\boldsymbol{e}_j,\\
b_{4,j} &= \mathrm{E}\left(\hat{\beta}_j\right) - \rho_{max}\sigma_r\mathrm{E}[(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_u]\boldsymbol{e}_j,
\end{aligned}
$$

for $j = 1, \ldots, p$, where $\boldsymbol{e}_j$ is a $(p+1)$ vector with all elements 0 except the $(j+1)$:th which is 1. Then the lower ($\beta_{l,j}$) and upper ($\beta_{u,j}$) bounds of the identification set are:

$$
[\beta_{l,j} = \min(b_{1,j}, b_{2,j}, b_{3,j}, b_{4,j}),\ \beta_{u,j} = \max(b_{1,j}, b_{2,j}, b_{3,j}, b_{4,j})] \tag{7}
$$

We can see that if we only know that $\rho \in [-1, 1]$ then the identification sets range from $-\infty$ to $+\infty$. In cases where we have knowledge on ρ, e.g., $\rho \in [\rho_{min}, \rho_{max}]$, where either $-1 < \rho_{min}$ and/or $\rho_{max} < 1$, we get a bounded identification set for β_j.

3.2 Relaxing distributional assumptions

Let $\mathrm{E}(y \mid \mathbf{x}) = \nu_2 + \mathbf{x}^T\boldsymbol{\beta}$, $\mathrm{Var}(y \mid \mathbf{x}) = \sigma_2^2$ and

$$
P(z = 1 \mid y, \mathbf{x}) = \exp\left[H\left(\alpha_0 + \frac{\rho}{\sqrt{1-\rho^2}}\frac{(y - \nu_2 - \mathbf{x}^T\boldsymbol{\beta})}{\sigma_2}\right)\right],
$$

where H is a known differentiable function and $\alpha_0 = \frac{\nu_1 + \mathbf{x}^T\boldsymbol{\delta}}{\sqrt{1-\rho^2}}$. Under some additional regularity assumptions we have (see Appendix):

$$
\begin{bmatrix}\nu_2\\ \boldsymbol{\beta}\end{bmatrix} = \mathrm{E}\left(\begin{bmatrix}\hat{\nu}_{2OLS}\\ \hat{\boldsymbol{\beta}}_{OLS}\end{bmatrix}\right) - \sigma_r\frac{\rho}{\sqrt{1-\rho^2}}\mathrm{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{H}'_{\alpha_0}] + O(\rho^2).
$$

Note that under the model assumptions of Sect. 2, i.e. $\boldsymbol{H}'_{\alpha_0} = \boldsymbol{\lambda}_{\alpha_0}$, we get:

$$
\begin{bmatrix}\nu_2\\ \boldsymbol{\beta}\end{bmatrix} = \mathrm{E}\left(\begin{bmatrix}\hat{\nu}_{2OLS}\\ \hat{\boldsymbol{\beta}}_{OLS}\end{bmatrix}\right) - \sigma_r\frac{\rho}{\sqrt{1-\rho^2}}\mathrm{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\lambda}_{\alpha_0}] + O(\rho^3)
$$

which corresponds to (7) up to convergence order.

3.3 Taking the sampling variability into account: uncertainty intervals

With $\hat{\beta}_j$ we denote the j-th element of $\hat{\boldsymbol{\beta}}_{OLS}$; see (4). The bounds (7) can be estimated from the observed data, by using $\hat{\beta}_j$ for $\mathrm{E}\left(\hat{\beta}_j\right)$, and by estimating $\mathrm{E}[(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_u]$

with $(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T\boldsymbol{\lambda}_{\hat{u}}$, where the parameters of u are estimated with a probit regression to yield $\hat{u}$. Also σ_r^2 can be estimated with the residual sample variance of the OLS fit, thereby implying a slight overestimation of σ_r^2. The latter can be seen with the following asymptotic argument:

$$\begin{aligned}\sigma_{OLS}^2 &= \mathrm{Var}(y - \nu_{2OLS} - \mathbf{x}^T\boldsymbol{\beta}_{OLS} \mid z = 1)\\ &= \sigma_r^2 + \mathrm{Var}(\mathrm{E}(y - \nu_{2OLS} - \mathbf{x}^T\boldsymbol{\beta}_{OLS} \mid \mathbf{x}, z = 1) \mid z = 1) > \sigma_r^2,\end{aligned}$$

where σ_{OLS}^2, ν_{2OLS} and $\boldsymbol{\beta}_{OLS}$ are the limits in probability of the OLS and where necessary regularity conditions are assumed for the first equality to hold. The overestimation is slight if $\mathrm{E}(y \mid \mathbf{x}, z = 1)$ is close to linear as a function of $\mathbf{x}$, which is often the case in applications (Puhani 2000).

These estimates will induce sampling variability into the identification set. The latter variability is incorporated to create uncertainty intervals with a confidence level of at least $(1 - \alpha)100\%$:

$$\left[\hat{\beta}_{l,j} - c_{\frac{\alpha}{2}}\mathrm{se}(\hat{\beta}_{l,j}),\ \hat{\beta}_{u,j} + c_{\frac{\alpha}{2}}\mathrm{se}(\hat{\beta}_{u,j})\right],$$

where $c_{\frac{\alpha}{2}}$ is the $(1 - \alpha/2)100\%$ percentile of the standard normal distribution, since $\hat{\beta}_{l,j}$ and $\hat{\beta}_{u,j}$ are asymptotically normal. This is a strong uncertainty region as defined in Vansteelandt et al. (2006), that is it covers all values in the identification set ($[\beta_{l,j}, \beta_{u,j}]$) with at least $(1 - \alpha)100\%$ probability.

Estimation of $\mathrm{se}(\hat{\beta}_{l,j})$ and $\mathrm{se}(\hat{\beta}_{u,j})$ can be performed with bootstrap techniques since all estimated quantities are identified. In this paper, however, we simply use the standard errors of the OLS estimates $\hat{\beta}_j$ to construct the uncertainty intervals. This implies an underestimation of the sampling variability but our simulations suggest that this is compensated by the otherwise conservative use of strong uncertainty intervals.

4 Predictors of BMI changes for middle age men

The analysis is performed on data collected via the Västerbotten Intervention Programme (VIP) (Norberg et al. 2010). VIP was initiated in 1985 to counter the high prevalence for cardiovascular disease in Västerbotten county, north of Sweden. From 1991 all residents turning 40, 50 and 60 have been asked to participate. We study all married or cohabiting 40 year old males born 1950–1956 who have chosen to participate, looking for predictors of BMI change from 40 to 50 years of age. By using Swedish personal numbers, these data are linked to socioeconomic and demographic information. At the 50 year call only 3,324 out of the 4,648 males that came to the 40 year call returned, so we have a dropout of 28.5 %. With such a level of dropout, we may question the reliability of standard OLS techniques, that rely on the missing at random assumption. In particular, a possibility could be that individuals that do not show up for the second check up have a larger increase in BMI than the ones that do (corresponding to a negative ρ).

Table 3 Width and center of 95 % uncertainty and confidence intervals

	$-0.9 < \rho < 0$			$-0.5 < \rho < 0$			$\lvert\rho\rvert < 0.5$			OLS			Extreme ρ
	Width	Sign.	Center	Width	Sign.	Center	Width	Sign.	Center	Width	Sign.	Center	
Intercept	5.30	1	7.97	4.20	1	7.42	4.63	1	7.20	3.77	1	7.20	0.99
Baseline BMI	0.11	−1	−0.13	0.07	−1	−0.15	0.08	−1	−0.16	0.05	−1	−0.16	−0.98
log(Earnings)	0.14	0	−0.03	0.10	0	−0.01	0.12	0	0.00	0.09	0	0.00	
log(Spouse earnings/earnings)	0.12	0	−0.02	0.08	0	0.00	0.09	0	0.01	0.06	0	0.01	
Spouse age - age	0.07	0	0.01	0.06	0	0.02	0.07	0	0.02	0.06	0	0.02	
Number of children	0.79	0	−0.15	0.59	0	−0.05	0.67	0	−0.01	0.51	0	−0.01	
Hospitalization (days)	0.97	0	0.17	0.74	0	0.05	0.83	0	0.01	0.65	0	0.01	
Children aged leq 3	0.70	0	0.19	0.53	0	0.11	0.59	0	0.08	0.46	0	0.08	
Education > 9 years	0.95	0	−0.34	0.81	0	−0.27	0.86	0	−0.24	0.76	0	−0.24	
Education difference	0.79	0	−0.15	0.67	0	−0.09	0.72	0	−0.07	0.63	0	−0.07	
Parent leave benefits > 0	0.68	0	−0.26	0.47	0	−0.16	0.55	0	−0.11	0.38	0	−0.11	
Student benefits > 0	1.14	0	−0.19	1.10	0	−0.22	1.11	0	−0.22	1.08	0	−0.22	
Sick leave benefit > 0	0.67	0	0.19	0.49	0	0.10	0.56	0	0.06	0.42	0	0.06	
Unemployment benefits > 0	0.95	0	−0.02	0.67	0	−0.15	0.78	0	−0.21	0.56	0	−0.21	
Tobacco use	0.35	1	0.22	0.23	1	0.16	0.28	1	0.14	0.18	1	0.14	0.52
Positive self-reported health	0.50	−1	−0.39	0.47	−1	−0.41	0.48	−1	−0.42	0.45	−1	−0.42	−0.99
Living in urban area	0.43	0	−0.14	0.39	0	−0.16	0.41	0	−0.16	0.38	0	−0.16	

The sign column takes value −1,0,1, if interval is below 0, contains 0, or above 0. Extreme ρ indicate for what values of ρ the uncertainty interval includes 0

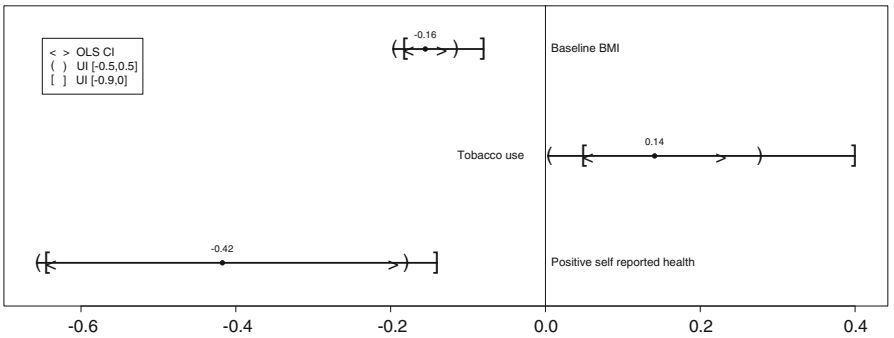

Fig. 2 Graphical display of the OLS estimates (dot), the 95 % confidence intervals (CI) and uncertainty intervals (UI) obtained with ρ in $[-0.5, 0.5]$ and $[0, 0.9]$, for the three variables of Table 3 which were significant at the 5 % level in the OLS fit.

In Table 3 we present uncertainty intervals that are built assuming three different sets of values for ρ: $(-0.9, 0)$, $(-0.5, 0)$ and $(-0.5, 0.5)$ and confidence intervals obtained by assuming missing at random (i.e., letting $\rho = 0$ and using OLS). The first two sets of values for ρ illustrate an assumption that ρ is not positive. We consider also an interval containing both negative and positive values. The uncertainty intervals are obtained from data as described in Sect. 3.3.

The results obtained are bestly displayed graphically as done in Fig. 2 for the three variables which are significant at the 5 % level in the OLS analysis. Results show that only the two covariates 'Baseline BMI' and 'Positive self-reported health' have a non-zero negative effect under all ranges of ρ considered. Their UI:s contain the value zero only under a rather extreme negative correlation (i.e. -0.98 or lower). On the other hand, 'Tobacco use' has a non-zero positive effect under all ranges of ρ considered, although the UI may contain zero for positive ρ:s larger than 0.52. Such considerations are the added value with respect to the analyses summarised in Table 2.

Note that 'Positive self-reported health' is the only significant variable that was not significant in the probit regression. If the component of $\boldsymbol{\delta}$ corresponding to "Positive self-reported health" is zero the corresponding component $\hat{\boldsymbol{\beta}}_{OLS,z=1}$ is not distorted and therefore our identification set will be reduced to a point, see Hutton and Stanghellini (2010). The corresponding uncertainty interval is then equivalent to a confidence interval with same level. On the other hand, if one of the element in $\boldsymbol{\delta}$ is small, it will reflect in the estimates and we will still get a rather narrow uncertainty interval. For that reason, when constructing the uncertainty intervals, we have used all available covariates in the probit regression, see Table 1.

5 A simulation study based on a wage offer study

5.1 Design of the study

The design is an attempt to mimic the characteristics of a case study on married women's wage mentioned in Sect. 2. The study focused on estimating the wage offer Eq. (1) given a set of observed covariates, with a selected sample since wage is

observed only for the women who work; see Mroz (1987) and for a more recent analysis (Wooldridge 2003, Chap. 17.4). The covariates used are 'Household income–woman's income' ($nwifeinc$), 'Educational attainment in years' ($educ$), 'Years of labour market experience' ($exper$), 'Age', 'Number of children 5 years or younger' ($kids5$) and 'Number of children 6–18 years old' ($kids618$).

The simulated samples in this study are obtained by drawing with replacement a given number of units out of the 753 women in the study. We use their true values on all explanatory variables, but simulate a new response variable using models (1) and (2), while setting $\rho = 0.1,\ 0.2$ or 0.4. The other parameters ($\boldsymbol{\delta}$, $\boldsymbol{\beta}$ and σ_2) are set to their estimated values obtained from TSLS applied to the original dataset with all covariates included in the selection equation and age, $kids5$ and $kids618$ are excluded from the outcome equation. More specifically, given $\mathbf{x}$ we simulate data from the following model:

$$y = -0.452 + \mathbf{x}\,[0.006, 0.097, 0.039, -0.001, 0, 0, 0]^T + \rho \cdot \sigma_2 \cdot \eta_1 + \varepsilon,$$
$$z^* = 0.270 + \mathbf{x}\,[-0.012, 0.131, 0.123, -0.002, -0.053, -0.868, 0.036]^T + \eta_1,$$

where $\mathbf{x} = [nwifeinc, educ, exper, exper^2, age, kids5, kidsge618]$, $\eta_1 \sim N(0, 1)$ and $\sigma_2 = 0.662$. In order to mimic the marginal distribution of the observed women's wages, the distribution of ε is chosen to be a centered gamma: $\varepsilon = \mathrm{E}(G) - G$, where G is gamma distributed with equal shape and scale parameters (i.e. both parameters are equal to $\mathrm{Var}(\varepsilon)^{1/3}$). From our model assumptions (see Sect. 3.1) we also impose that $\sqrt{\mathrm{Var}(\varepsilon)} = \sigma_2\sqrt{1-\rho^2}$, thereby implying that $\sqrt{\mathrm{Var}(\varepsilon)} = 0.659, 0.649, 0.607$ for the different values of ρ respectively. In this study we build 10,000 replicates of samples with sizes 100, 350 and 753.

For the identification sets (7) we let $\rho \in [0,\ 0.5]$ and compute uncertainty intervals as described in Sect. 3.3. We apply the TSLSs procedure without restrictions and with two different exclusion restrictions: TSLS E1, where we exclude three variables (age, $kids5$ and $kids618$) from the outcome equation, i.e. TSLS E1 corresponds to the data generating mechanism of the study; TSLS E2, where four variables are excluded (i.e., also $nwifeinc$) from the outcome equation, i.e. TSLS E2 is a misspecified model. Finally, OLS estimates are also produced.

5.2 Results

Results for the $\boldsymbol{\beta}$ coefficient corresponding to $educ$ are summarized in Fig. 3, where the width of the uncertainty interval and confidence intervals for the 10,000 replicates[1] are reported with box plots. Empirical coverage are also given in the figure. As expected TSLS implies confidence intervals due to collinearity problems (variance inflation

[1] TSLS will sometimes estimate ρ outside of $[-1,\ 1]$ which will lead to unstable results. These replicates are removed. Thus, about 49, 23 and 9 % of the replicates are removed for TSLSs procedure without exclusion restrictions with sample size 100, 350 and 753 respectively. The same problem occurred in about 6 % of the replicates for the smallest sample size and only at a few occasions for the other sample sizes when using exclusion restrictions (TSLS E1 and TSLS E2). The simulations are performed with R software.

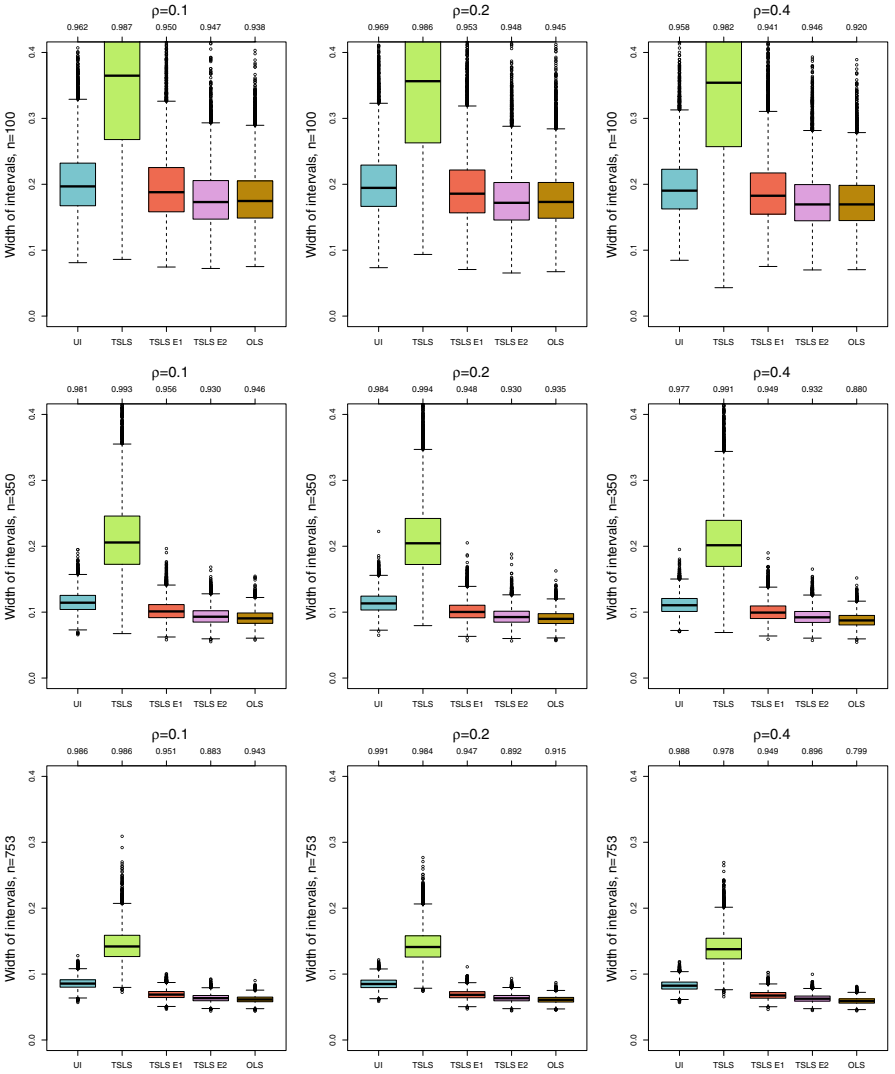

Fig. 3 Box plot of the width of 95 % uncertainty intervals and 95 % confidence intervals for the regression coefficient of *educ* when varying ρ and sample sizes. The empirical coverage of the intervals are above each box.

factor ranging from around 10 to 100). TSLS E1 (correctly specified model) yields tighter confidence intervals and empirical coverage close to the nominal level. TSLS E2 gives too low empirical coverage for the parameter due to model misspecification (a problem that increases with sample size), as does OLS in all cases for the same reason.

Uncertainty intervals are not directly comparable to confidence intervals since they converge to a non-degenerate interval as sample size grows. Thus, uncertainty intervals are expected to be wider than confidence intervals with correctly specified model

(TSLS E1). Uncertainty intervals should—and in our simulations do—imply a higher empirical coverage rate than the nominal level since they are constructed to take into account the uncertainty due to the unknown parameter ρ. By letting ρ be uncertain we avoid the need to have prior knowledge about exclusion restrictions, and we see that using an incorrect exclusion restriction (TSLS E2) can lead to serious under-coverage. However, one should also note that the coverage of the uncertainty intervals relies on correct a priori information on ρ, i.e. an interval for ρ containing the true value. Using an interval for ρ not containing the true value will typically yield too low coverage. Using $\rho \in [-0.5, 0]$ instead of $\rho \in [0, 0.5]$ in the above simulations yielded empirical coverages often below 95 % although higher than coverages obtained with OLS, since $\rho = 0$ is included.

Finally, it is worth noting that the p values of the inverse Mills' ratio (obtained in the second stage of TSLS) are not significant in most of the replicates (even with non-zero ρ), making the corresponding test of no selection not reliable in practice, i.e. the data carry little information on whether the sample is selected or not.

6 Discussion

We have shown how to compute bounds on the parameters of a regression model with missing continuous outcome without making strong untestable assumptions about missing data. The bounds make evident which inference can be made with reasonably mild restrictions on the value of ρ, which expresses the correlation between the unmeasured factor that drives the missingness mechanism and the residuals of the regression model under study. This is especially important with large datasets, where the sampling variation will be small and therefore the lack of knowledge on ρ is the major cause for uncertainty. Furthermore, these bounds can be computed without imposing any exclusion restriction and contain the missing at random assumption as a particular case. Therefore, they provide an indication of the impact that the untestable assumptions have on the inference of the parameters. Note that simulations show that correct coverage of the uncertainty intervals relies on specifying an interval for ρ containing the true value. An alternative to bounds is to use Bayesian inference, where a posterior distribution of the parameters of interest is deduced by integrating out the nuisance parameter ρ (Daniels and Hogan 2008; Rubin 1977). Our approach has the advantage of relaxing distributional assumptions. Since the bounds are based on standard OLS techniques, they are also easy to compute using standard statistical softwares.

Acknowledgments We are grateful to the referees for their detailed and constructive criticism. We thank Per Johansson, Anders Lundquist, and Mathias Lundin for helpful comments. We also acknowledge the financial support of the Swedish Research Council through the Swedish Initiative for Research on Microdata in the Social and Medical Sciences (SIMSAM), and the Ageing and Living Condition Program.

Appendix

Derivation of identification set under the assumptions of Sect. 3.2

Let $q = \alpha_0 + \frac{\rho}{\sqrt{1-\rho^2}} \frac{(y-\nu_2-\mathbf{x}^T\boldsymbol{\beta})}{\sigma_2}$. Then by Taylor expanding $P(z = 1 \mid y, \mathbf{x})$ around $q = \alpha_0$ we get:

$$P(z = 1 \mid y, \mathbf{x}) = \exp H\left(\alpha_0 + \frac{\rho}{\sqrt{1-\rho^2}} \frac{(y - \nu_2 - \mathbf{x}^T\boldsymbol{\beta})}{\sigma_2}\right) =$$
$$= e^{H(\alpha_0)}\left(1 + \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0)\left(\frac{y - \nu_2 - \mathbf{x}^T\boldsymbol{\beta}}{\sigma_2}\right) + D\right)$$

where $D = \left(\frac{\rho}{\sqrt{1-\rho^2}}\right)^2 \frac{H''(a)+H'(a)^2}{2}\left(\frac{y-\nu_2-\mathbf{x}^T\boldsymbol{\beta}}{\sigma_2}\right)^2$ for some a between α_0 and q. From this we get:

$$P(z = 1 \mid \mathbf{x}) = \mathrm{E}_y(P(z = 1 \mid y, \mathbf{x})) =$$
$$= e^{H(\alpha_0)}\left(1 + \mathrm{E}\left(\frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0)\left(\frac{y - \nu_2 - \mathbf{x}^T\boldsymbol{\beta}}{\sigma_2}\right)\right) + D'\right) =$$
$$= e^{H(\alpha_0)}\left(1 + D'\right)$$

where $D' = \mathrm{E}(D)|_{\mathbf{x}}$. Since

$$f(y \mid z = 1, \mathbf{x}) = \frac{1}{P(z = 1 \mid \mathbf{x})} f(y \mid \mathbf{x}) P(z = 1 \mid y, \mathbf{x})$$

we have

$$\mathrm{E}[y \mid z = 1, \mathbf{x}] =$$
$$= \frac{\int y f(y \mid \mathbf{x}) e^{H(\alpha_0)}\left(1 + \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0)\left(\frac{y-\nu_2-\mathbf{x}^T\boldsymbol{\beta}}{\sigma_2}\right) + D\right) dy}{e^{H(\alpha_0)}\left(1 + D'\right)} =$$
$$= \frac{\int A + B + C dy}{(1 + D')}$$

with

$$\int A dy = \int y f(y \mid \mathbf{x}) dy = \mathrm{E}(y \mid \mathbf{x}) = \nu_2 + \mathbf{x}^T \boldsymbol{\beta}$$

$$\int B dy = \int y f(y \mid \mathbf{x}) \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0) \left(\frac{y - \nu_2 - \mathbf{x}^T \boldsymbol{\beta}}{\sigma_2} \right) dy =$$

$$= \frac{\frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0)}{\sigma_2} \int y f(y \mid \mathbf{x}) \left(y - \nu_2 - \mathbf{x}^T \boldsymbol{\beta} \right) dy =$$

$$= \sigma_2 \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0)$$

$$\int C dy = \int y f(y \mid \mathbf{x}) D dy = D''.$$

So we have:

$$\mathrm{E}[y \mid z = 1, \mathbf{x}] = \frac{\nu_2 + \mathbf{x}^T \boldsymbol{\beta} + \sigma_2 \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0) + D''}{(1 + D')} =$$

$$= \nu_2 + \mathbf{x}^T \boldsymbol{\beta} + \sigma_2 \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0) + D'''$$

where $D''' = \frac{D'' - D' \left(\nu_2 + \mathbf{x}^T \boldsymbol{\beta} + \sigma_2 \frac{\rho}{\sqrt{1-\rho^2}} H'(\alpha_0) \right)}{1 + D'}$.

Let $\mathbf{X}_s$ be as in Sect. 3.1. If we let $(\boldsymbol{H}'_{\alpha_0})^T = [H'(\alpha_{01}) \quad H'(\alpha_{02}) \quad \cdots \quad H'(\alpha_{0n})]$ then:

$$\mathrm{E}\left(\begin{bmatrix} \hat{\nu}_{2OLS} \\ \hat{\boldsymbol{\beta}}_{OLS} \end{bmatrix} \right) = \mathrm{E}[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathrm{E}(y \mid \mathbf{X}_s)] =$$

$$= \mathrm{E}\left[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \left(\mathbf{X}_s \begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} + \sigma_2 \frac{\rho}{\sqrt{1-\rho^2}} \boldsymbol{H}'_{\alpha_0} + D''' \right) \right] =$$

$$= \begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} + \frac{\rho}{\sqrt{1-\rho^2}} \sigma_2 \mathrm{E}\left[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \boldsymbol{H}'_{\alpha_0} \right] + O(\rho^2). \tag{8}$$

For the last step to be valid we need to assume that $(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T D'''$ is dominated by some B_n and that $\mathrm{E}(B_n) < \infty$. By using (6) and (8) we get:

$$\begin{bmatrix} \nu_2 \\ \boldsymbol{\beta} \end{bmatrix} = \mathrm{E}\left(\begin{bmatrix} \hat{\nu}_{2OLS} \\ \hat{\boldsymbol{\beta}}_{OLS} \end{bmatrix} \right) - \sigma_r \frac{\rho}{\sqrt{1-\rho^2}} \mathrm{E}[(\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \boldsymbol{H}'_{\alpha_0}] + O(\rho^2)$$

since

$$\frac{\sigma_r}{\sqrt{1-\rho^2}}\frac{\rho}{\sqrt{1-\rho^2}}\mathrm{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{H}'_{\alpha_0}] = \\ = \sigma_r\frac{\rho}{\sqrt{1-\rho^2}}\mathrm{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{H}'_{\alpha_0}] + O(\rho^2).$$

References

Andridge R, Little R (2011) Proxy pattern-mixture analysis for survey nonresponse. J Off Stat 27:153–180

Copas J, Eguchi S (2005) Local model uncertainty and incomplete data bias. J R Stat Soc Ser B 67(4): 459–513

Copas J, Li H (1997) Inference for non-random samples. J R Stat Soc Ser B 59(1):55–95

Copas JB (2013) A likelihood-based sensitivity analysis for publication bias in meta-analysis. J R Stat Soc Ser C 62:47–66

Daniels M, Hogan J (2008) Missing data In longitudinal studies: strategies for Bayesian modeling and sensitivity analysis. Chapman and Hall/CRC, Boca Raton

de Luna X, Lundin M (2014) Sensitivity analysis of the unconfoundedness assumption with an application to an evaluation of college choice effects on earnings. J Appl Stat 41(8):1767–1784

Heckman J (1979) Sample selection bias as a specification error. Econometrica 47(1):153–161

Henmi M, Copas JB, Eguchi S (2007) Confidence intervals and p-values for meta-analysis with publication bias. Biometrics 63:475G482

Horowitz J, Manski C (2006) Identification and estimation of statistical functionals using incomplete data. J Econom 132(2):445–459

Hutton J, Stanghellini E (2010) Modelling bounded health scores with censored skew-normal distributions. Stat Med 30(4):368–376

Imbens G (2003) Sensitivity to exogeneity assumptions in program evaluation. Am Econ Rev 93:126–132

Imbens G, Manski C (2004) Confidence intervals for partially identified parameters. Econometrica 72(6):1845–1857

Jonsson R (2012) When does heckmans two-step procedure for censored data work and when does it not? Stat Pap 53:33–49

Lennox C, Francis J, Wang Z (2012) Selection models in accounting research. Account Rev 87(2):589–616

Little R (1985) A note about models for selectivity bias. Econometrica 53(6):1469–1474

Little R (2009) Selection and pattern-mixture models. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) Longitudinal data analysis. Chapman and Hall/CRC, Boca Raton

Little R, Rubin D (2002) Statistical analysis with missing data, 2nd edn. Wiley, Hoboken, New Jersey

Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Stern JPSH (2012) Special report: the prevention and treatment of missing data in clinical trials. N Engl J Med 367:1355–1360

Manski C (2003) Partial identification of probability distributions. Springer, New York

Mroz T (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica 55(4):765–799

Norberg M, Wall S, Boman K, Weinehall L (2010) The Västerbotten intervention programme: background, design and implications. Glob Health Action 3. doi:10.3402/gha.v3i0.4643

Olsen R (1982) Distributional tests for selectivity bias and a more robust likelihood estimator. Int Econ Rev 23(1):223–240

Puhani (2000) The Heckman correction for sample selection and its critique. J Econ Surv 14(1):53–68

Rosenbaum P (2010) Design of observational studies. Springer, New York

Rubin D (1977) Formalizing subjective notions about the effect of nonrespondents in sample surveys. J Am Stat Assoc 72:538–543

Scharfstein D, Rotnitsky A, Robins J (1999) Adjusting for non-ignorable drop-out using semiparametric models. J Am Stat Assoc 94:1096–1120

Vansteelandt S, Goetghebeur E (2001) Analyzing the sensitivity of generalized linear models to incomplete outcomes via the ide algorithm. J Comput Graph Stat 10(4):656–672

Vansteelandt S, Goetghebeur E, Kenward M, Molenberghs G (2006) Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Stat Sin 16:953–979
Wooldridge J (2003) Econometric analysis of cross section and panel data. The MIT Press, Cambridge