Robustness in Sample Selection Models Based on the t Distribution and its Influence Functions

Dylan Halfhide

A Thesis presented for the degree of Master of Science in Econometrics and Management Science

Erasmus School of Economics
Erasmus University Rotterdam
Netherlands
23 April 2021

Supervisor: Mikhail Zhelonkin
Co-reader: Andrea Naghi

Abstract

The Heckman selection model is a frequently used model when there is concern of sample selection bias. In this paper the selection model is adapted under a biavariate-t distribution and its robustness properties are studied. The classical and robust two-stage estimators are adapted under this distribution and its influence functions are constructed. Monte Carlo studies are done to show performance of the robust two-stage estimator and its influence functions. The Monte Carlo studies show the robustness issues when testing for sample selectivity.

Keywords: Heckman model; Influence function; EM algorithm; M-estimator; Robust estimation; Robust inference; Sample selection; Gosset; Student-t distribution.

1 Introduction

Sample selection bias is a frequent occurring problem in statistical and econometric modelling. This bias typically emerges when observations have been partially/fully omitted according to some selection rule. For example, bias might arise when modelling the household expenditure on a new house when the data only contains households that have bought a house. Here, the decision to buy and the expenditure amount are not independent and selection bias might occur. Selection bias occurs in multiple fields of study, such as finance, political science, biology, sociology, and economics. The classical Heckman (1979) sample selection model is well known and used for data with this bias problem. This model is also known as a Type 2 tobit model (Amemiya (1985)). It is expressed as

$$y_{1i} = I(x_{1i}^\top \beta_1 + \varepsilon_{1i} > 0), \tag{1}$$

$$y_{2i} = \begin{cases} x_{2i}^\top \beta_2 + \varepsilon_{2i}, & \text{if } y_{1i} = 1, \\ \text{NA}, & \text{if } y_{1i} = 0, \end{cases} \tag{2}$$

where x_{ji} is a vector of explanatory variables, β_j is a $p_j \times 1$ vector of parameters, $j = 1, 2$, and $I(\cdot)$ is the indicator function. The error terms ε_1 and ε_2 are assumed to follow a bivariate normal distribution with correlation ρ, and variances equal to 1 and σ^2, respectively. The expression in (1) is known as the selection equation. The expression in (2) is known as the outcome/primary equation or the equation of interest. The model is typically estimated by either maximum likelihood (MLE) or by two-step estimator (also known as control function approach). Both estimators are well-known to be sensitive to departures from normality assumption (Paarsch 1984, Zhelonkin et al. 2016, and references therein). It is often assumed that the bivariate-t distribution is a robust alternative. Marchenko and Genton (2012) proposed the Selection-t (SLt) model, where the Heckman model is estimated by MLE with the assumption of bivariate-t errors. Although this is a more flexible approach, since the SLt model has an additional parameter ν, which is the number of degrees of freedom and it can capture the long tails, it is still a parametric model. The protection against contamination in the neighborhood of the model is not guaranteed.

Liu (2005) proposed a robit regression model for the selection equation and an EM-type algorithm to estimate its parameters. Zhao et al. (2020) proposed three new algorithms (ECM, ECM(NR), and ECME) for the Heckman Selection model. They have the advantages of the EM algorithm, namely an easy implementation and numerical stability. These newly proposed algorithms give slightly better estimates than the two-step estimator. However, they can be easily influenced by their starting values. The new algorithms also take more time to estimate. In this paper I study the robustness of the two-step estimator with bivariate-t errors. In Section 2 the influence function (IF) of the MLE is derived. Section 3 discusses the two-step estimator and its influence function. In Section 4 the robust two-step estimator and its influence functions (IF) are proposed. Proposed by Hampel (1974), the sensitivity curve or the empirical influence function (EIF), a finite-sample version of the IF. The EIF is further described in Section 4.4 and used as an approximate for the IF in the second simulation study. In Section 5 two Monte Carlo simulation studies are performed. The first study investigates the bias, variance, and mean squared errors of the classical and robust two-step estimators. The second study investigates the IF of the classical and robust estimators. This paper does not propose a Two-Step estimator of the degrees of freedom ν of the error distribution. Instead the parameter ν is is estimated using two different methods. These are described in Section 4.3. The following error distribution for the Heckman selection model is assumed throughout this paper

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim t_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}, \nu \right\}, \tag{3}$$

where ε_{1i} and ε_{2i} are the errors of the selection and outcome equation respectively, t_2 a bivariate-t distribution, ρ the correlation, and σ^2 the variance.

2 Influence Functions of the ML estimators

In this section the IF of the maximum likelihood estimators are derived. Hampel et al. (1986) describes the IF as "the (approximate and standardized) effect of an additional observation in any point x on a statistic T, given a (large) sample with distribution F". The IF can be loosely defined as the marginal effect of an observation $\tilde{x} \in \mathbb{R}$ on an estimator $\hat{\theta}(x_1, ..., x_{n-1}, \tilde{x})$. Let $z_i^* = \{y_{1i}, x_{1i}, y_{2i}, x_{2i}\} \in \mathbb{R}^{p+q+2}$ be a set of observations, where $i = 1, ..., n$, and $\theta^T = (\beta_1^T, \beta_2^T, \rho, \sigma, \nu)^T \in \mathbb{R}^{p+q} \times [-1, 1] \times \mathbb{R}^+ \times \mathbb{R}^+$ a set of parameters. Then the log-likelihood function can be written as a M-estimator with loss function $\rho(z_i^*, \theta) = -l(\theta|z_i^*)$. The IF of an M-estimator is given by:

$$IF(z_i^*; \theta^{ML}, T_2) = M(\Psi, T_2)^{-1}\Psi(z_i^*, \theta^{ML}), \tag{4}$$

where $M(\Psi, T_2) \equiv E[-\frac{\partial}{\partial\theta}\Psi], \Psi(z_i^*, \theta^{ML}) \equiv -\frac{\partial}{\partial\theta}l(\theta|z_i^*)$, and $l(\theta|z_i^*)$ is the log-likelihood function. The score functions and second-order derivatives of the SLt MLE with errors terms described in (3) are derived in Marchenko and Genton (2012). These derivations are used in the rest of Section 2.

2.1 Influence under truncation

Let $\Psi(z_i^*, \theta^{ML})$ be the vector of all the score functions of the SLt MLE times -1 (described in the Appendix of Marchenko and Genton 2012) . Let $M(\Psi, T_2)^{-1}$ be the inverse matrix of the expectation of the second derivatives of the score functions times -1. Then, the elements of the IF of β_1 are the linear combination of the first row of $M(\Psi, T_2)^{-1}$ and $\Psi(z_i^*, \theta^{ML})$. The same holds for the IF of all other variables in $\theta^T = (\beta_1^T, \beta_2^T, \rho, \sigma, \nu)^T$. Under truncation (for one observation $z^* = (y_1, x_1^T, y_2, x_2^T))$ the Ψ functions become the following:

$$\Psi(z^*, \theta)|_{y_1=0} = \left\{ M_\nu(-x_1^T\beta_1)x_1, 0, 0, 0, -\frac{\partial \ln T(-x_1^T\beta_1; \nu)}{\partial\nu} \right\}. \tag{5}$$

where $M_\nu(-x_1^T\beta_1) = \frac{t(-x_1^T\beta_1;\nu)}{T(-x_1^T\beta_1;\nu)}$, and $t(.)$ and $T(.)$ are the probability density and cumulative distribution function of the univariate-t distribution respectively. The derivation above shows that under truncation, the IF of all MLE parameters are proportional to the linear combination of $M_\nu(-x_1^T\beta_1)x_1$ and $-\frac{\partial \ln T(-x_1^T\beta_1;\nu)}{\partial\nu}$. Figure 1 plots multiple values of the variable x_1 against the non-zero score functions (5). For simplicity $\beta_1 = 1$. These plots show that the score function of ν is unbounded for large values of x_1. Thus, the IF of the MLE are also unbounded for large values of x_1.

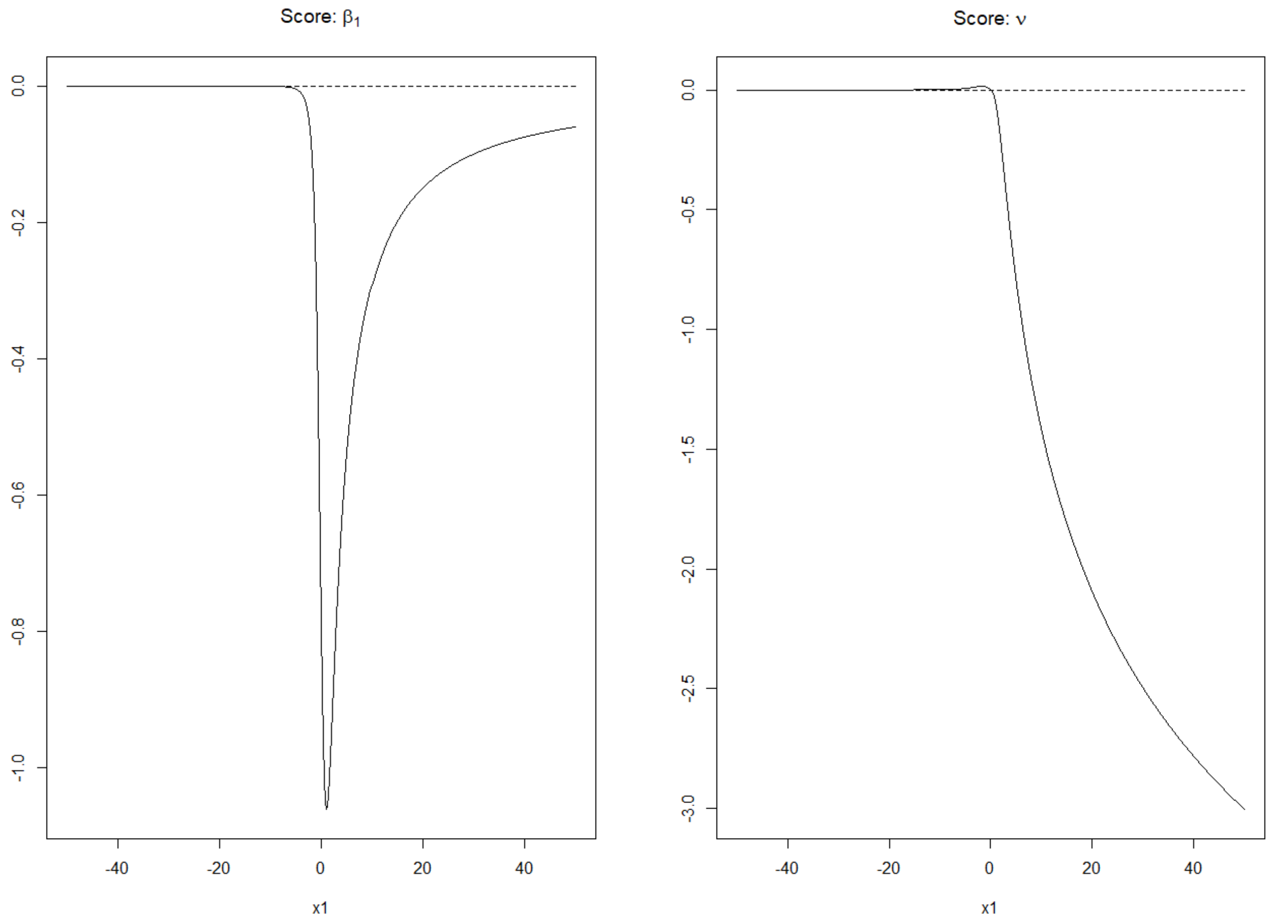

Figure 1: Values of the score functions of β_1 and ν given $y_1 = 0$

2.2 From Score Function to IF

This section describes the method to derive the elements the matrix $M(\Psi, T_2)$. Let $\theta^T = (\beta_1^T, \beta_2^T, \rho, \sigma, \nu)^T \in \mathbb{R}^{p+q} \times [-1, 1] \times \mathbb{R}^+ \times \mathbb{R}^+$ be the vector of parameters. In particular the equations for β_2 and ν are derived. The procedure used to obtain values for $M(\Psi, T_2)$ is similar to the procedure used in Serrudo (2008). The second-order derivative $S_{\beta_1\beta_1}$ (including those for the other parameters in θ) can be found in the Appendix of Marchenko and Genton (2012). The following equations describes derivation for the first element (upper left corner) of the M matrix.

$$\begin{aligned}
E\left(-\frac{\partial}{\partial\beta_1}\Psi_{\beta 1}\right) =& E(-S_{\beta_1\beta_1}) \\
=& -\frac{1}{\sigma^2} E\left(y_{1i}x_{1i}x_{1i}^T Q_\nu^2 \ \left[\left\{2z_i^2(\nu+z_i^2)^{-1} - 1\right\} + \left\{\zeta_{\theta_{-\nu}}(\nu+z_i^2)^{-1}z_i - A_{\rho\rho})^2\right\} M'_{\nu+1}(\eta_\theta)\right]\right) \\
& -\frac{1}{\sigma^2} E\left(y_{1i}x_{1i}x_{1i}^T Q_\nu(\nu+z_i^2)^{-1}\left[\zeta_{\theta_{-\nu}}\left\{3(\nu+z_i^2)^{-1}z_i^2 - 1\right\} - 2A_{\rho\rho}z_i\right] M_{\nu+1}(\eta_\theta)\right),
\end{aligned} \tag{6}$$

where $Q_\nu = \sqrt{\frac{\nu+1}{\nu+z^2}}$, $\zeta_{\theta_{-\nu}} = A_{\rho\rho}z + A_\rho x_{1i}^T\beta_1$, $A_\rho = 1/\sqrt{1-\rho^2}$, $A_{\rho\rho} = \rho A_\rho$, $z_i = \frac{y_{2i} - x_{2i}^T\beta_2}{\sigma}$, and $\eta_\theta = Q_\nu\zeta_{\theta_{-\nu}}$. Let $g(\varepsilon_{2i}, x_{1i}, x_{2i}) = x_{1i}x_{1i}^T Q_\nu^2 \ \{(2z_i^2(\nu+z_i^2)^{-1} - 1) + (\zeta_{\theta_{-\nu}}(\nu+z_i^2)^{-1}z_i - A_{\rho\rho})^2 M'_{\nu+1}(\eta_\theta)\}$ be the expression in the first expectation of (6). Then,

$$
\begin{aligned}
&E\left[y_{1i}\cdot x_{1i}x_{1i}^TQ_\nu^2\ \{(2z_i^2(\nu+z_i^2)^{-1}-1)+(\zeta_{\theta-\nu}(\nu+z_i^2)^{-1}z_i-A_{\rho\rho})^2M'_{\nu+1}(\eta_\theta)\}\right.\\
&\quad=E\left[y_{1i}\cdot g(\varepsilon_{2i},x_{2i},x_{1i})\right]\\
&\quad=\frac{1}{N}\frac{1}{N_1}\sum_{t=1}^{N}\sum_{s=1}^{N_1}E_{\varepsilon_{2i}|y_{1i}}\left[g(\varepsilon_{2i},x_{2s},x_{1t})|y_{1i}=1\right]\cdot T(x'_{1t}\beta_1;\nu),
\end{aligned}
\tag{7}
$$

where $T(\cdot)$ is the cumulative t distribution function, $N_1=\sum_{i=1}^N y_{1i}$, and

$$
\begin{aligned}
E_{\varepsilon_{2i}|y_{1i}}\left[g(\varepsilon_{2i},x_{2s},x_{1t})|y_{1i}=1\right]&=E_{\varepsilon_{2i}|y_{1i}}\left[g(\varepsilon_{2i},x_{2s},x_{1t})|y_{1i}^*>0\right]\\
&=E_{\varepsilon_{2i}|y_{1i}}\left[g(\varepsilon_{2i},x_{2s},x_{1t})|\varepsilon_{1i}>-x'_{1i}\beta_1\right]\\
&=\int_{-\infty}^{\infty}g(\varepsilon_{2i},x_{2s},x_{1t})f_{h_i}(h_i)dh_i,
\end{aligned}
\tag{8}
$$

where $h_i=(\varepsilon_{2i}|\varepsilon_{1i}>-x_{1i}^T\beta_1)$ and $f_{h_i}(\cdot)=f_{EST}(\cdot)$ the pdf of the distribution of $Y_2^*|Y_1^*>0$. Marchenko and Genton (2012) derive $f_{EST}(\cdot)$ as the extended skew-t distribution, such that

$$
f_{EST}(y,\mu,\sigma,\alpha,\tau,\nu)=\frac{1}{\sigma}t(z,\nu)\frac{T\left\{(\alpha z+\tau)(\frac{\nu+1}{\nu+z^2})^{1/2};\nu+1\right\}}{T(\tau/\sqrt{1-\alpha^2};\nu)}
$$

with $z=\frac{y_2-\mu}{\sigma},\mu=x_2^T\beta_2,\alpha=\frac{\rho}{\sqrt{1-\rho^2}},\tau=\frac{x_1^T\beta_1}{\sqrt{1-\rho^2}},y\in\mathbb{R}$. Numerical integration is used to evaluate the expectation in (8). The second expectation in (6) can be evaluated in the same manner as described above. For the lower right corner of the M matrix, its expectation is as follows,

$$
\begin{aligned}
E[-S_{\nu^2}]=&-\frac{\partial^2c(\nu,\sigma)}{\partial^2\nu}E[y_1]-E\left[y_{1i}\cdot\left(\frac{1}{2\nu}\frac{z_i^2}{\nu+z_i^2}-\frac{Q_\nu^2}{2\nu}\left(\frac{1}{\nu+1}-\frac{1}{\nu+z_i^2}\right)-\frac{z_i^4}{2\nu^2}+\frac{\partial^2\mathrm{ln}T(\eta_\theta;\nu)}{\partial^2\nu}\right)\right]\\
&-E\left[\frac{\partial^2\mathrm{ln}T(-x_1^T\beta_1;\nu)}{\partial^2\nu}\right]+E\left[y_{1i}\cdot\frac{\partial^2\mathrm{ln}T(-x_{1i}^T\beta_1;\nu)}{\partial^2\nu}\right],
\end{aligned}
\tag{9}
$$

where $E[y_{1i}]=\frac{1}{N}\sum_{i=1}^N T(x_{1i}^T\beta_1;\nu)$, $E\left[\frac{\partial^2\mathrm{ln}T(-x_1^T\beta_1;\nu)}{\partial^2\nu}\right]=\frac{1}{N}\sum_{i=1}^N\frac{\partial^2\mathrm{ln}T(-x_{1i}^T\beta_1;\nu)}{\partial^2\nu}\cdot T(x_{1i}^T\beta_1;\nu)$, and $E\left[y_1\frac{\partial^2\mathrm{ln}T(-x_1^T\beta_1;\nu)}{\partial^2\nu}\right]=\frac{1}{N}\sum_{i=1}^N y_{1i}\frac{\partial^2\mathrm{ln}T(-x_{1i}^T\beta_1;\nu)}{\partial^2\nu}T(x_{1i}^T\beta_1;\nu)$. These expectations are only dependent on the variables of the selection equation. For the expression (9), the procedure of computation is similar to that of the upper left corner of the M matrix. So, first define the part of the expectation dependent on the variables of the outcome equation (and indirectly the selection equation), then apply the computation method of (7) and (8). The rest of the M matrix can be computed using the procedure described above.

3 Two-step Estimator

3.1 Two-step Estimator linear function

Proposed by Heckman (1979), the two-step estimator is another method of estimating a linear model when a selection component is present. Consider the conditional expectation of y_{2i}, given the variables x_{1i} and the selection rule y_{1i}, such that

$$
E(y_{2i}|x_{2i},y_{1i}=1)=x_{2i}^T\beta_2+E(\varepsilon_{1i}|\varepsilon_{2i}>-x_{1i}^T\beta_1). \tag{10}
$$

The conditional expectation of the error term is typically non-zero. This expectation is expressed as

$$E(\varepsilon_{1i}|\varepsilon_{2i} > -x_{1i}^T\beta_1) = \rho\sigma\frac{\nu + (x_{1i}^T\beta_1)^2}{\nu - 1}\frac{t(x_{1i}^T\beta_1, \nu)}{T(x_1^T\beta_1, \nu)}, \tag{11}$$

where $t(\cdot)$ and $T(\cdot)$ denote the density and the cumulative distribution function of a univariate-t distribution respectively with ν degrees of freedom. Given the conditional expectation the adjusted expression of the outcome equation becomes

$$y_{2i} = x_{2i}^T\beta_2 + \beta_\lambda\lambda(x_{1i}^T\beta_1; \nu) + v_i, \tag{12}$$

where v_i is a zero mean error term, $\beta_\lambda = \rho\sigma, \lambda(x_{1i}^T\beta_1; \nu) = \frac{\nu+(x_{1i}^T\beta_1)^2}{\nu-1}\frac{t(x_{1i}^T\beta_1;\nu)}{\mathbf{T}(x_{1i}^T\beta_1;\nu)}$ is the Inverse Mills Ratio (IMR) with correction factor $c(x_{1i}^T\beta_1; \nu) = \frac{\nu+(x_{1i}^T\beta_1)^2}{\nu-1}$, i=1,...,N. The selection equation is estimated by MLE using the link function based on a cumulative distribution of the t distribution, i.e., the Gosset link for generalized linear model (Koenker and Yoon (2009)). This link function is also known as the robit function Liu (2005). The second stage can be estimated by OLS with additional regressor $\lambda(x_{1i}^T\beta_1; \nu)$

3.2 IF of the Two-Step Estimators

In this section we derive the first- and second-stage IF of the two-step estimators. The procedure is similar to the one used in Zhelonkin et al. (2016). Recall that first the parameters $\beta_1 \in \mathbb{R}^q$ of the selection equation have to be estimated using the MLE with the Gosset link function. This results in the estimated parameters $\hat{\beta}_1^{ML}$ such that $T^{-1}(E[y_{1i}|x_{1i}^T]; \nu) = x_{1i}^T\beta_1$ where $y_{1i} \sim Bernoulli(T(x_{1i}^T\beta_1; \nu))$ and $(y_{1i}, x_{1i}^T)^T \in \mathbb{R}^{q+1}$, for $i = 1, ..., N$. Suppose that we have an empirical distribution F_N which puts mass $\frac{1}{N}$ on the observations $z_i = (y_{1i}, x_{1i}^T, y_{1i}y_{2i}, y_{1i}x_{1i}^T)^T$. This is similar to defining the Heckman selection model where the outcome equation is truncated for $y_{1i} = 0$. Denote a random vector $(Y_1, X_1, Y_1Y_2, Y_1X_2) \in \mathbb{R}^{p+q+2}$. Then the probability distribution F on $\mathbb{R}^{p+q+2}$ has a Heckit structure if $E_F[Y_1|x_1] = T(x_1^T\beta_1; \nu)$ and $E_F[Y_1Y_2|x_1^T, \lambda, Y_1] = Y_1x_2^T\beta_2 + Y_1\beta_\lambda\lambda$. Given the values of $Y_1 \in \{0, 1\}$ the latter expectation leads to equation (2). Let $S(F)$ be the estimator for the first stage. Let $\lambda\{(x_1, y_1); S(F)\}$ denote the dependence of λ on $S(F) = \beta_1$. Let $T(F)$ be the estimator for the second stage where the estimator is directly dependent on F and indirectly dependent on F through $S(F)$. The two-stage estimator is the solution of the of the following set of equations

$$\int \Psi_1 \{(x_1, y_1); S(F)\}\, dF \quad = \quad 0, \tag{13}$$

$$\int \Psi_2 [(x_2, y_2); \lambda\{(x_1, y_1); S(F)\}, T(F)]\, dF \quad = \quad 0, \tag{14}$$

where the score functions Ψ_1 and Ψ_2 are defined by

$$\Psi_1 \{(x_1, y_1); S(F)\} \quad = \quad (y_1 - T(x_1^T\beta_1; \nu))\frac{t(x_1^T\beta_1; \nu)}{T(x_1^T\beta_1; \nu)(1 - T(x_1^T\beta_1; \nu))}x_1, \tag{15}$$

$$\Psi_2 [(x_2, y_2); \lambda\{(x_1, y_1); S(F)\}, T(F)] \quad = \quad (y_2 - x_2^T\beta_2 - \lambda\beta_\lambda)\begin{pmatrix} x_2 \\ \lambda \end{pmatrix} y_1. \tag{16}$$

The IF for first-stage estimator is $IF(z, S, F) = M(\Psi_1)^{-1}\Psi_1 \{(x_1, y_1); S(F)\}$, where Ψ_1 is given by (15) and

$$M(\Psi_1) = \int \left[\frac{t(x_1^T\beta_1;\nu)^2}{T(x_1^T\beta_1;\nu)\{1-T(x_1^T\beta_1;\nu)\}} \right] x_1 x_1^T dF, \tag{17}$$

such that

$$IF(z,S,F) = \left(\int \left[\frac{t(x_1^T\beta_1;\nu)^2}{T(x_1^T\beta_1;\nu)\{1-T(x_1^T\beta_1;\nu)\}} \right] x_1 x_1^T dF \right)^{-1} \{y_1 - T(x_1^T\beta_1;\nu)\} \frac{t(x_1^T\beta_1;\nu)}{T(x_1^T\beta_1;\nu)\{1-T(x_1^T\beta_1;\nu)\}} x_1. \tag{18}$$

The score function for the classical first-stage estimator is unbounded. This results in an unbounded first-stage IF. The IF for the second-stage estimator is

$$IF(z,T,F) = M(\Psi_2)^{-1} \left[\Psi_2\{(x_2,y_2);\lambda,T(F)\} + \int \frac{\partial}{\partial\lambda} \Psi_2\{(x_2,y_2);\lambda,T(F)\}\lambda' \, dF \cdot IF(z,S,F) \right], \tag{19}$$

where $\lambda = \lambda\{(x_1,y_1);S(F)\}$. The matrix $M(\Psi_2)$ is described by

$$M(\Psi_2) = \int \left[\frac{\partial}{\partial T(F)} \Psi_2\{(x_2,y_2);\lambda,T(F)\} \right] dF = \int \begin{pmatrix} x_2 x_2^T & \lambda x_2 \\ \lambda x_2^T & \lambda^2 \end{pmatrix} y_1 \, dF, \tag{20}$$

$$\frac{\partial}{\partial\lambda} \Psi_2\{(x_2,y_2);\lambda,T(F)\} = \begin{pmatrix} 0 \\ y_2 - x_2^T\beta - \lambda\beta_\lambda \end{pmatrix} y_1 - \begin{pmatrix} x_2 \\ \lambda \end{pmatrix} \beta_\lambda y_1, \tag{21}$$

and

$$\lambda' = \frac{\partial}{\partial\beta_1} \lambda\{(x_1,y_1);S(F)\} = -\lambda\{(x_1,y_1);S(F)\} \left[x_1^T\beta_1 \frac{\nu+1}{\nu+(x_1^T\beta_1)^2} + \lambda\{(x_1,y_1);S(F)\} \right]. \tag{22}$$

The score function of the classical second-stage estimator is unbounded. This results in an unbounded second-stage IF.

3.3 Asymptotic Variance

The asymptotic variance is a key component when testing for Sample Selection Bias (SSB). The asymptotic variance is given by (see Hampel et al. (1986)):

$$V(T,F) = \int IF(z;T,F)^2 dF(z). \tag{23}$$

Zhelonkin et al. (2016) derived the asymptotic variance of the robust Heckman two-stage estimator under the assumption of normally distributed errors. Under the assumption of bivariate-t errors the classical asymptotic variance becomes the following

$$V(T,F) = (X^TX)^{-1} \left[\sigma^2 \left\{ X^T \left(I - \frac{\beta_\lambda^2}{\sigma^2} \right) X \right\} + \beta_\lambda^2 X^T \Delta X_1 Var(S,F) X_1^T \Delta X \right] (X^TX)^{-1}, \tag{24}$$

with

$$
\begin{aligned}
a(z) &= \Psi_2((x_2, y_2); \lambda, T(F)), \\
b(z) &= \int \frac{\partial}{\partial \lambda} \Psi_2((x_2, y_2); \lambda, T(F)) \frac{\partial}{\partial \beta_1} \lambda \, dF \cdot IF(z, S, F),
\end{aligned} \tag{25}
$$

where $M(\Psi_2)$ is defined by (20), Δ is a diagonal matrix with elements $\delta_{ii} = \frac{\partial \lambda(x_{1i}\beta_1;\nu)}{\partial(x_{1i}\beta_1)}$, ν the degrees of freedom, the matrix X consisting of vectors $x_i = \binom{x_{2i}}{\lambda_i}$, and Var(S,F) is the asymptotic variance of the robit MLE.

3.4 Sample Selection Bias (SSB) test

Proposed by Heckman (1979) the test for SSB uses a standard t-test for the variable β_λ. The test is defined as $\tau_n = \sqrt{n}\hat{\beta}_\lambda / \sqrt{V(\beta_\lambda, F)}$. Zhelonkin et al. (2016) proposes a method for approximating the bias of the test statistic under contamination using the von Mises expansion (von Mises (1947)) of the test statistic. The test statistic is expressed as

$$
\frac{T(F_\epsilon)}{\sqrt{V(F_\epsilon)/n}} = \frac{T(F)}{\sqrt{V(F)/n}} + \epsilon \left[\frac{IF(z;T,F)}{\sqrt{V(F)/n}} + \frac{1}{2n} T(F) \frac{CVF(z;T,F)}{\{V(F)/n\}^{5/2}} \right] + o(\epsilon). \tag{26}
$$

where F_ϵ is the data distribution with contamination proportion ϵ. Since this expansion depends on the IF and Change-of-Variance Function (CVF), derived in Zhelonkin et al. (2016), it is easily notable that if either of the functions are unbounded, the bias under contamination will also be unbounded. This means that the bias can become arbitrarily large. By robustifying the IF and CVF, which is achieved by robustifying the estimators, an unbounded bias test can be formulated.

4 Robust Two-Step Estimator

4.1 Robust first stage

Let the two-stage M-estimators (13) and (14) be the starting estimators. These can be robustified by bounding their IF's. From Cantoni and Ronchetti (2001): "We consider a general class of M-estimators of Mallows's type, where the influence of deviations on y_1 and on x_1 are bounded separate". The score function for the robust first stage estimator is defined as

$$
\Psi_1^R(z_1, S(F)) = \nu(z_1; \mu_1)\omega_1(x_1)\mu_1' - \alpha(\beta_1), \tag{27}
$$

where $\alpha(\beta_1) = \frac{1}{N}\sum_{i=1}^N E(\nu(z_{1i}; \mu_{1i}))\omega_1(x_{1i})\mu_{1i}'$, $\nu(z_{1i}, \mu_{1i}) = \psi_{c_1}(r_i)\frac{1}{V^{1/2}(\mu_{1i})}$, and $r_i = \frac{y_{1i} - \mu_{1i}}{V^{1/2}(\mu_{1i})}$. For the robit function we have $\mu_{1i} = T(x_{1i}^T\beta_1; \nu)$ with $\mu_{1i}' = \frac{\partial}{\partial \beta_1}\mu_{1i}$ and $V(\mu_{1i}) = P(Y_{1i} = 1) \cdot P(Y_{1i} = 0) = T(x_{1i}^T\beta_1; \nu) \cdot (1 - T(x_{1i}^T\beta_1; \nu))$. The quasi-likelihood estimation equation becomes

$$
\sum_{i=1}^N \left(\psi_c(r_i)\omega_1(x_{1i}) \frac{1}{\{T(x_{1i}^T\beta_1;\nu)(1 - T(x_{1i}^T\beta_1;\nu)\}^{1/2}} t(x_{1i}'\beta_1;\nu)x_{1i} - \alpha(\beta_1) \right) = 0, \tag{28}
$$

with

$$
\psi_{c_1}(r) = \begin{cases} r, & \text{if } |r| \le c_1, \\ c_1 \text{sign}(r), & \text{if } |r| > c_1. \end{cases} \tag{29}
$$

The weight function is $\omega_{1i} = \sqrt{1 - H_{ii}}$ with $H = X_1(X_1^T X_1)^{-1} X_1^T$. "The tuning constant c is typically chosen to ensure a given level of asymptotic efficiency" - Cantoni and Ronchetti (2001). The choice for the tuning constant is $c_1 = 1.345$. As for the constant $\alpha(\beta_1)$, $E[\psi_c(r_i)]$ becomes

$$
\begin{aligned}
E[\psi_{c_1}(r_i)] = E\left[\psi_{c_1}\left(\frac{Y_{1i} - \mu_{1i}}{V^{1/2}(\mu_{1i})}\right)\right] &= \psi_{c_1}\left[\frac{1 - \mu_{1i}}{\{T(x_{1i}^T\beta_1; \nu)(1 - T(x_{1i}^T\beta_1; \nu)\}^{1/2}}\right] T(x_{1i}'\beta_1; \nu) \\
&+ \psi_{c_1}\left[\frac{-\mu_{1i}}{\{T(x_{1i}^T\beta_1; \nu)(1 - T(x_{1i}^T\beta_1; \nu)\}^{1/2}}\right] (1 - T(x_{1i}'\beta_1; \nu)).
\end{aligned}
\tag{30}
$$

4.2 Robust second stage

The robust second stage uses the following M-estimator of Mallows's type

$$
\Psi_2^R(z_2, \lambda, T) = \Psi_{c_2}(y_2 - x_2^T\beta_2 - \lambda\beta_\lambda)\omega_2(x_2, \lambda)y_1, \tag{31}
$$

where $\Psi_{c_2}(.)$ is the classical Huber function with tuning constant c_2, $\omega_2(.)$ is the weight function described in (32), and $d(x_2, \lambda)$ the Mahalanobis distance such that

$$
\omega_2(x_2, \lambda) = \begin{cases} x_2, & \text{if } d(x_2, \lambda) < c_m, \\ \frac{x_2 c_m}{d(x_2, \lambda)}, & \text{if } d(x_2, \lambda) \geq c_m, \end{cases} \tag{32}
$$

The squared Mahalanobis distance follows a χ^2-distribution. The value of c_m is chosen based on the critical value of 5%. Here the tuning constant is $c_2 = 1.345$. The decisions for this weight function and tuning constants are based on the theory of Linear regression, described in Hampel et al. (1986).
In case of possible multicollinearity split the observations into two different sets $(x_2^{(1)}, ..., x_2^{(k)})$ and $(x_2^{(k+1)}, ..., x_2^{(q)}, \lambda)$, such that the weights $\omega(x_2, \lambda) = \omega(x_2^{(1)}, ..., x_2^{(k)})\omega(x_2^{(k+1)}, ..., x_2^{(q)}, \lambda)$ can be computed as the combination of the weights form the two nearly orthogonal sets.

4.3 Estimating ν

When estimating the Heckman Two-step model with bivariate-t errors the parameters can be estimated using the techniques presented in Sections 3 and 4. These estimators, however, do not account for the estimation of the parameter ν. For the numerical studies and the empirical example this parameter is estimated using two different methods.
The first estimation method requires the idea of profile likelihood, proposed in the book of McCullagh and Nelder (1989). Let $\hat{\theta}_\nu$ be the ML estimate of θ_ν for fixed ν. Let $l(\cdot)$ be the (log) likelihood function and y be a set of observations. Then, the profile (log) likelihood for ν is the the partially maximized (log) likelihood function $l^\dagger(\nu; y) = l(\nu, \hat{\theta}_\nu; y) = \sup_\theta l(\nu, \theta; y)$. The best estimate for ν thus becomes the estimate which (given the rest of the parameters) maximizes the profile (log) likelihood. The profile likelihood method is applicable to the Two-step estimation method, since the degrees of freedom is estimated in the first step. The robust estimate of ν is estimated using the Quasi-likelihood of the robust estimator.
The second estimation method requires the use of an EM algorithm. Davila et al. (2020) proposed an EM algorithm for the MLE of the SLt model. The values for ν are estimated using this EM algorithm, after which the Two-stage estimator is applied for the rest of the parameters. Note that since the score function for the MLE of ν is unbounded, it is possible to get biased estimates for the Two-stage estimators if the wrong starting values for the EM algorithm are chosen.

4.4 EIF

Let $x \in \mathbb{R}^N$ be a vector of N observations and $\theta \in \mathbb{R}^p$ be a vector of p parameters. Let F be a probability distribution on $\mathbb{R}^p$ and $\hat{\theta}$, an estimator of θ, be a function of the distribution F. Let $\tilde{x} \in \mathbb{R}$ be an observation. Then function for the change of $\tilde{x}$ on the estimator $\hat{\theta}$ is

$$\text{EIF}_i(\tilde{x}; \hat{\theta}, x_N) = N\left\{\hat{\theta}(x_1, ..., x_{i-1}, \tilde{x}, x_{i+1}, ..., x_N) - \hat{\theta}(x_1, ..., x_N)\right\}$$

where $i = 1, ..., N$. If $x_i \sim F$ then $\text{EIF}_i(\tilde{x}; \hat{\theta}, x_N)$ should be close to the IF of the estimator $\hat{\theta}$ for large N. The EIF can be used as an approximate for the IF for the classical and robust two-stage estimators. The EIF will be used in the sensitivity analysis in section 5.2.

4.5 Robust SSB Test

The robust SSB test ($H_0 : \beta_\lambda = 0$ vs $H_A : \beta_\lambda \neq 0$) is a simple t-test is performed on the robust estimates of β_λ and its robust standard errors. The Heckman's two-stage estimator defined in sections (4.1) and (4.2) is robust, consistent, and asymptotically normal, with asymptotic variance

$$V(T, F) = M(\Psi_2^R)^{-1} \int \{a_R(z)a_R(z) + b_R(z)b_R(z)\} dF M(\Psi_2^R)^{-1}, \tag{33}$$

where $M(\Psi_2^R) = -\int \frac{\partial}{\partial \beta_2}\Psi_2^R(z; \lambda, T)dF$, $b_R(z) = \int \frac{\partial}{\partial \lambda}\Psi_2^R(z; \lambda, T)\frac{\partial}{\partial \beta_1}dF\left\{\int \Psi_1^R(z; S)dF\right\}^{-1}\Psi_1^R(z; S)$, and $a_R(z) = \Psi_2^R(z; \lambda, T)$. The first term of the asymptotic variance is comparable to the the asymptotic variance of a standard linear regression with heteroscedasticity. This part can be estimated using the Eicker-Huber-White Heteroskedasticity Variance Estimator (Eicker (1967), Huber et al. (1967), and White (1980)). The second term contains the asymptotic variance of the robit MLE. Consistent estimators for the first and second term (respectively) are:

$$\begin{aligned} &\hat{M}(\Psi_2^R)^{-1}\frac{1}{n}\sum \hat{a}_R(z_i)\hat{a}_R(z_i)^T \hat{M}(\Psi_2^R)^{-1}, \text{ and} \\ &\hat{M}(\Psi_2^R)^{-1}\frac{1}{n}\sum \hat{b}_R(z_i)\hat{b}_R(z_i)^T \hat{M}(\Psi_2^R)^{-1} = \hat{M}(\Psi_2^R)^{-1}\frac{1}{n}\sum \frac{\partial \Psi_{2i}^R}{\partial \beta_1}\hat{Var}(S, F)\left(\frac{1}{n}\sum \frac{\partial \Psi_{2i}^R}{\partial \beta_1}\right)^T \hat{M}(\Psi_2^R)^{-1}, \end{aligned} \tag{34}$$

where $\hat{M}$ and $\hat{a}(z)$ are the sample versions of M and $a(z)$ respectively and $\frac{\partial \Psi_{2i}^R}{\partial \beta_1} = \frac{\partial \Psi_2(z_{2i}; \lambda, T(F))}{\partial \lambda}\frac{\partial \lambda(z_{1i}; S(F))}{\partial \beta_1}$.

5 Simulation Study

5.1 Simulation 1: Sample Selection bias

A Monte Carlo Simulation study is done to exhibit the robustness issues of the models defined in Section 4. In this study the following specifications are generated:

$$y_{ji}^* = \beta_{j1}x_{j1i} + \beta_{j2}x_{j2i} + \beta_{j3}x_{j3i} + \varepsilon_{ji} \tag{35}$$

where i = 1,...,N and j = 1,2. The selection model is generated when $j = 1$. The model of interest (primary equation) is generated when $j = 2$. The number of observations generated is N = 1000. For bigger sample sizes the results of the estimators remain similar. When the exclusion restriction is unavailable the variables for both equations are the same, i.e. $x_1 = x_2$. When this exclusion restriction is available the variable

x_{23i} is generated independently. The data is generated as follows: $x_{11i} \sim N(0,1)$, $x_{12i} \sim N(-2, 0.1)$, $x_{13i} \sim N(2, 1.5)$. For the selection equation the parameters are $\beta_1 = (0.5, 0.75, 1)^T$. For the primary equation the parameters are $\beta_2 = (1.25, 1, 0.75)^T$. In this study $\beta_{10} = 0$ and $\beta_{20} = 0$. The value of the selection equation intercept coincides to a degree of censoring of roughly 55%. The errors ε_1 and ε_2 are bivariate-t distributed with expectation 0, $\sigma_1 = 1, \sigma_2 = 1, \rho = 0.6$. So $\beta_\lambda = 0.6$. The degrees of freedom $\nu = 3$ is used. The number of Monte Carlo replications is 500.

This simulation study comprises of 3 different cases. Case I contains simulations with uncontaminated data. Case II contains contaminated data in x_1. The data is generated as described above and replaced with probability $p = 0.01$ at a point mass $(x_{11}, x_{12}, x_{13}, y_1, y_2) = (3, -1, 5, 1, 0)$. Here the effects of leverage outliers are studied when they are not present in the primary equation. Case III is generated as described above and replaced with probablity $p = 0.01$ at a point mass $(x_{11}, x_{12}, x_{13}, y_1, y_2) = (-3, -3, -1, 0, 1)$. Here the effects of leverage outliers are studied when they are present in the primary equation.

I compare 2 different first-stage estimators: the classical robit, and the robust robit (Section 4.1). I compare 3 different second-stage estimators in this study. The first second-stage estimator is the Classical estimator, where the first stage is estimated using the robit function and the second stage is estimated using OLS. The second second-stage estimator is the Robust 1S estimator, where the first stage is estimated using the robust robit function and the second stage is estimated using OLS. The third second-stage estimator is the Robust 2S estimator, where the first- and second-stage estimators are the robust estimators described in Sections 4.1 and 4.2.

Table 1 shows the average bias, variance, and MSE for ν estimated from the profile likelihood. Figure 2 shows the boxplots of the estimated ν. The robust robit model has low bias when no contamination is present. Under contamination the robust model becomes more biased, but performs better than the classical estimator. With the exclusion the bias, variance, and MSE increased slightly. Table 2 shows the bias, variance, and MSE for the EM algorithm ν. Figure 3 shows the boxplots for the estimated ν. The estimator has low bias under contamination. The bias is lowest with the exclusion restriction when the contamination is only in the first stage. Under contamination ν estimated using the EM algorithm has overall lower bias, variance, and MSE than ν estimated from the profile likelihood. However, the EM algorithm estimates are dependent on the starting value. It did not obtain an estimate lower than the true parameter, as shown in the boxplots, and is therefore still a biased estimator.

Table 1: Bias, Variance, MSE for ν estimated from the profile likelihood.

	No Contamination			x_1 Contaminated, $y_1 = 0$			x_1 Contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
With ex. restriction									
Classical robit	0.874	6.066	12.883	-1.499	0.001	2.247	-1.495	0.003	2.238
Robust robit	0.063	0.302	0.608	0.427	0.559	1.300	0.243	0.390	0.838
Without ex. restriction									
Classical robit	1.077	7.010	15.166	-1.497	0.002	2.245	-1.490	0.005	2.231
Robust robit	0.098	0.315	0.640	0.459	0.549	1.308	0.269	0.398	0.867

Table 2: Bias, Variance, MSE for ν estimated by EM

	With exclusion res.			Without exclusion res.		
	Bias	Var	MSE	Bias	Var	MSE
No contamination	0.248	0.125	0.311	0.327	0.175	0.456
Contamination, y1=0	0.008	0.003	0.005	0.040	0.018	0.037
Contamination, y1=1	0.026	0.013	0.028	0.015	0.005	0.010

Tables 4 and 5 show the bias, variance, and MSE of the first stage estimators, where ν was estimated using the profile likelihood. Figures 4 - 6 show the boxplots of the estimates. The bias the first stage robust robit estimator remains low for all cases. Without the exclusion restriction, the classical robit estimator has lower bias, but still performs poorly compared to the robust robit estimator. Tables 6 and 7 show the results of the second stage estimators, where ν was estimated using the profile likelihood. With the exclusion restriction and no contamination second stage estimators have low bias, variance, and MSE. The bias for the models increases as contamination gets introduced, with the exception for the Robust 2S estimator in the case of contamination with $y_1 = 0$. Without the exclusion restriction the second-stage estimators quickly become more unbiased. With contamination in the second stage the Robust 2S has the performs the best. The IMR estimates of the Robust 2S estimator have the lowest bias under contamination. Tables 8 and 9 show the results of the first stage estimators where ν was estimated using the EM algorithm. Figures 7 - 9 show boxplots of the estimates. Here the estimators with the exclusion restriction have slightly higher bias. With and without contamination the robust robit has the lowest overall bias. Tables 10 and 11 show the results of the second stage estimators where ν was estimated using the EM algorithm. With the exclusion restriction and no contamination, all second-stage estimators have low bias, variance, and MSE. Under contamination in the selection equation the bias increases slightly, with the exception of β_λ for Robust 1S. When contamination takes place in the primary equation the bias increases for the Classical and Robust 1S estimators. The Robust 2S estimator retains low bias. Without the exclusion restriction and no contamination all estimators become more unbiased. With the introduction of contamination this bias increases for all estimators. The Robust 2S estimator is the least affected, but is still unbiased.
In the case of uncontaminated data the Robust 2S estimator has low bias with the exclusion restriction. However, under contamination this estimator performs better than the other estimators.

5.2 Simulation 2: Sensitivity Analysis

A Monte Carlo Simulation study is done to exhibit robustness issues of the score and Influence functions models defined in section 3.2. In this study the following models are generated:

$$y^*_{ji} = \beta_{j1} x_{j1i} + \varepsilon_{ji}, \tag{36}$$

where i = 1,...,N and j=1,2. The selection model is generated when j = 1. The primary model is generated when j = 2. The number of observations is 1000. Similar to the previous simulation study, comparisons are also made between models with and without the exclusion restriction. The data is generated as follows: $x_{11i} \sim N(1, 1)$ and $x_{21i} \sim N(1, 1)$. For the selection equation the parameter used is $\beta_{11} = -1$. For the primary equation the parameter is $\beta_{21} = 1$. Similar to the previous simulation study $\beta_{10} = \beta_{20} = 0$. The errors of both equations are bivariate-t distributed with expectation 0, $\sigma_1 = 1, \sigma_2 = 1$, and $\rho = 0.5$. So, $\beta_\lambda = 0.5$. The degrees of freedom for the distribution of the errors is $\nu = 3$. The number of Monte Carlo simulations is 500.
This simulation study comprises of 3 different cases. In the first case (case 1) the variable x_{1k} with $k \in [1, N]$ gets replaced by values between -15 and 15. This shows the bias and influence that leverage outliers have on the two-step estimators when transferred through the selection equation. In the second case (case 2) the variable x_{2k} gets replaced by values between -15 and 15. This shows the bias and influence that leverage outliers have when they are present in the primary equation. In the third case (case 3) the variable y_{2k} gets replaced by values between -15 and 15. The results are in the Appendix.
Figures 10, 11, and 12 show boxplots of the EIF values of the first-stage estimators with the exclusion restriction for cases 1, 2, and 3 respectively. In case 1 we see that there is some influence in the estimators. However, we see that the classical first-stage estimator is unbounded as the EIF values become arbitrarily large as x_{1k} increases. The robust first-stage estimator is bounded (this is also shown in its score function). In cases 2 and 3 the first-stage estimators are not influenced, thus the results remain the same for each

change in observation. Figures 16-19 show the boxplots of the EIF values of the first-stage estimators wtihout the exclusion restriction for cases 1, 2. and 3 respectively. These results are similar to the results with the addition of the exclusion restriction.
Figures 13, 14, and 15 show boxplots of the EIF values of the second-stage estimators with the exclusion restriction for cases 1, 2, and 3 respectively. In case 1 the EIF values of the classical estimator increase as x_{1k} increases. The change in observation has some influence on the robust estimator, but it remains bounded. Case 2 shows more the influence of the unbounded classical estimator. The change in EIF values indicates that for large positive and negative changes in x_{2k} the bias can become arbitrarily large. As for the robust estimator, there is more influence than in case 1, but the results show that the robust estimator is bounded. Case 3 , again, shows the influence on the unbounded classical and bounded robust second-stage estimators. For large changes in observation y_{2k} the bias of the classical estimator can become very large. As for the robust estimator, again, there is some influence, but the results still indicate a bounded estimator. Figures 19, 20, and 21 show boxplots of the EIF values of the second-stage estimators without the exclusion restriction for cases 1, 2, and 3 respectively. These results are similar to the results with the addition of the exclusion restriction.
This sensitivity analysis shows that the robust two-stage estimator is bounded and thus not (heavily) influenced by outliers. When the exclusion restriction is not available this influence increases.

6 Ambulatory Expenditures Example

To demonstrate the behavior of the robust Two-Step estimator with bivariate-t errors, the methodology is implemented on the 2001 Medical Expenditure Panel Survey, analyzed by Cameron and Trivedi (2009). The survey consist of 3385 observations, with 526 observations citing 0 expenditure. The expenditure variable has a skewed distribution. In this case, the log scale of the variable is used. For the selection equation the following explanatory variables are used: age, insurance status (ins), the total number of chronic diseases (totchr), ethnicity (blhisp), number of years of education (educ), and gender (female). For the outcome equation, the same explanatory variables are used. In case of the addition of the exclusion restriction, the variable income (inc) is added to the explanatory variables of the selection equation. The values for the degrees of freedom are estimated using the profile likelihood.
The estimation results of the Ambulatory Expenditures data are reported in Table 3. For comparison, the estimation results under normally distributed errors are reported in Table 12 in the Appendix. The results of the Classical two-stage estimates show that both with and without the exclusion restriction the variables for the decision to spend are significant. As for estimating the expenditure, the variables education and insurance status are not significant in both cases.
The SSB t-tests for the classical estimator have p-values 0.001 (with ex. restriction) and 0.109 (without ex. restriction). Only with the exclusion restriction is this test significant at the 5% level. Theoretically this indicates that there is a presence of sample selectivity when the exclusion restriction is present. Without the exclusion restriction the SSB t-test for the classical estimator is not significant. This non-significance indicates that there is an absence of selectivity when the exclusion restriction is not present. This is in line with the conclusion of Zhelonkin et al. (2016) and Marchenko and Genton (2012).
The robust two-stage estimates show that for both with and without the exclusion restriction the pattern for the decision to spend is similar. The difference here is that the estimate for income is significant at a 10% level. The SSB t-tests for the robust estimator have p-values 0.002 (with ex. restriction) and 0.070 (without ex. restriction). The tests show significance at the 5% level with the exclusion restriction and at the 10% level without the exclusion restriction. The SSB t-test for the robust model indicates that there is a presence of selectivity at the 10% significance level. When using a significance level of 5% the test indicates that there is an absence of selectivity when the exclusion restriction is not included.
The estimation results of the IMR variable differs depending on the error distribution. This is then reflected in the p-values of the SSB-t tests. The SSB-t test of the IMR (of the classical estimator) under the t-distribution with the exclusion restriction is significant, compared to under the normal distribution. The

SSB-t test of the IMR (of the robust estimator) under the t-distribution without the exclusion restriction is only significant at the 10% significance level, compared to under the normal distribution. These results further shows the concerns of Cameron and Trivedi (2009) about robustness problems when searching for sample selectivity.

Table 3: Estimation result of the ambulatory expenditures data for the classical and robust Two-stage estimators. ν is estimated using the profile-likelihood. The standard errors are given in the parenthesis. The significance codes are ”***” 0.001, ”**” 0.01, ”*” 0.05, ”·”0.10.

	With ex. restriction		Without ex. restriction	
	Classical	Robust	Classical	Robust
ν	3.979	2.208	3.625	2.208
Selection				
Intercept	-0.772	-0.822	-0.821	-0.853
	(0.211)*	(0.224)*	(0.210)**	(0.223)*
age	0.099	0.113	0.111	0.124
	(0.031)*	(0.033)*	(0.030)*	(0.032)*
female	0.752	0.803	0.733	0.793
	(0.069)***	(0.075)**	(0.068)***	(0.075)**
educ	0.067	0.069	0.076	0.074
	(0.013)**	(0.014)*	(0.012)**	(0.013)*
blhisp	-0.411	-0.440	-0.422	-0.452
	(0.068)**	(0.072)*	(0.068)**	(0.072)**
totchr	0.938	0.962	0.937	0.966
	(0.089)***	(0.106)**	(0.089)***	(0.107)*
ins	0.185	0.178	0.197	0.190
	(0.070)*	(0.075)·	(0.069)*	(0.075)·
income	0.003	0.002	-	-
	(0.001)·	(0.002)		
Outcome				
Intercept	5.183	6.356	5.223	5.328
	(0.521)***	(0.448)**	(0.271)***	(0.270)***
age	0.179	0.167	0.205	0.205
	(0.045)**	(0.039)*	(0.024)***	(0.024)**
female	0.347	0.096	0.308	0.279
	(0.131)*	(0.113)	(0.068)**	(0.068)*
educ	0.023	0.008	0.015	0.016
	(0.021)	(0.018)	(0.011)	(0.011)
blhisp	-0.224	-0.102	-0.194	-0.171
	(0.119)·	(0.102)	(0.063)*	(0.063)·
totchr	0.533	0.235	0.511	0.492
	(0.088)**	(0.076)*	(0.046)***	(0.046)**
ins	0.024	-0.053	-0.040	-0.060
	(0.098)	(0.084)	(0.052)	(0.052)
IMR	-3.017	-3.909	-0.303	-0.462
	(0.360)***	(0.309)**	(0.204)	(0.204)·

7 Conclusion

This paper adapts the classical and robust Heckman selection model to the bivariate-t distribution. The IF for the classical and robust estimators are constructed under this error distribution. A Monte Carlo study on the bias, variance, and MSE shows the good performance of the robust two-stage estimators with different types of contamination. Under contamination the robust second-stage estimator produces the lowest sample selection bias. A Monte Carlo study on the EIF shows the marginal effects on the robust two-stage estimator. The boundedness of the robust score functions are reflected in the results of the simulation. An analysis of the ambulatory expenditures data shows that, under bivariate-t errors, there is presence of selectivity, according to the significant SSB-t test. Although an estimator for the degrees of freedom is not constructed, the profile likelihood method and EM algorithm (for the MLE model) is used to approximate a value for this parameter. Both ways of estimating ν produces low bias for the two-stage estimators. However, the score functions for the MLE are unbounded and estimator based on the MLE can produce bias in the sample selection model. Constructing an estimator for the degrees of freedom for the sample selection model can be a subject for further research.

References

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, MA.

Cameron, C. A. and Trivedi, P. K. (2009). *Microeconometrics Using Stata*. College Station, TX: Stata Press.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96:1022–1030.

Davila, V. H. L., Prates, M. O., and Dey, D. K. (2020). Heckman selection-t model: parameter estimation via the em-algorithm.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. 1(1):59–82.

Hampel, F., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley and Sons.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

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

Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. 1(1):221–233.

Koenker, R. and Yoon, J. (2009). Parametric links for binary choice models: a fisherian–bayesian colloquy. *Journal of Econometrics*, 152:120–130.

Liu, C. (2005). Robit regression: A simple robust alternative to logistic and probit regression. pages 227 – 238.

Marchenko, Y. V. and Genton, M. G. (2012). A Heckman selection-t model. *Journal of the American Statistical Association*, 107:304–317.

McCullagh and Nelder (1989). *Generlalized Linear Models*, volume 2nd. Chapman Hall, London, 2nd edition.

Paarsch, H. J. (1984). A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24:197–213.

Serrudo, L. S. (2008). A robustness study of heckman's model.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics*, 18:309–348.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838.

Zhao, J., Kim, H.-J., and Kim, H.-M. (2020). New em-type algorithms for the heckman selection model. *Computational Statistics Data Analysis*, 146:106930.

Zhelonkin, M., Genton, M. G., and Ronchetti, E. (2016). Robust inference in sample selection models. *Journal of the Royal Statistical Society Series B*, 78:805–827.

Appendix

Results for the Two-stage estimators, where ν was estimated using the profile likelihood

Table 4: Bias, Variance, MSE of classical and robust models, first-stage, with the exclusion restriction, profile

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical Robit									
β_{10}	-0.003	0.056	0.056	-0.003	0.056	0.056	-0.003	0.056	0.056
β_{11}	0.051	0.010	0.013	0.051	0.010	0.013	0.051	0.010	0.013
β_{12}	0.118	0.016	0.030	0.118	0.016	0.030	0.118	0.016	0.030
β_{13}	0.177	0.012	0.043	0.177	0.012	0.043	0.177	0.012	0.043
Robust Robit									
β_{10}	-0.001	0.040	0.040	-0.001	0.040	0.040	-0.001	0.040	0.040
β_{11}	-0.018	0.008	0.008	-0.018	0.008	0.008	-0.018	0.008	0.008
β_{12}	-0.018	0.014	0.014	-0.018	0.014	0.014	-0.018	0.014	0.014
β_{13}	-0.019	0.013	0.014	-0.019	0.013	0.014	-0.019	0.013	0.014

Table 5: Bias, Variance, MSE of classical and robust models, first-stage, without the exclusion restriction, profile

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical Robit									
β_{10}	-0.002	0.041	0.041	-0.028	0.057	0.057	-0.003	0.056	0.056
β_{11}	0.021	0.012	0.012	0.060	0.010	0.013	0.051	0.010	0.013
β_{12}	0.030	0.023	0.024	0.124	0.016	0.031	0.118	0.016	0.030
β_{13}	0.041	0.029	0.030	0.181	0.012	0.045	0.177	0.012	0.043
Robust Robit									
β_{10}	-0.000	0.043	0.043	-0.005	0.039	0.039	-0.001	0.040	0.040
β_{11}	0.013	0.008	0.008	-0.015	0.008	0.008	-0.018	0.008	0.008
β_{12}	0.019	0.016	0.016	-0.018	0.014	0.014	-0.018	0.014	0.014
β_{13}	0.025	0.015	0.015	-0.020	0.013	0.014	-0.019	0.013	0.014

Table 6: Bias, Variance, MSE of classical and robust models, second-stage. With the exclusion restriction, profile

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical									
β_{20}	-0.004	0.043	0.043	-0.074	0.048	0.053	-0.167	0.044	0.072
β_{21}	-0.001	0.005	0.005	-0.014	0.005	0.005	-0.024	0.005	0.006
β_{22}	-0.007	0.009	0.009	-0.019	0.009	0.010	-0.003	0.010	0.010
β_{23}	-0.001	0.002	0.002	-0.000	0.002	0.002	-0.013	0.002	0.002
β_{λ}	-0.018	0.055	0.055	-0.298	0.010	0.099	-0.171	0.004	0.033
Robust 1S									
β_{20}	-0.002	0.042	0.042	-0.006	0.043	0.043	-0.101	0.044	0.054
β_{21}	-0.000	0.005	0.005	-0.000	0.005	0.005	-0.018	0.005	0.006
β_{22}	-0.005	0.009	0.009	-0.004	0.009	0.009	0.018	0.010	0.010
β_{23}	-0.000	0.002	0.002	-0.000	0.002	0.002	-0.023	0.002	0.002
β_{λ}	-0.005	0.040	0.040	0.032	0.044	0.046	0.358	0.033	0.161
Robust 2S									
β_{20}	-0.056	0.026	0.029	-0.054	0.026	0.029	-0.069	0.026	0.031
β_{21}	0.002	0.003	0.003	0.002	0.003	0.003	0.001	0.003	0.003
β_{22}	0.001	0.005	0.005	0.002	0.005	0.005	0.003	0.005	0.005
β_{23}	0.000	0.001	0.001	0.001	0.001	0.001	-0.000	0.001	0.001
β_{λ}	-0.030	0.020	0.021	-0.011	0.021	0.021	0.018	0.020	0.020

Table 7: Bias, Variance, MSE of classical and robust models, second-stage. Without the exclusion restriction, profile

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical									
β_{20}	-0.028	0.157	0.158	-0.170	0.212	0.241	-0.797	0.126	0.762
β_{21}	0.003	0.008	0.008	0.002	0.008	0.008	0.045	0.007	0.009
β_{22}	0.001	0.017	0.017	0.017	0.020	0.020	0.135	0.013	0.032
β_{23}	0.006	0.014	0.014	0.034	0.018	0.019	0.211	0.009	0.053
β_{λ}	0.043	0.141	0.143	0.149	0.185	0.207	0.788	0.063	0.684
Robust 1S									
β_{20}	-0.020	0.154	0.154	-0.075	0.161	0.166	-0.864	0.144	0.891
β_{21}	0.002	0.008	0.008	0.011	0.008	0.008	0.075	0.008	0.014
β_{22}	-0.001	0.017	0.017	0.012	0.017	0.017	0.166	0.015	0.043
β_{23}	0.004	0.014	0.014	0.021	0.014	0.015	0.247	0.011	0.072
β_{λ}	0.036	0.134	0.136	0.100	0.142	0.152	0.933	0.095	0.965
Robust 2S									
β_{20}	-0.144	0.102	0.122	-0.141	0.104	0.123	-0.215	0.106	0.153
β_{21}	0.015	0.005	0.005	0.018	0.005	0.005	0.027	0.005	0.006
β_{22}	0.020	0.009	0.009	0.022	0.009	0.010	0.037	0.009	0.011
β_{23}	0.028	0.009	0.009	0.030	0.009	0.010	0.049	0.009	0.011
β_{λ}	0.075	0.081	0.086	0.071	0.081	0.086	0.160	0.082	0.107

Results for the Two-stage estimators, where ν was estimated using the EM algorithm

Table 8: Bias, Variance, MSE of classical and robust models, first-stage, with the exclusion restriction, EM

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical Robit									
β_{10}	0.005	0.033	0.033	-0.034	0.029	0.030	0.004	0.028	0.028
β_{11}	-0.002	0.005	0.005	-0.106	0.005	0.017	-0.117	0.006	0.019
β_{12}	-0.004	0.009	0.009	-0.105	0.008	0.019	-0.110	0.008	0.020
β_{13}	-0.007	0.006	0.006	-0.120	0.006	0.020	-0.123	0.006	0.021
Robust Robit									
β_{10}	0.002	0.037	0.037	-0.003	0.037	0.037	0.001	0.036	0.036
β_{11}	0.001	0.005	0.005	-0.009	0.006	0.006	-0.017	0.006	0.006
β_{12}	0.000	0.010	0.010	-0.009	0.010	0.010	-0.017	0.010	0.010
β_{13}	-0.000	0.007	0.007	-0.010	0.007	0.007	-0.019	0.007	0.007

Table 9: Bias, Variance, MSE of classical and robust models, first-stage, without the exclusion restriction, EM

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical Robit									
β_{10}	-0.003	0.036	0.036	-0.041	0.031	0.033	-0.003	0.032	0.032
β_{11}	-0.005	0.006	0.006	-0.108	0.006	0.018	-0.117	0.006	0.020
β_{12}	-0.009	0.010	0.010	-0.109	0.009	0.020	-0.112	0.009	0.021
β_{13}	-0.011	0.006	0.006	-0.122	0.006	0.021	-0.121	0.006	0.021
Robust Robit									
β_{10}	-0.000	0.040	0.040	-0.006	0.040	0.040	-0.001	0.040	0.040
β_{11}	-0.001	0.006	0.006	-0.010	0.006	0.006	-0.017	0.006	0.006
β_{12}	-0.001	0.011	0.011	-0.010	0.011	0.011	-0.016	0.011	0.011
β_{13}	-0.001	0.007	0.007	-0.009	0.007	0.007	-0.016	0.007	0.007

Table 10: Bias, Variance, MSE of classical and robust models, second-stage. With the exclusion restriction, EM

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical									
β_{20}	-0.003	0.043	0.043	-0.056	0.047	0.050	-0.180	0.048	0.080
β_{21}	0.001	0.005	0.005	-0.011	0.005	0.005	-0.041	0.006	0.007
β_{22}	-0.004	0.009	0.009	-0.005	0.009	0.010	0.015	0.010	0.010
β_{23}	-0.000	0.002	0.002	-0.000	0.002	0.002	-0.025	0.002	0.003
β_{λ}	0.015	0.033	0.034	0.056	0.040	0.043	0.461	0.023	0.236
Robust 1S									
β_{20}	-0.001	0.042	0.042	-0.011	0.044	0.044	-0.104	0.044	0.055
β_{21}	0.000	0.005	0.005	-0.001	0.005	0.005	-0.017	0.005	0.006
β_{22}	-0.005	0.009	0.009	-0.005	0.009	0.009	0.018	0.010	0.010
β_{23}	-0.000	0.002	0.002	-0.000	0.002	0.002	-0.022	0.002	0.002
β_{λ}	0.012	0.033	0.034	0.004	0.034	0.034	0.331	0.014	0.123
Robust 2S									
β_{20}	-0.051	0.026	0.028	-0.060	0.027	0.030	-0.072	0.026	0.032
β_{21}	0.002	0.003	0.003	0.002	0.003	0.003	0.001	0.003	0.003
β_{22}	0.001	0.005	0.005	0.001	0.005	0.005	0.003	0.005	0.005
β_{23}	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001
β_{λ}	-0.020	0.019	0.020	-0.029	0.019	0.019	0.007	0.017	0.017

Table 11: Bias, Variance, MSE of classical and robust models, second-stage. Without the exclusion restriction, EM

	No contamination			x_1 contaminated, $y_1 = 0$			x_1 contaminated, $y_1 = 1$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
Classical									
β_{20}	-0.032	0.159	0.160	-0.173	0.212	0.242	-0.795	0.125	0.756
β_{21}	0.004	0.008	0.008	0.002	0.008	0.008	0.045	0.007	0.009
β_{22}	0.002	0.017	0.017	0.018	0.020	0.020	0.134	0.013	0.031
β_{23}	0.008	0.014	0.014	0.035	0.018	0.019	0.211	0.009	0.053
β_{λ}	0.054	0.140	0.143	0.155	0.185	0.210	0.784	0.061	0.675
Robust 1S									
β_{20}	-0.024	0.155	0.156	-0.076	0.161	0.167	-0.863	0.143	0.888
β_{21}	0.003	0.008	0.008	0.011	0.008	0.008	0.075	0.008	0.014
β_{22}	0.001	0.017	0.017	0.013	0.017	0.018	0.166	0.015	0.043
β_{23}	0.006	0.014	0.014	0.022	0.014	0.015	0.246	0.011	0.071
β_{λ}	0.046	0.134	0.136	0.104	0.143	0.154	0.929	0.094	0.957
Robust 2S									
β_{20}	-0.144	0.102	0.123	-0.141	0.104	0.124	-0.215	0.106	0.153
β_{21}	0.016	0.005	0.005	0.019	0.005	0.005	0.027	0.005	0.006
β_{22}	0.021	0.009	0.009	0.022	0.009	0.010	0.036	0.009	0.011
β_{23}	0.029	0.009	0.010	0.030	0.009	0.010	0.049	0.009	0.011
β_{λ}	0.082	0.084	0.091	0.074	0.082	0.088	0.159	0.081	0.106

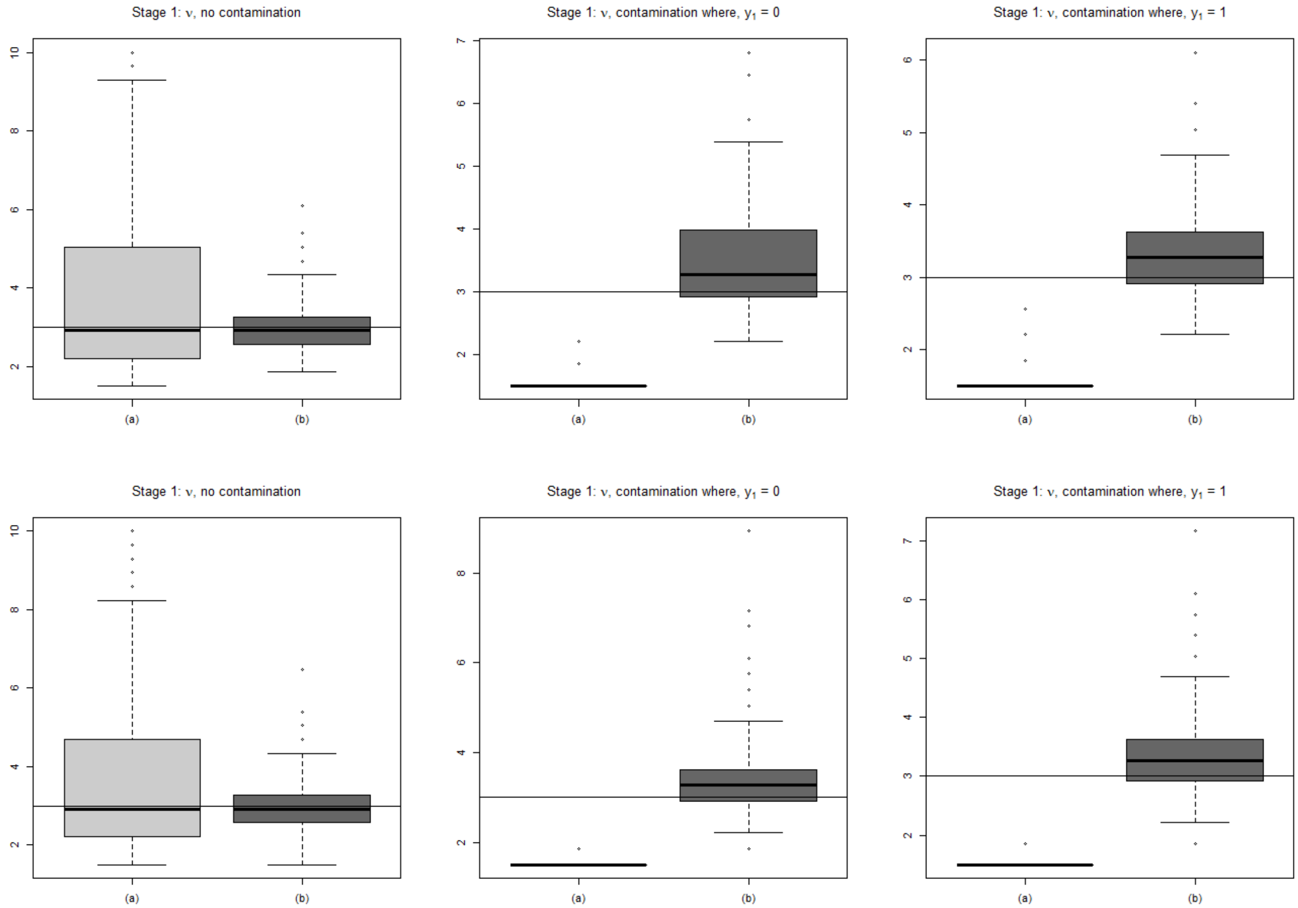

Figure 2: Boxplots of the estimates of ν from the profile likelihoods. The top boxplots contain the exclusion restriction. The bottom plots do not contain exclusion restriction. (a) are the values gained from the classical robit function. (b) are the values gained from the robust robit function.

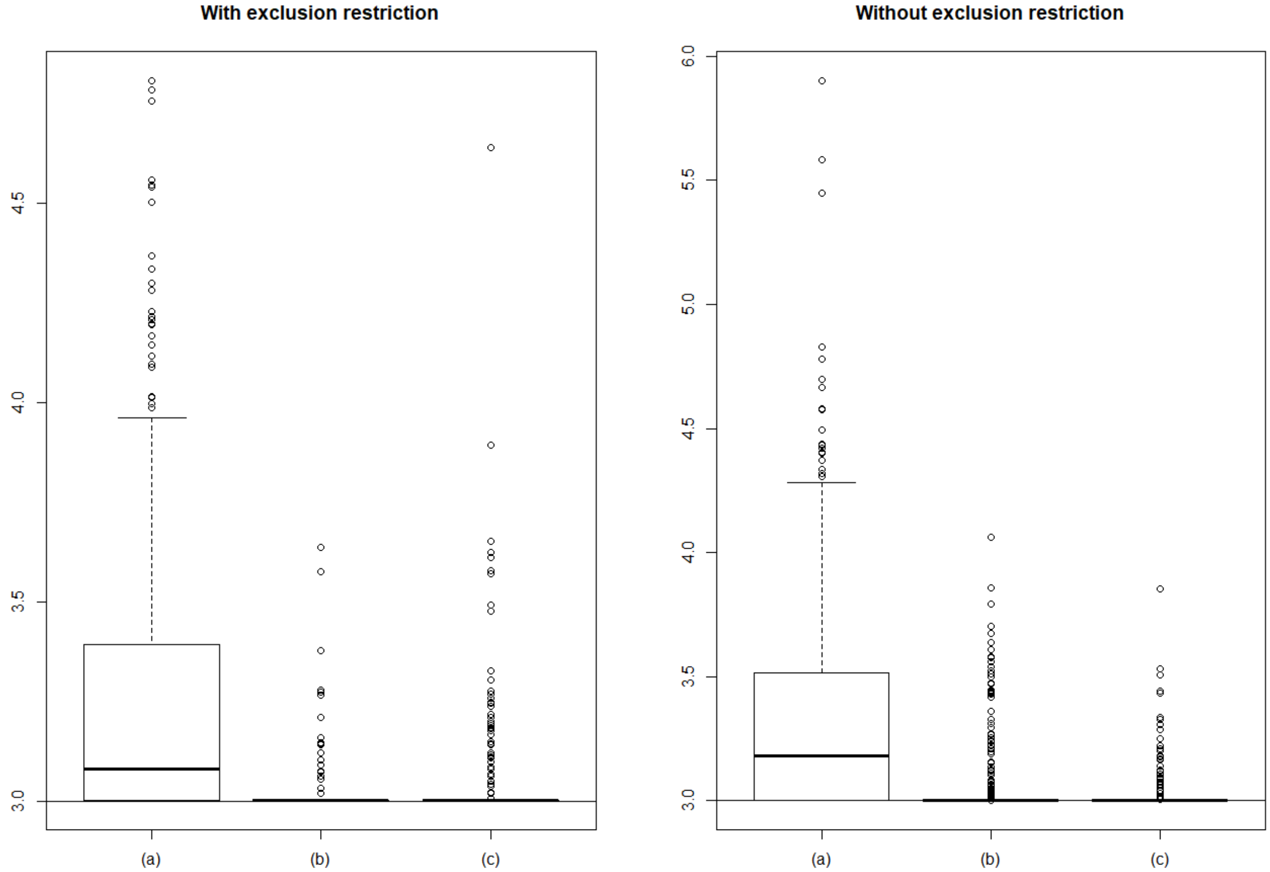

Figure 3: Boxplots of the MLE estimates of ν from the EM algorithm. The left boxplots contain the exclusion restriction. The right boxplots do not have this restriction. (a) contains estimations from uncontaminated data. (b) contains estimations contaminated data, with $y_1 = 0$. (c) contains estimations from contaminated data, with $y_1 = 1$.

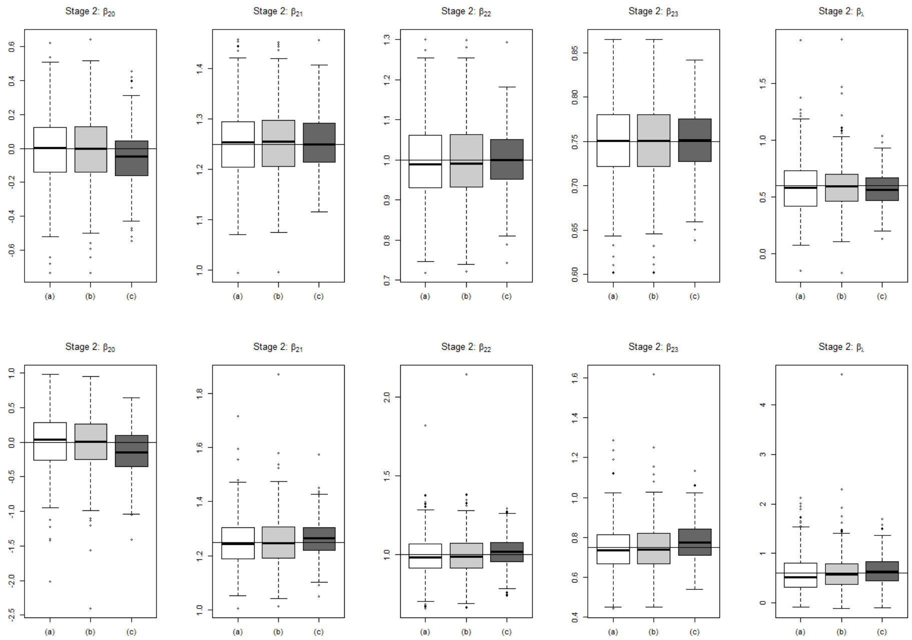

Figure 4: Boxplots of the parameters of the second stage with no contamination. The values of ν are gained based on the profile likelihood. The top panel shows the parameters with no exclusion restriction. The bottom panel shows the parameters with the exclusion restriction. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

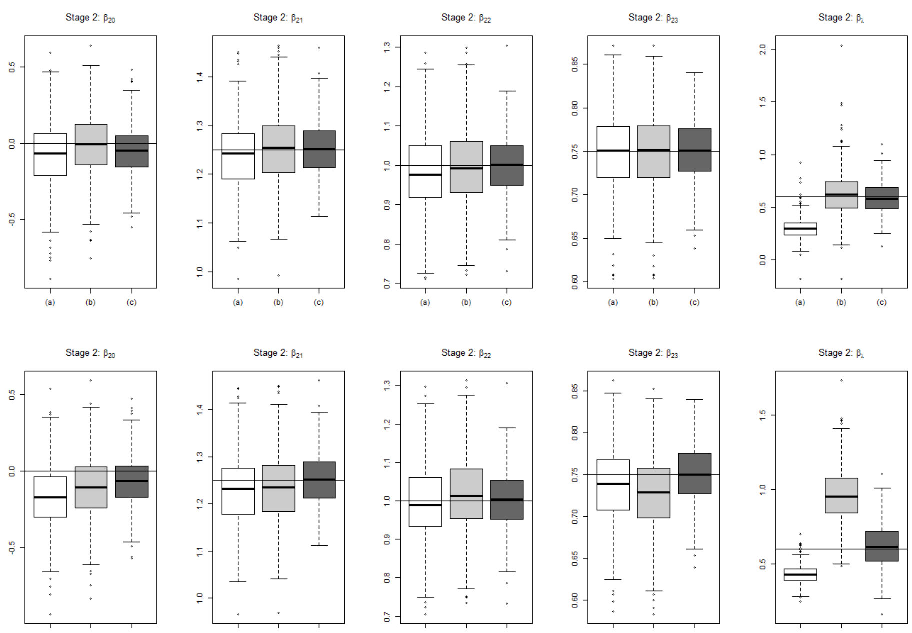

Figure 5: Boxplots of the parameters of the second stage without the exclusion restriction. The values of ν are gained based on the profile likelihood. The top panel shows the parameters with contaminated x_1 and $y_1 = 0$. The bottom panel shows the parameters with contaminated x_1 and $y_1 = 1$. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

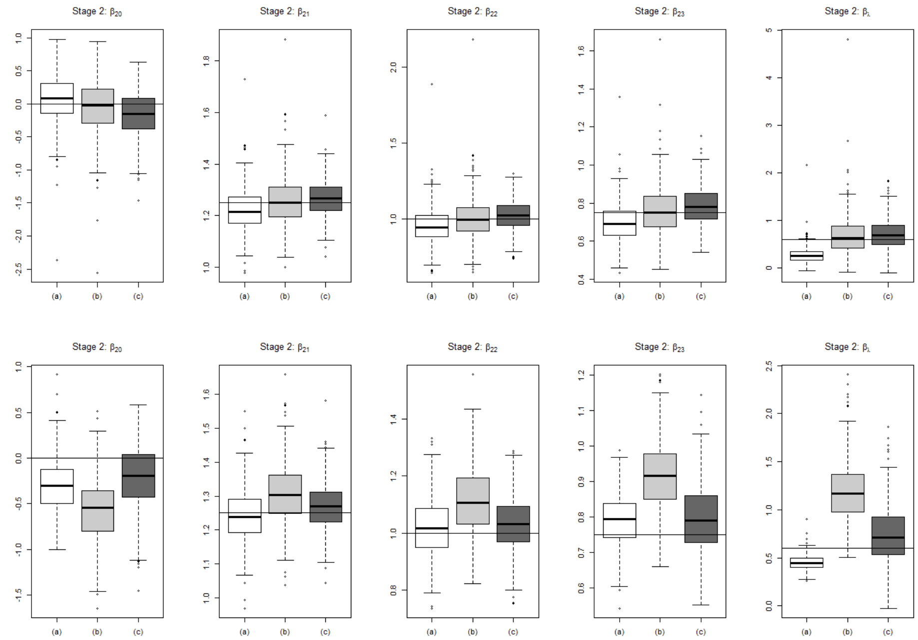

Figure 6: Boxplots of the parameters of the second stage with the exclusion restriction. The values of ν are gained based on the profile likelihood. The top panel shows the parameters with contaminated x_1 and $y_1 = 0$. The bottom panel shows the parameters with contaminated x_1 and $y_1 = 1$. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

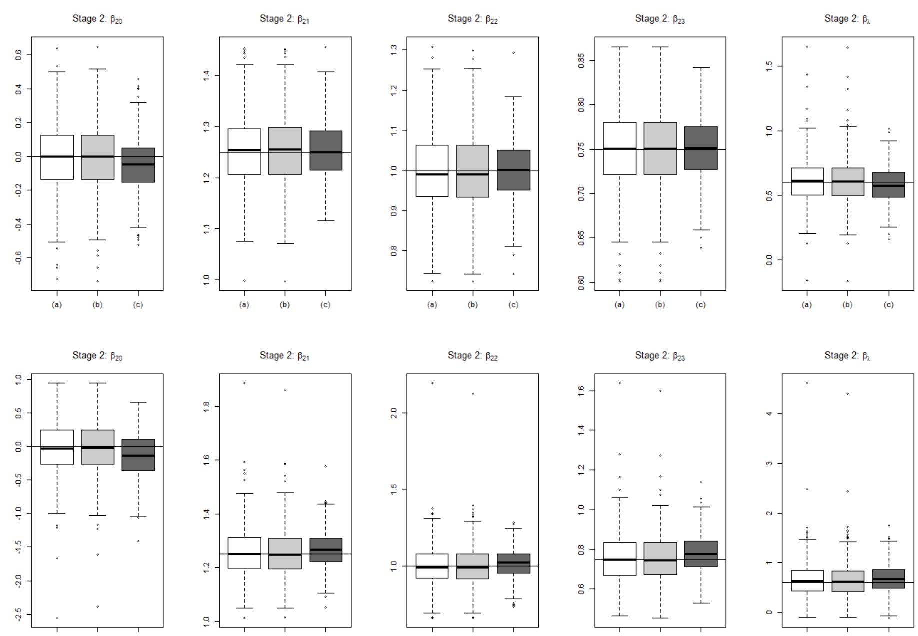

Figure 7: Boxplots of the parameters of the second stage with no contamination. The top panel shows the parameters with no exclusion restriction. The bottom panel shows the parameters with the exclusion restriction. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

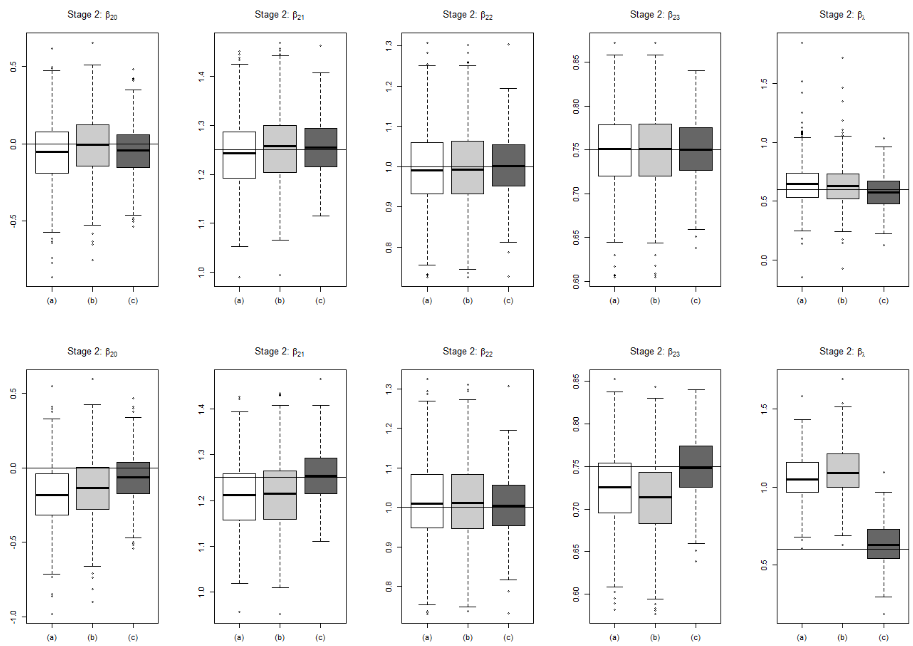

Figure 8: Boxplots of the parameters of the second stage without the exclusion restriction. The top panel shows the parameters with contaminated x_1 and $y_1 = 0$. The bottom panel shows the parameters with contaminated x_1 and $y_1 = 1$. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

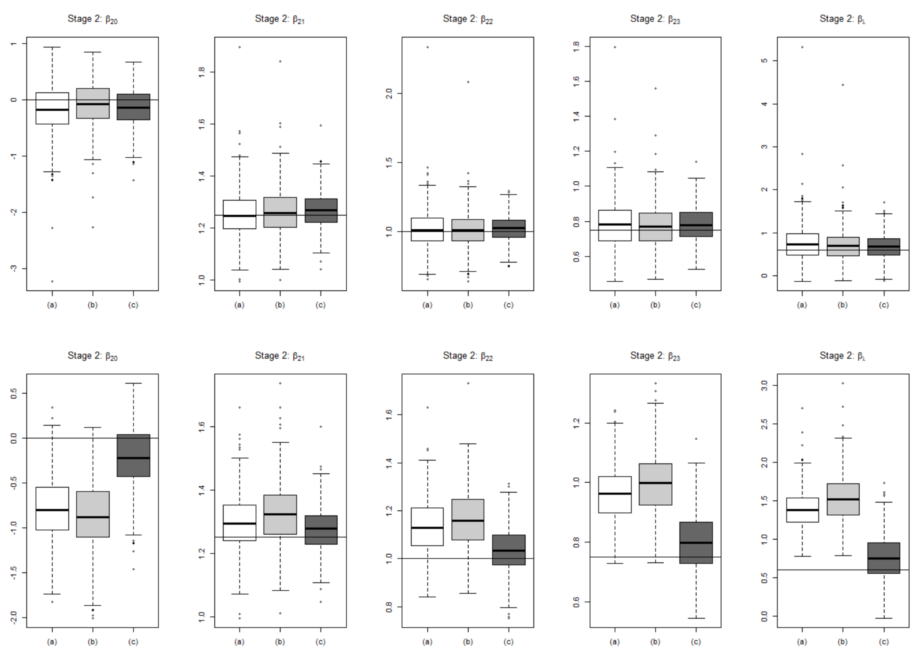

Figure 9: Boxplots of the parameters of the second stage with the exclusion restriction. The top panel shows the parameters with contaminated x_1 and $y_1 = 0$. The bottom panel shows the parameters with contaminated x_1 and $y_1 = 1$. (a) is obtained using the standard two-stage estimator method. For (b) the robust first stage is used. For (c) the robust first and second stages are used.

Sensitivity Analysis results

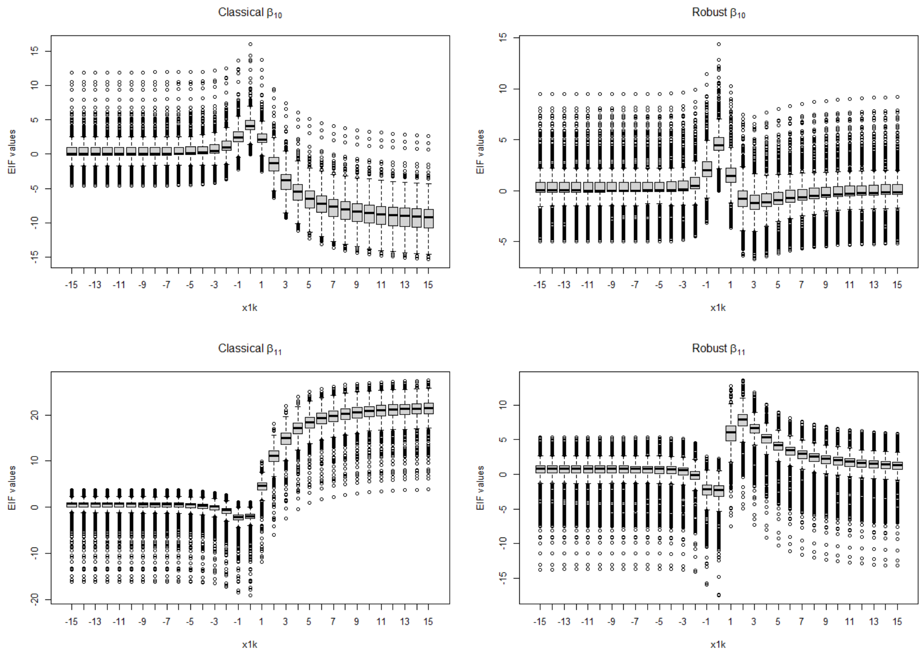

Figure 10: Boxplots of the EIF values for the first stage parameters, where observation x_{1k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

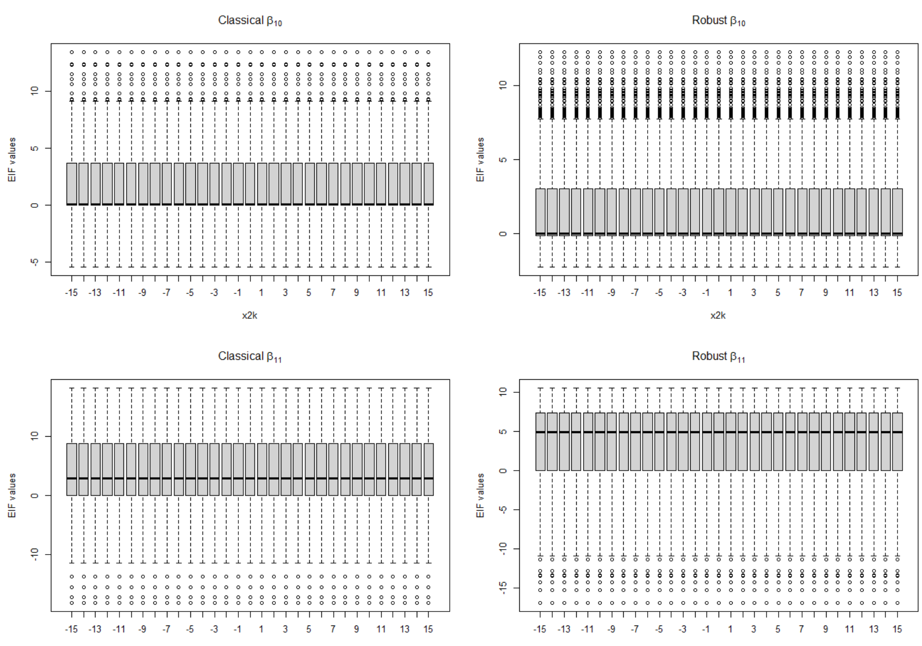

Figure 11: Boxplots of the EIF values for the first stage parameters, where observation x_{2k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

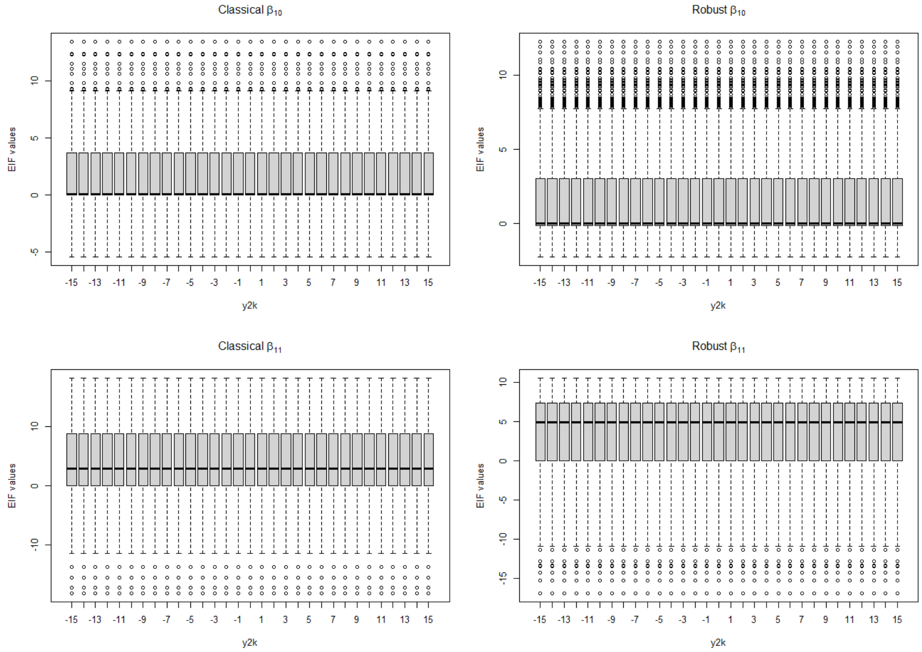

Figure 12: Boxplots of the EIF values for the first stage parameters, where observation y_{2k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

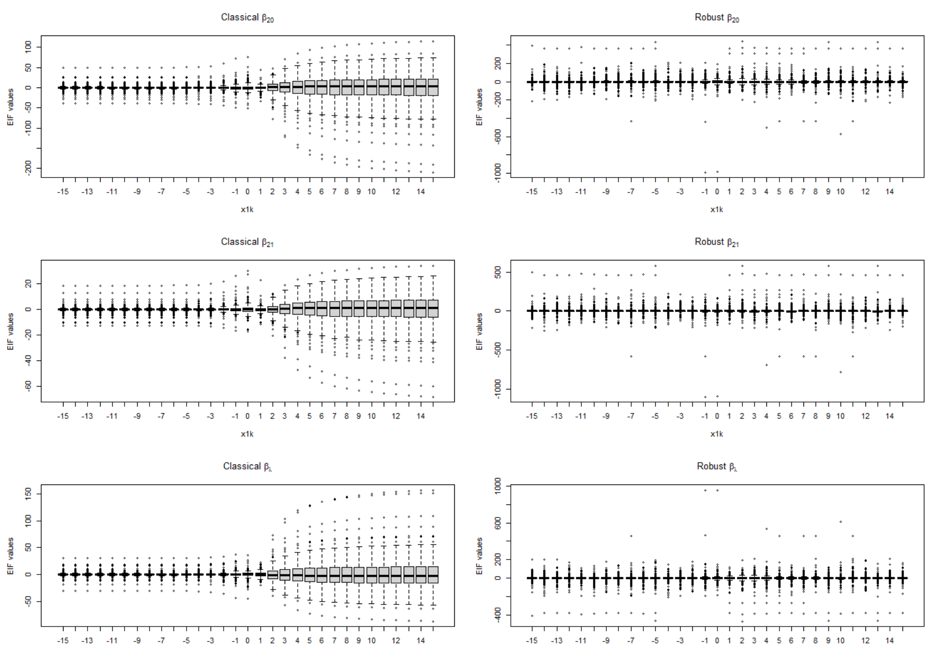

Figure 13: Boxplots of the EIF values for the second stage parameters, where observation x_{1k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

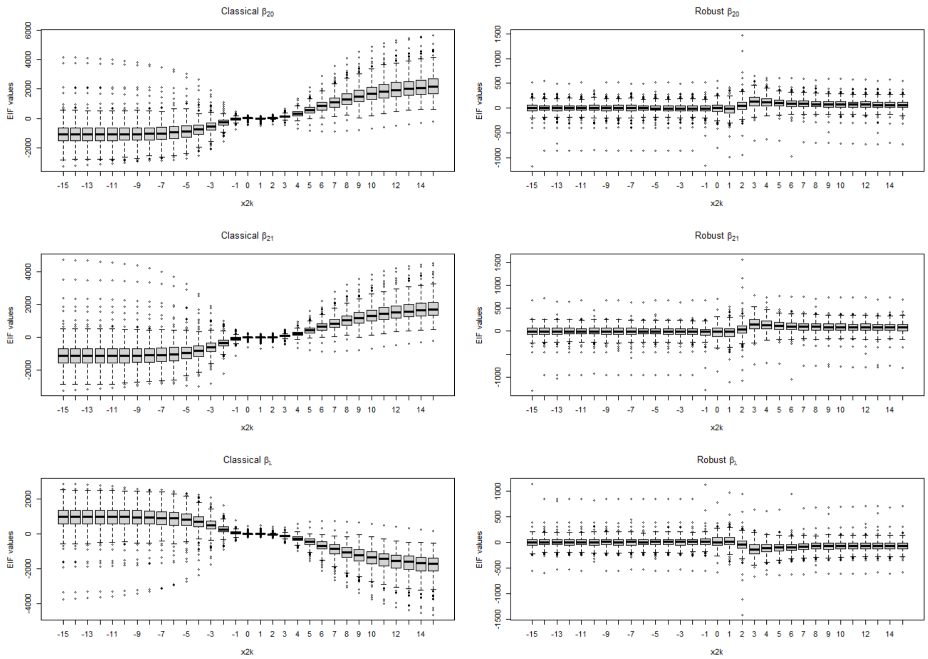

Figure 14: Boxplots of the EIF values for the second stage parameters, where observation x_{2k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

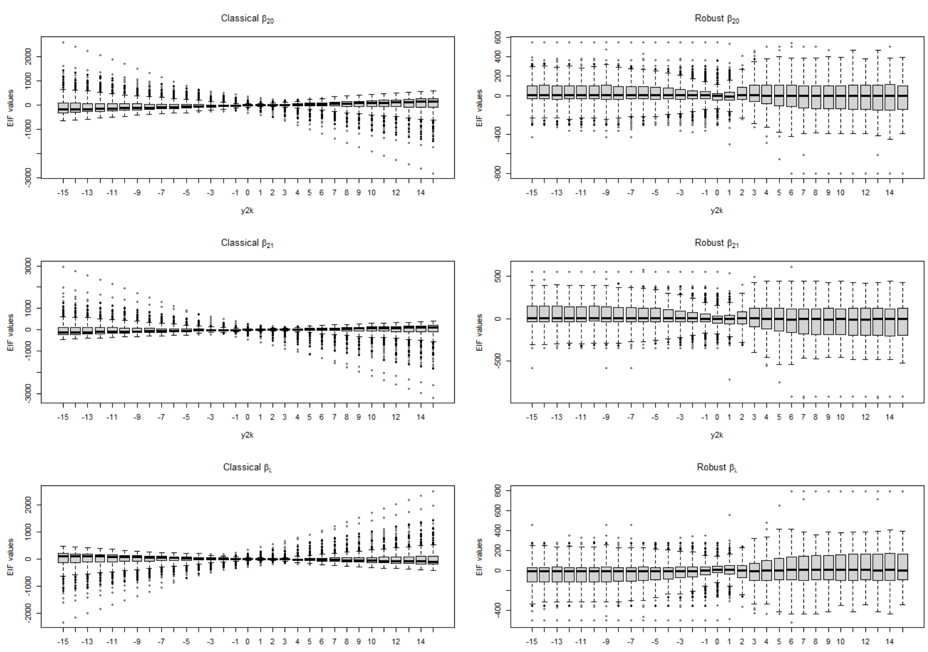

Figure 15: Boxplots of the EIF values for the second stage parameters, where observation y_{2k} is replaced. Here the exclusion restriction is included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

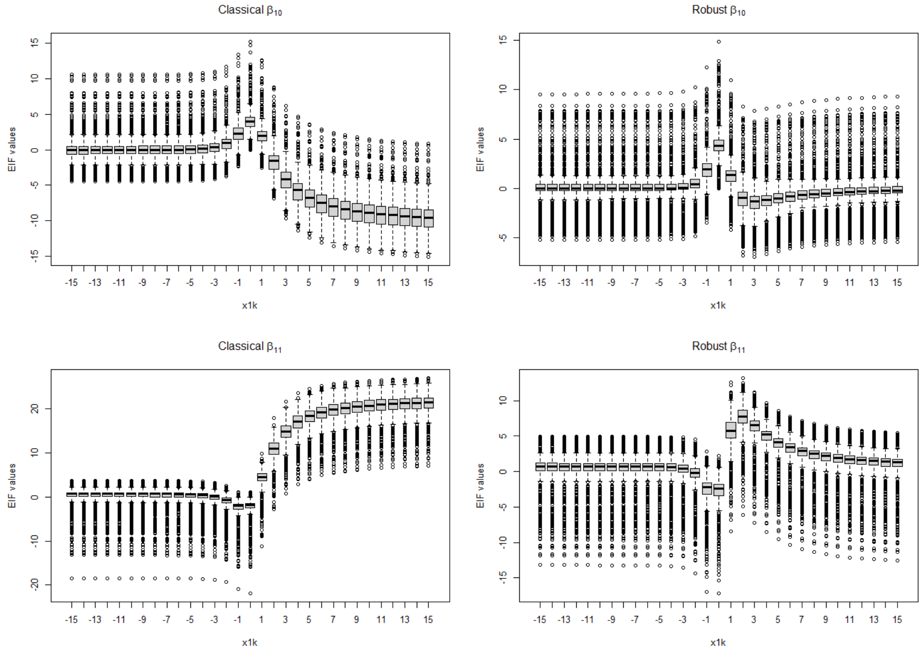

Figure 16: Boxplots of the EIF values for the first stage parameters, where observation x_{1k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

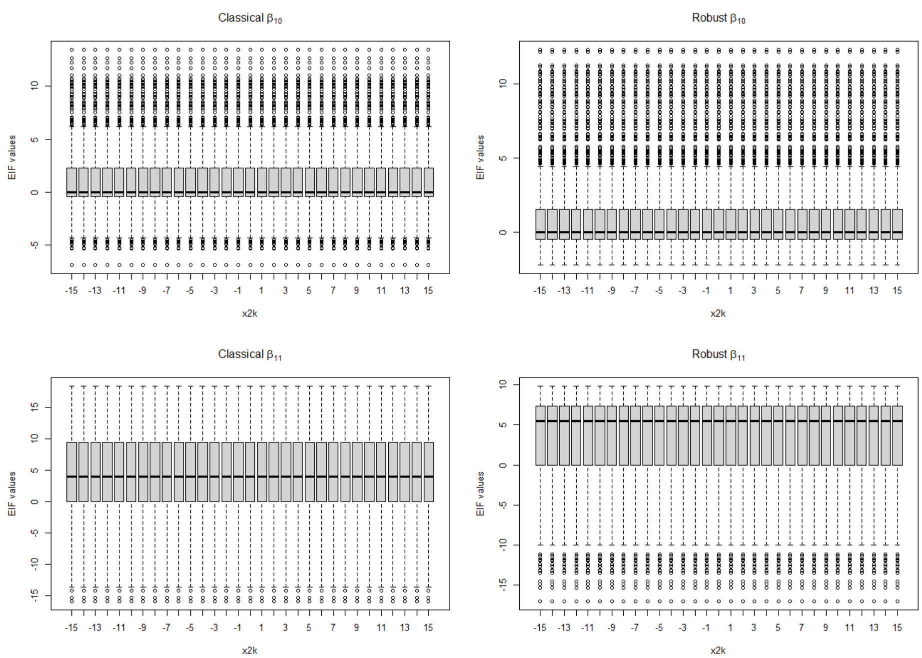

Figure 17: Boxplots of the EIF values for the first stage parameters, where observation x_{2k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

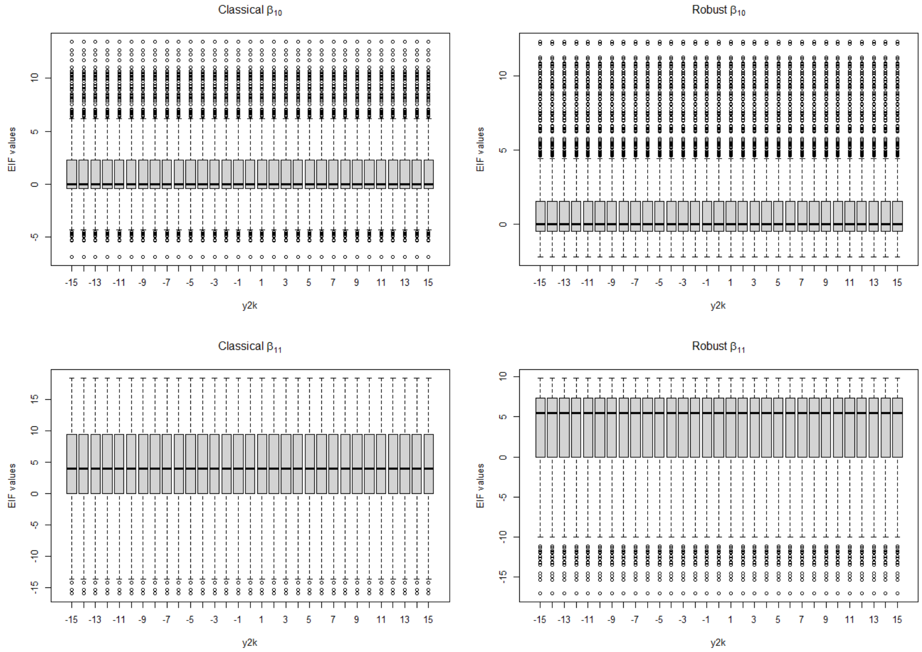

Figure 18: Boxplots of the EIF values for the first stage parameters, where observation y_{2k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

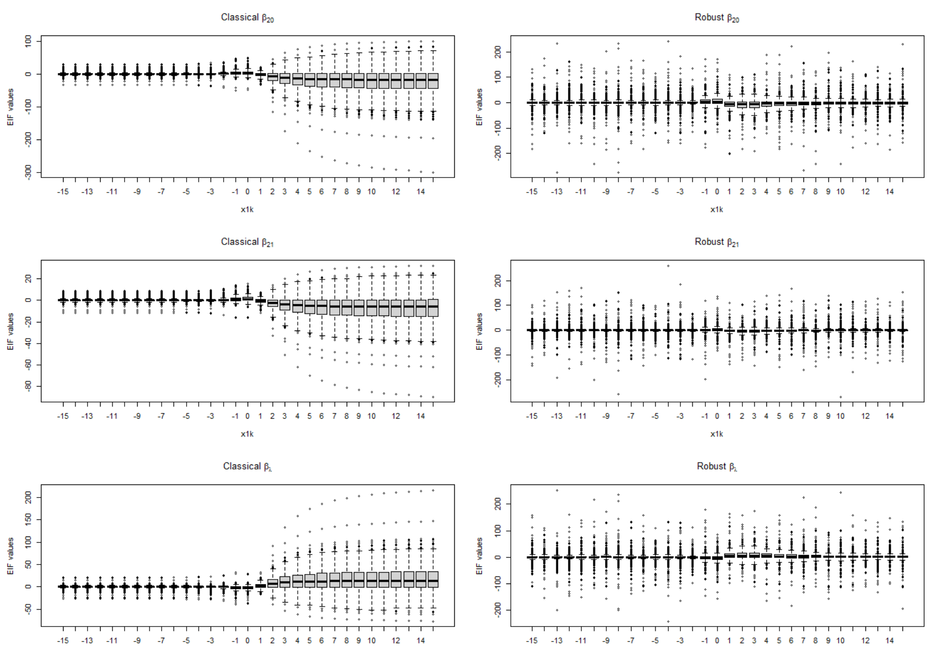

Figure 19: Boxplots of the EIF values for the second stage parameters, where observation x_{1k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

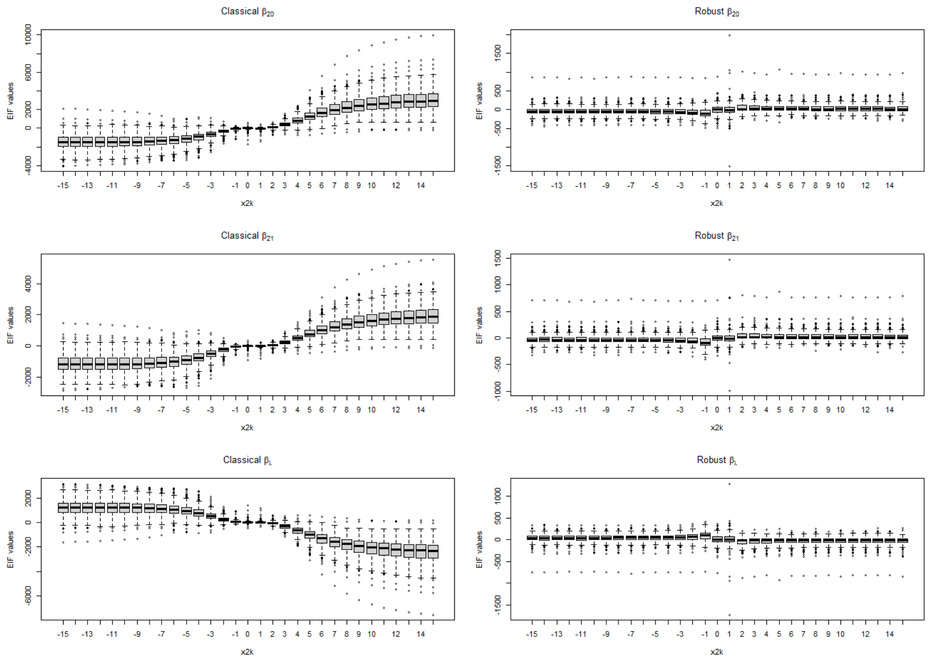

Figure 20: Boxplots of the EIF values for the second stage parameters, where observation x_{2k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

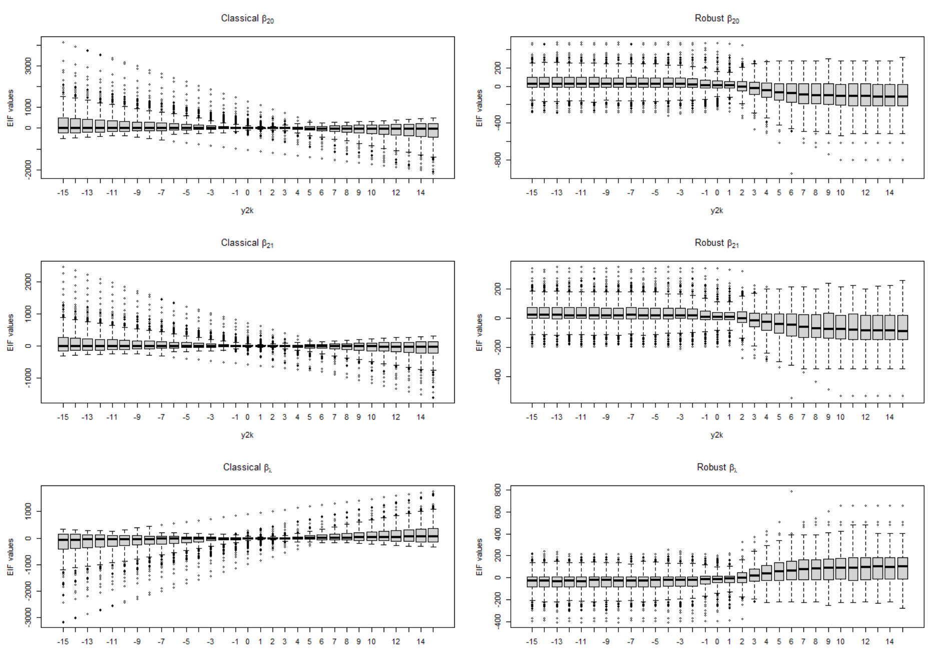

Figure 21: Boxplots of the EIF values for the second stage parameters, where observation y_{2k} is replaced. Here the exclusion restriction is not included. The figures on the left represents the EIF of the Classical Two-stage model. The figures on the right represent the EIF of the Robust Two-stage models.

Ambulatory Expenditure Results Under the Normal Distribution

Table 12: Estimation result of the ambulatory expenditures data for the classical and robust Two-stage estimators. The errors have been estimated under the assumption of a bivariate Normal distribution. The standard errors are given in the parenthesis. The significance codes are ”***” 0.001, ”**” 0.01, ”*” 0.05, ”·”0.10.

	With ex. restriction		Without ex. restriction	
	Classical	Robust	Classical	Robust
Selection				
Intercept	-0.668 (0.194)***	-0.700 (0.196)***	-0.718 (0.192)***	-0.749 (0.195)***
age	0.086 (0.027)**	0.094 (0.028)***	0.097 (0.027)***	0.105 (0.027)***
female	0.663 (0.061)***	0.704 (0.063)***	0.644 (0.060)***	0.687 (0.062)***
educ	0.062 (0.012)***	0.062 (0.012)***	0.070 (0.011)***	0.070 (0.011)***
blhisp	-0.366 (0.062)***	-0.389 (0.063)***	-0.375 (0.062)***	-0.398 (0.063)***
totchr	0.796 (0.071)***	0.834 (0.080)***	0.793 (0.071)***	0.793 (0.080)***
ins	0.169 (0.063)**	0.173 (0.064)**	0.181 (0.063)**	0.183 (0.064)**
income	0.003 (0.001)*	0.003 (0.001)·	-	-
Outcome				
Intercept	5.289 (0.289)***	5.409 (0.273)***	5.303 (0.294)***	5.402 (0.277)***
age	0.202 (0.024)***	0.200 (0.024)***	0.202 (0.024)***	0.201 (0.025)***
female	0.292 (0.073)***	0.252 (0.069)***	0.289 (0.074)***	0.255 (0.069)***
educ	0.0124 (0.012)	0.013 (0.012)	0.012 (0.012)	0.013 (0.012)
blhisp	-0.183 (0.065)**	-0.153 (0.065)*	-0.181 (0.066)**	-0.155 (0.065)*
totchr	0.501 (0.049)***	0.479 (0.038)***	0.498 (0.049)***	0.481 (0.038)***
ins	-0.047 (0.053)	-0.068 (0.052)	-0.047 (0.053)	-0.067 (0.053)
IMR	-0.464 (0.283)	-0.689 (0.255)**	-0.480 (0.291)·	-0.677 (0.259)**