

Birnbaum-Saunders sample selection model

Fernando de Souza Bastos^{*†}  and Wagner Barreto-Souza^{#†} 

^{*}Universidade Federal de Viçosa - Campus UFV - Florestal, Instituto de Ciências Exatas e Tecnológicas, Florestal, Brazil; [#]Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [†]Universidade Federal de Minas Gerais, Departamento de Estatística, Belo Horizonte, Brazil

ARTICLE HISTORY

Compiled July 8, 2020

ABSTRACT

The sample selection bias problem occurs when the outcome of interest is only observed according to some selection rule, where there is a dependence structure between the outcome and the selection rule. In a pioneering work, J. Heckman proposed a sample selection model based on a bivariate normal distribution for dealing with this problem. Due to the non-robustness of the normal distribution, many alternatives have been introduced in the literature by assuming extensions of the normal distribution like the Student-t and skew-normal models. One common limitation of the existent sample selection models is that they require a transformation of the outcome of interest, which is common \mathbb{R}^+ -valued, such as income and wage. With this, data are analyzed on a non-original scale which complicates the interpretation of the parameters. In this paper, we propose a sample selection model based on the bivariate Birnbaum-Saunders distribution, which has the same number of parameters that the classical Heckman model. Further, our associated outcome equation is \mathbb{R}^+ -valued. In other words, our model is parsimonious and does not require the transformation of positive outcomes. We discuss estimation by maximum likelihood and present some Monte Carlo simulation studies. An empirical application to the ambulatory expenditures data from the 2001 Medical Expenditure Panel Survey is presented.

KEYWORDS

Bivariate Birnbaum-Saunders distribution; Heckman sample selection model; Skewness; Maximum likelihood estimation; Nonignorable missing mechanism.

1. Introduction

In many practical situations, it is common to find data missing not at random (MNAR; [34]), which means that the observed data do not represent a random sample from the population. Non-random sampling here means that the outcome of interest is only observed when some selection rule holds, where there is a dependence structure between the outcome of interest and the selection rule. This problem is also known in the literature as *sample selection bias* and *nonignorable missing data*.

Please cite as: Bastos, F.S. and Barreto-Souza, W. (2020). Birnbaum-Saunders sample selection model. *Journal of Applied Statistics*. doi: 10.1080/02664763.2020.1780570

Email: fsbmat@gmail.com; fernando.bastos@ufv.br

Email: wagner.barretosouza@kaust.edu.sa

In a pioneering work, [14] proposed a model for dealing with the sample selection bias problem based on a bivariate normal distribution to construct a relationship between the outcome and selection rule. In that paper, estimation by maximum likelihood is proposed. This model is known in the literature as Heckman sample selection and has a nonignorable missing mechanism [24]. Due to the non-robustness of the normal distribution in the presence of outliers, [15] proposed a semiparametric alternative known as Heckman’s two-step method.

Several papers arose to investigate the performance of the parametric and semiparametric Heckman models. For instance, see [27], [30], [22], [35], [13], [25, 26] and [20]. These authors showed that the Heckman models can reduce or eliminate the selection bias when its assumptions are met, that is when the model is correctly specified. However, they also showed that the departure from normality, which can be caused by the presence of outliers, multimodality, mixtures, or skewness, is a problem that the Heckman model can not handle.

In this direction, [23] introduced a sample selection model based on the bivariate Student-t distribution to associate the outcome with the selection rule. The advantage of this model is the robustness concerning the normal assumption. Disadvantages of this model are the difficulty to estimate the degrees of freedom (additional parameter to the Heckman model) and that the correlation equal to zero does not imply independence between the outcome and the selection rule (unless when the degrees of freedom goes to ∞ , which corresponds to Heckman model). A Bayesian approach of this Student-t sample selection model was proposed by [10]. An approach based on copulas was proposed by [19] and semiparametric alternatives were suggested by [1], [31] and [28]. A non-parametric approach was studied by [9].

An important advantage of the parametric models over the non-parametric and semiparametric models is that they can identify the intercept of the outcome equation (mean regression structure of the outcome of interest), mainly when a prediction is of great interest, as argued by [29]. These authors introduced a skew-normal sample selection model by adding a skewness parameter to the Heckman model (through the bivariate skew-normal distribution) for treating skewed data with sample selection bias. One motivation for this model is that the widely-used logarithmic transformation could not be sufficient to eliminate the skewness of the outcome. Another recent important paper is due to [37], where a robust inference procedure was proposed. This robust method combines the Heckman’s two-step method and the M-estimators [12, 16].

One common limitation of the existent models aforementioned is that they require a transformation of the outcome of interest, which is common \mathbb{R}^+ -valued; for example, income and wage. The logarithmic transformation is a common choice. In practice, it is not clear which transformation is adequate. Moreover, data are analyzed on a non-original scale to the applied researches which complicate the interpretation of the parameters (for instance, see [29]).

Another disadvantage of the existent models is the lack of parsimony. Many of the parametric alternatives to the classical Heckman model have additional parameters. As mentioned by [37], to add parameters to take into account a few possible outliers is only justifiable when observations contain important information, which is hard to check in practice.

With these problems in mind, in this paper we propose a sample selection model based on the bivariate Birnbaum-Saunders (BS) distribution [18], having the same number of parameters that the classical Heckman model and with associated outcome equation being \mathbb{R}^+ -valued. In other words, our proposed model is parsimonious and

does not require the transformation of positive outcomes, so overcoming some drawbacks of existent sample selection models. Below, we list some additional features of the Birnbaum-Saunders sample selection model:

(i) *Robustness*. This is illustrated through simulated and real data sets. In the empirical illustration, we analyze the ambulatory expenditures from the 2001 Medical Expenditure Panel Survey (MEPS2001) and show that our model indicates that there is the sample selection bias in agreement with robust models considered in the literature and in contrast with the classical Heckman model.

(ii) *Independence*. For our proposed model, sample selection bias equal to 0 implies independence between the outcome and selection rule as in the Heckman model; for instance, this does not happen in the Student-t sample selection model by [23].

(iii) *Skewness*. The BS model naturally accommodates the skewness of positive outcomes.

(iv) *Interpretation*. Since our model does not require the transformation of positive outcomes such as income and wage, the interpretation of the parameters is made on the original scale familiar to the applied researches.

Inspired by [29], whose recommendation is the use of a range of suitable parametric models for analyzing data with sample selection bias, we expect that our model may be included in this range for practitioners, mainly due to the advantages aforementioned, which overcome existent unsolved issues in the literature.

The paper is organized in the following way. In Section 2 we describe important tools for our purposes in this paper, which are the classical Heckman model and the bivariate Birnbaum-Saunders distribution by [18]. In Section 3 we introduce the Birnbaum-Saunders (BS) sample selection model and propose an estimation of the parameters by maximum likelihood. Further, we discuss the goodness-of-fit of the BS model through residual analysis. Section 4 is devoted to Monte Carlo simulations in order to study the finite-sample behavior of the maximum likelihood estimators and the size and power of the likelihood ratio test for checking the presence of sample selection bias. We investigate the performance of the BS model under misspecification in Section 5. In Section 6, we present an empirical application to the ambulatory expenditures from the 2001 Medical Expenditure Panel Survey (MEPS). This paper contains Supplementary Material.

2. Preliminaries

In this section we introduce some notations and describe the Heckman selection model and the bivariate Birnbaum-Saunders distribution, which plays an important rule in this paper.

2.1. Heckman sample selection model

Denote by $Y_{1i}^* \in \mathbb{R}$, for $i = 1, \dots, n$, the outcomes of interest (response variable) and assume that they are linearly related to the covariate vectors $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^p$ through the following regression:

$$Y_{1i}^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_{1i}, \quad i = 1, \dots, n, \quad (1)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients and $\{\epsilon_{1i}\}_{i=1}^n$ are error terms. Assume that there is a latent selection mechanism given by the regression

$$Y_{2i}^* = \mathbf{z}_i^\top \boldsymbol{\gamma} + \epsilon_{2i}, \quad i = 1, \dots, n, \quad (2)$$

where $\{\mathbf{z}_i\}_{i=1}^n \in \mathbb{R}^q$ are covariate vectors, $\boldsymbol{\gamma} \in \mathbb{R}^q$ is a vector of unknown parameters and $\{\epsilon_{2i}\}_{i=1}^n$ are also error terms. It is also assumed that the error terms follow a bivariate normal distribution, that is

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \stackrel{iid}{\sim} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right), \quad i = 1, \dots, n, \quad (3)$$

where $\rho \in (-1, 1)$ is the correlation parameter and $\sigma^2 > 0$. Assume that we only observe the outcome of interest Y_{1i}^* when $Y_{2i}^* > 0$, for $i = 1, \dots, n$. That is, in practice, we observe the variables $\{(Y_i, U_i)\}_{i=1}^n$, which are defined by

$$\begin{cases} U_i = I\{Y_{2i}^* > 0\} \\ Y_i = U_i Y_{1i}^*, \end{cases} \quad \text{for } i = 1, \dots, n, \quad (4)$$

where $I\{A\}$ is an indicator function assuming the value 1 if the event A holds; otherwise it equals 0. Equations (1) and (2) are called primary interest and selection equations, respectively. The Heckman sample selection model is defined by (4) and the normal assumption given in (3). The parameter ρ controls the sample selection bias. If $\rho = 0$, there is not sample selection bias, otherwise there is a nonignorable missing mechanism or sample selection bias.

The density function associated to (4) has a continuous component, which is given by the conditional density of Y_i given $U_i = 1$, denoted by $f(y_i|U_i = 1)$, as follows

$$f(y_i|U_i = 1) = \frac{1}{\sigma} g \left(\frac{y - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \frac{\Phi \left(\frac{\rho}{\sqrt{1 - \rho^2}} \left(\frac{y - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) + \frac{\mathbf{z}_i^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}} \right)}{\Phi(\mathbf{z}_i^\top \boldsymbol{\gamma})},$$

and by a discrete component defined by the probit model as $P(U_i = u_i) = \Phi(\mathbf{z}_i^\top \boldsymbol{\gamma})^{u_i} \Phi(-\mathbf{z}_i^\top \boldsymbol{\gamma})^{1-u_i}$, for $u_i = 0, 1$ and $i = 1, \dots, n$, where $g(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution functions of the standard normal distribution, respectively. The log-likelihood function is given by $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta})$, where

$$\begin{aligned} \mathcal{L}_i(\boldsymbol{\theta}) &= u_i \log f(y_i|U_i = 1) + u_i \log \Phi(\mathbf{z}_i^\top \boldsymbol{\gamma}) + (1 - u_i) \log \Phi(-\mathbf{z}_i^\top \boldsymbol{\gamma}) \\ &= u_i \left\{ \log \Phi \left(\frac{\mathbf{z}_i^\top \boldsymbol{\gamma} + \rho(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/\sigma}{\sqrt{1 - \rho^2}} \right) + \log g \left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) - \log \sigma \right\} \\ &\quad + (1 - u_i) \log \Phi(-\mathbf{z}_i^\top \boldsymbol{\gamma}), \end{aligned}$$

for $i = 1, \dots, n$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \sigma^2, \rho)^\top$ is the parameter vector.

When the assumption of bivariate normality is satisfied, the maximum likelihood estimators (MLEs) obtained from the maximization of the above log-likelihood function are consistent, asymptotically normally distributed, and efficient. However, the function $\mathcal{L}(\boldsymbol{\theta})$ is nonlinear and thus requires iterative methods for its maximization,

which require a good starting point for the numerical algorithm, since $\mathcal{L}(\boldsymbol{\theta})$ may have local maxima. Therefore, the second estimation method proposed by [15], known as a two-step method, is suggested as a good starting point in the optimization procedure for finding the maximum likelihood estimates, which does not require complicated algorithms to fit [21].

It is important to emphasize that both methods suffer from possible multicollinearity problems when the selection equation has the same covariates that the regression equation. In this case, to reduce the impact of this problem in the parameter estimation, the exclusion restriction is suggested, where at least one significant covariate for the selection equation is not included in the primary regression.

2.2. Bivariate Birnbaum-Saunders distribution

We say that a random vector $\mathbf{Y} = (Y_1, Y_2)^\top$ follows a bivariate Birnbaum-Saunders (BS) distribution with parameters $\alpha_1, \beta_1, \alpha_2, \beta_2 > 0$ and $-1 < \rho < 1$, denoted by $\mathbf{Y} \sim BS_2(\alpha_1, \alpha_2, \beta_1, \beta_2, \rho)$ or $\mathbf{Y} \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho)$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, if its joint cumulative distribution function is given by

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = \Phi_2 \left(\frac{1}{\alpha_1} \left(\sqrt{\frac{y_1}{\beta_1}} - \sqrt{\frac{\beta_1}{y_1}} \right), \frac{1}{\alpha_2} \left(\sqrt{\frac{y_2}{\beta_2}} - \sqrt{\frac{\beta_2}{y_2}} \right); \rho \right),$$

for $y_1, y_2 > 0$, where $\Phi_2(\cdot, \cdot; \cdot)$ is the joint cumulative distribution function of a standard bivariate normal distribution with joint density function given by

$$g_2(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (z_1^2 + z_2^2 - 2\rho z_1 z_2) \right\},$$

with $(z_1, z_2) \in \mathbb{R}^2$.

This bivariate Birnbaum-Saunders distribution was introduced and studied by [18]. This bivariate distribution can be obtained through a transformation of a random vector following a bivariate normal distribution. For further details, please refer to [18].

The mean and variance of Y_i are respectively given by $E(Y_i) = \frac{\beta_i}{2} (2 + \alpha_i^2)$ and $Var(Y_i) = \frac{\beta_i^2 \alpha_i^2}{4} (4 + 5\alpha_i^2)$, for $i = 1, 2$. We here assume the following reparametrization: $\alpha_i = \sqrt{\frac{2}{\phi_i}}$ and $\beta_i = \frac{\phi_i \mu_i}{(\phi_i + 1)}$, for $i = 1, 2$. With this, we obtain

$$E(Y_i) = \mu_i \quad \text{and} \quad Var(Y_i) = \mu_i^2 \frac{(2\phi_i + 5)}{(\phi_i + 1)^2}, \quad i = 1, 2.$$

The parameter ρ controls the dependence between Y_1 and Y_2 . For $\rho = 0$, we have that Y_1 and Y_2 are independent. This follows directly from the relationship between the bivariate Birnbaum-Saunders distribution and the bivariate normal distribution discussed above. An explicit expression for the correlation between Y_1 and Y_2 can be found in the paper by [18]. Under this parameterization, we denote $\mathbf{Y} \sim BS_2(\boldsymbol{\mu}, \boldsymbol{\phi}; \rho)$,

with $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ and $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$. In this case, the density of \mathbf{Y} can be written as

$$f_{Y_1, Y_2}(y_1, y_2) = g_2(a_{y_1}(\mu_1, \phi_1), a_{y_2}(\mu_2, \phi_2); \rho) \\ \times \frac{\phi_1 + 1}{2\phi_1\mu_1} \sqrt{\frac{\phi_1}{2}} \left\{ \left[\frac{\phi_1\mu_1}{(\phi_1 + 1)y_1} \right]^{\frac{1}{2}} + \left[\frac{\phi_1\mu_1}{(\phi_1 + 1)y_1} \right]^{\frac{3}{2}} \right\} \\ \times \frac{\phi_2 + 1}{2\phi_2\mu_2} \sqrt{\frac{\phi_2}{2}} \left\{ \left[\frac{\phi_2\mu_2}{(\phi_2 + 1)y_2} \right]^{\frac{1}{2}} + \left[\frac{\phi_2\mu_2}{(\phi_2 + 1)y_2} \right]^{\frac{3}{2}} \right\},$$

for $y_1, y_2 > 0$, where

$$a_{y_i}(\mu_i, \phi_i) = \sqrt{\frac{\phi_i}{2}} \left(\sqrt{\frac{(\phi_i + 1)y_i}{\phi_i\mu_i}} - \sqrt{\frac{\phi_i\mu_i}{(\phi_i + 1)y_i}} \right), \quad \text{for } i = 1, 2.$$

We now have all ingredients to introduce the Birnbaum-Saunders sample selection model, which is one of the aims of the next section.

3. Model definition and inference

Consider a sample of bivariate vectors $(Y_{11}^*, Y_{21}^*)^\top, \dots, (Y_{1n}^*, Y_{2n}^*)^\top$ following $BS_2(\boldsymbol{\mu}_i, \boldsymbol{\phi}, \rho)$ distribution, where $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})^\top \in \mathbb{R}^+ \times \mathbb{R}^+$, $i = 1, 2, \dots, n$, $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top \equiv (\phi, 1)$, with $\phi > 0$, and $\rho \in (-1, 1)$. Here, the sequences $\{Y_{1i}^*\}_{i=1}^n$ and $\{Y_{2i}^*\}_{i=1}^n$ represent the outcomes of interest and the selection variables, respectively. Assume that we only observe $(Y_1, U_1), \dots, (Y_n, U_n)$, which are defined by

$$\begin{cases} U_i = I\{Y_{2i}^* > 1\} \\ Y_i = U_i Y_{1i}^*, \end{cases} \quad \text{for } i = 1, \dots, n, \quad (5)$$

That is, we observe the outcome of interest Y_{1i}^* only when $Y_{2i}^* > 1$, for $i = 1, \dots, n$.

Remark 1. Without loss of generality we can define $U_i = I\{Y_{2i}^* > 1\}$ instead of $U_i = I\{Y_{2i}^* > c\}$, for $c > 0$ arbitrary, since this constant c will be absorbed by the parameters of the Birnbaum-Saunders distribution, which is invariant by scale. Therefore, we work with $c = 1$ to avoid non-identifiability problems, similarly as done in the Heckman and associated sample selection models. We have also fixed $\phi_2 = 1$ above to avoid non-identifiability problems, also as done in the Heckman model. Note we cannot define $U_i = I\{Y_{2i}^* > 0\}$ for our BS model like in the Heckman model since $P(Y_{2i}^* > 0) = 1$.

We assume the following regression structures for the mean vector of our model:

$$\log \mu_{1i} = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{and} \quad \log \mu_{2i} = \mathbf{w}_i^\top \boldsymbol{\gamma}, \quad (6)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top \in \mathbb{R}^q$ are vectors of unknown parameters with $p + q < n$, and $\mathbf{x}_i = (\mathbf{x}_{1i}, \dots, \mathbf{x}_{pi})^\top$ and $\mathbf{w}_i = (\mathbf{w}_{1i}, \dots, \mathbf{w}_{qi})^\top$ being explanatory variables not necessarily mutually exclusive, for $i = 1, \dots, n$.

The Birnbaum-Saunders sample selection model is defined by (5) and the regression structures in (6). Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \phi, \rho)^\top$ be the parameter vector.

We now discuss estimation of the parameters by maximum likelihood. For this, we need to find the log-likelihood function. Consider a single pair (Y, U) of the BS sample selection model as above, with (y, u) denoting the observed value of (Y, U) . We have that

$$\begin{aligned} P(Y \leq y) &= P(Y \leq y|U = 1)P(U = 1) + P(Y \leq y|U = 0)P(U = 0) \\ &= I\{y > 0\}F_{Y_1^*|Y_2^* > 1}(y) (1 - F_{Y_2^*}(1)) + I\{y = 0\}F_{Y_2^*}(1), \end{aligned}$$

where $(Y_1^*, Y_2^*) \sim \text{BS}_2(\boldsymbol{\mu}, \boldsymbol{\phi}; \rho)$, $F_{Y_1^*|Y_2^* > 1}(y)$ is the conditional cumulative distribution of Y_1^* given $Y_2^* > 1$ and

$$F_{Y_2^*}(1) = P(Y_2^* \leq 1) = \Phi \left(\sqrt{\frac{1}{2}} \left(\sqrt{\frac{2}{\mu_2}} - \sqrt{\frac{\mu_2}{2}} \right) \right) = \Phi \left(\mu_2^{-1/2} - \mu_2^{1/2}/2 \right),$$

with $\Phi(\cdot)$ denoting the cumulative distribution function of the univariate standard normal distribution. Then, we obtain that the density function of Y is

$$f_Y(y) = \begin{cases} f_{Y_1^*|Y_2^* > 1}(y) \Phi \left(\mu_2^{1/2}/2 - \mu_2^{-1/2} \right), & \text{if } y > 0, \\ \Phi \left(\mu_2^{-1/2} - \mu_2^{1/2}/2 \right), & \text{if } y = 0, \end{cases}$$

where $f_{Y_1^*|Y_2^* > 1}(\cdot)$ is the conditional density function of Y_1^* given $Y_2^* > 1$. That is, Y is a random variable with discrete and continuous components.

The continuous component can be obtained from the paper by [3], in which authors developed a general formula for the distributions that arise from the selection mechanism $\mathbf{Y}|\mathbf{U} \in C$, where $\mathbf{Y} \in \mathbb{R}^k$, $\mathbf{U} \in \mathbb{R}^l$ and C is a measurable subset of \mathbb{R}^l , such that $0 < P(\mathbf{U} \in C) < 1$. In our case, we obtain that

$$f_{Y_1^*|Y_2^* > 1}(y) = f_{Y_1^*}(y) \frac{P(Y_2^* > 1|Y_1^* = y)}{P(Y_2^* > 1)}, \quad (7)$$

where $f_{Y_1^*}(\cdot)$ is the density function of $Y_1^* \sim \text{BS}(\mu_1, \phi)$ and an explicit expression for the term $P(Y_2^* > 1|Y_1^* = y)$ is given in [18]; the general result (7) it can be also found in [4], [2] and [5]. With this, the density function of Y can be written as

$$\begin{aligned} f_Y(y; \boldsymbol{\theta}) &= I\{y > 0\} \frac{1}{2\sqrt{2\pi}} \exp \left(-\frac{\phi}{4} \left(\sqrt{\frac{y(\phi+1)}{\phi\mu_1}} - \sqrt{\frac{\phi\mu_1}{y(\phi+1)}} \right)^2 \right) \\ &\times \Phi \left(\frac{\mu_2 - 2}{2\sqrt{\mu_2(1-\rho^2)}} + \frac{\sqrt{\phi}\rho}{\sqrt{2(1-\rho^2)}} \left(\sqrt{\frac{(\phi+1)y}{\phi\mu_1}} - \sqrt{\frac{\phi\mu_1}{y(\phi+1)}} \right) \right) \\ &\times \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\mu_1 y}} + \frac{\sqrt{\phi\mu_1}}{\sqrt{y^3(\phi+1)}} \right) \sqrt{\frac{\phi}{2}} + I\{y = 0\} \Phi \left(\frac{1}{\sqrt{\mu_2}} - \frac{\sqrt{\mu_2}}{2} \right). \quad (8) \end{aligned}$$

The model (8) is identifiable in the sense that for any pair of parameters $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ we have that $f_Y(y; \boldsymbol{\theta}_1) \neq f_Y(y; \boldsymbol{\theta}_2)$.

The conditional density given in (7), which corresponds to the continuous component of the mixed variable Y , have already appeared in the literature in the recent work

by [17]. They proposed a generalization of the multivariate Birnbaum-Saunders distribution based on the multivariate skew-normal distribution and present the marginal, reciprocal, and conditional distributions obtained from the proposed distribution.

The log-likelihood function for the BS sample selection model is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta}), \quad (9)$$

where $\mathcal{L}_i(\boldsymbol{\theta}) = \log f_{Y_i}(y_i; \boldsymbol{\theta})$, with $f_{Y_i}(\cdot; \boldsymbol{\theta})$ assuming the form (8), for $i = 1, \dots, n$. More explicitly for $i = 1, \dots, n$, we have that

$$\begin{aligned} \mathcal{L}_i(\boldsymbol{\theta}) = & u_i \left\{ \log \Phi \left(\frac{\mu_{2i} - 2}{2\sqrt{\mu_{2i}(1-\rho^2)}} + \frac{\sqrt{\phi}\rho}{\sqrt{2(1-\rho^2)}} \left(\sqrt{\frac{(\phi+1)y_i}{\phi\mu_{1i}}} - \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right) \right) \right. \\ & - \frac{\phi}{4} \left(\sqrt{\frac{y_i(\phi+1)}{\phi\mu_{1i}}} - \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right)^2 + \log \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\mu_{1i}y_i}} + \frac{\sqrt{\phi\mu_{1i}}}{\sqrt{y_i^3(\phi+1)}} \right) \\ & \left. + \log \left(\frac{1}{4} \sqrt{\frac{\phi}{\pi}} \right) \right\} + (1-u_i) \log \Phi \left(\sqrt{\frac{1}{2}} \left(\sqrt{\frac{2}{\mu_{2i}}} - \sqrt{\frac{\mu_{2i}}{2}} \right) \right), \quad (10) \end{aligned}$$

where μ_{1i} and μ_{2i} are the regression structures defined above. The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\beta}}^\top, \hat{\phi}, \hat{\rho})^\top$, is given by $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$, where $\mathcal{L}(\cdot)$ is the log-likelihood function given by (9)-(10). The score function $\mathbf{U}(\boldsymbol{\theta}) = \partial \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and the observed information matrix $\mathbf{I}(\boldsymbol{\theta}) = -\partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ are presented in the Supplementary Material. The standard errors of the estimates of the parameters can be obtained through $\mathbf{I}(\hat{\boldsymbol{\theta}})$.

Since there is no closed-form expression for $\hat{\boldsymbol{\theta}}$, we need to use some numerical maximization method to obtain it. In this paper we use the quasi-Newton algorithm Broyden-Fletcher-Goldfarb-Shanno (BFGS). For details, see [32]. The advantage of this method is that it eliminates the necessity to calculate the information matrix; the evaluation of this matrix in each iteration of the optimization process can be very cumbersome.

For testing hypotheses, we can perform likelihood ratio, score, or Wald tests. Our main interest here is to test the hypotheses $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$. That is, the null hypothesis states that there is no sample selection bias whereas the alternative hypothesis states that there is. We investigate the power of the likelihood ratio test for checking the presence of sample selection bias in the following section through Monte Carlo simulation.

We now conclude this section by discussing on residual analysis, which is frequently used to evaluate the validity of model assumptions. Residuals carry important information concerning the appropriateness of statistical models, and thereby play an important role in checking model adequacy. In the context of regression models, Pearson and studentized residuals are often used. Nevertheless, in a censored scenario, these residuals are not adequate even under normality; for instance, see [6]. In the Heckman-BS case, we consider the generalized Cox-Snell (GCS) residual given by

$$r_i^{GCS} = -\log \left(1 - F(y_i; \hat{\mu}_{1i}, \hat{\mu}_{2i}, \hat{\phi}, \hat{\rho}) \right)$$

and the Normalized Quantile residuals introduced by [11], which is defined by

$$r_i = \Phi^{-1} \left(F(y_i; \hat{\mu}_{1i}, \hat{\mu}_{2i}, \hat{\phi}, \hat{\rho}) \right),$$

for $i = 1, \dots, n$, where F is the cumulative distribution function of the response variable. Under a correct model specification, the GCS and normalized quantile residuals are approximately standard exponential and standard normal distributed, respectively. We illustrate the usefulness of these residuals in the empirical application of this paper.

4. Monte Carlo simulation

4.1. Finite-sample behavior of the estimators

In this section we evaluate the finite-sample behavior of the maximum likelihood estimators of the BS sample selection model through Monte Carlo simulations. A simulated study of the Heckman-BS model under misspecification (involving a gamma assumption) is presented in the next section. We consider two scenarios as follows.

- *Scenario 1.* We here consider exclusion restriction with the following regression structures:

$$\begin{aligned} \mu_{1i} &= \exp(1 + 0.7x_{2i} + 1.1x_{3i}) \quad \text{and} \\ \mu_{2i} &= \exp(1.6 + 0.8x_{2i} + 0.2x_{3i} + 0.7x_{4i}), \quad \text{for } i = 1, 2, \dots, n \end{aligned}$$

- *Scenario 2.* Here, exclusion restriction is not assumed, where μ_{1i} is given as in *Scenario 1* and $\mu_{2i} = \exp(1.6 + 0.8x_{2i} + 0.2x_{3i})$, $i = 1, 2, \dots, n$.

For each scenario, we consider three different configurations. In Configuration 1, we consider covariates $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, $j = 2, 3, 4$. For Configuration 2, we assume $x_{2i} \stackrel{iid}{\sim} \text{Bernoulli}(0.5)$, $x_{3i} \stackrel{iid}{\sim} \text{Uniform}(0, 1)$ and $x_{4i} \stackrel{iid}{\sim} \mathcal{N}(0, 3.5)$, $i = 1, \dots, n$. Configuration 3 considers that x_{2i} and x_{3i} follow a bivariate normal distribution (null mean) with correlation $\rho = 0.7$ and $x_{4i} \stackrel{iid}{\sim} \mathcal{N}(0, 3.5)$, $i = 1, \dots, n$. We set the sample sizes $n = 500, 1000, 2000$. The covariates were kept fixed along the simulation. The choices of parameter values for such explanatory variables were made in order to keep approximately 30% of censoring.

Under the above scenarios, we consider 1000 Monte Carlo replicas and generate

$$\{(Y_{1i}^*, Y_{2i}^*)\}_{i=1}^n \stackrel{ind}{\sim} \text{BS}(\mu_{1i}, \mu_{2i}, \phi, 1, \rho)$$

and compute $Y_i = Y_{1i}^* U_i$, (which is the response variable, with $U_i = I\{Y_{2i}^* > 1\}$) in each Monte Carlo loop. One of the main interest is the estimation of the parameter ρ , which controls the sample selection bias. In this simulation, we consider $\rho = 0, \pm 0.2, \pm 0.5, \pm 0.7$ and also $\phi = 1.2$. For the estimation of the parameters, we used the package `optim` implemented in software R ([33]). Initial values for the maximization procedure were obtained by the Heckman's two-step method.

To examine the MLE performance of the models, we compute the empirical mean of the parameter estimates and the root mean square error (RMSE) for each scenario considered. The results for Scenario 1 with $\rho = 0.7$ is presented in Table 1. Table 2

gives the simulated results for Scenario 2 with $\rho = 0.7$. More simulated results (with $\rho = 0, \pm 0.2, \pm 0.5, -0.7$) can be found in the Supplementary Material.

Table 1. Empirical mean of the maximum likelihood estimates with their respective root mean square error (RMSE) for the Birnbaum-Saunders sample selection model under Scenario 1 with $\rho = 0.7$. Sample size $n = 500, n = 1000$ and $n = 2000$ with $N = 1000$ Monte Carlo replicates.

Parameters	n	Configuration 1		Configuration 2		Configuration 3	
		Mean	RMSE	Mean	RMSE	Mean	RMSE
$\gamma_1 = 1.600$	500	1.609	0.083	1.603	0.214	1.606	0.086
	1000	1.605	0.060	1.610	0.155	1.606	0.062
	2000	1.602	0.042	1.607	0.108	1.601	0.044
$\gamma_2 = 0.800$	500	0.814	0.095	0.823	0.199	0.809	0.125
	1000	0.806	0.064	0.800	0.143	0.808	0.086
	2000	0.805	0.045	0.806	0.100	0.805	0.061
$\gamma_3 = 0.200$	500	0.208	0.080	0.216	0.344	0.206	0.117
	1000	0.199	0.057	0.190	0.241	0.199	0.078
	2000	0.200	0.042	0.197	0.167	0.199	0.056
$\gamma_4 = 0.700$	500	0.712	0.089	0.713	0.055	0.709	0.089
	1000	0.704	0.065	0.706	0.038	0.707	0.065
	2000	0.703	0.043	0.704	0.028	0.703	0.041
$\beta_1 = 1.000$	500	1.004	0.068	0.994	0.123	1.003	0.070
	1000	1.000	0.046	0.998	0.091	1.000	0.046
	2000	1.000	0.033	0.999	0.064	1.000	0.033
$\beta_2 = 0.700$	500	0.695	0.066	0.700	0.114	0.703	0.089
	1000	0.697	0.046	0.703	0.082	0.698	0.061
	2000	0.699	0.033	0.700	0.056	0.700	0.042
$\beta_3 = 1.100$	500	1.100	0.056	1.104	0.185	1.099	0.078
	1000	1.099	0.036	1.096	0.134	1.100	0.054
	2000	1.100	0.029	1.098	0.097	1.100	0.039
$\phi = 1.200$	500	1.220	0.144	1.227	0.108	1.218	0.142
	1000	1.216	0.100	1.209	0.071	1.214	0.099
	2000	1.207	0.064	1.205	0.048	1.205	0.064
$\rho = 0.700$	500	0.690	0.115	0.702	0.091	0.690	0.117
	1000	0.691	0.083	0.701	0.064	0.690	0.084
	2000	0.696	0.053	0.701	0.045	0.696	0.053

In Table 1 we observe a good performance of the maximum likelihood estimators of the BS model. The bias is small and tends to 0 as the sample size increases for all cases considered. The RMSE goes to 0 as the sample size increases, as expected. This indicates the consistency of the maximum likelihood estimators of our proposed sample selection model.

Despite observing a similar result in Table 2, we can see an inconsistency in Scenario 2, Configuration 3 for $\rho = -0.7$, presented in Table 7 of the Supplementary Material. In this case, the RMSE does not decrease as the sample size increases. This fact reinforces the importance of exclusion selection, as indicated for the classic Heckman, the Heckman-t, and the Heckman skew-normal models.

Table 2. Empirical mean of the maximum likelihood estimates with their respective root mean square error (RMSE) for the Birnbaum-Saunders sample selection model under Scenario 2 with $\rho = 0.7$. Sample size $n = 500, n = 1000$ and $n = 2000$ with $N = 1000$ Monte Carlo replicates.

Parameters	n	Configuration 1		Configuration 2		Configuration 3	
		Mean	RMSE	Mean	RMSE	Mean	RMSE
		$\gamma_1 = 1.600$	500	1.606	0.085	1.616	0.192
	1000	1.605	0.061	1.610	0.131	1.603	0.061
	2000	1.601	0.042	1.603	0.091	1.600	0.043
$\gamma_2 = 0.800$	500	0.815	0.089	0.798	0.330	0.808	0.120
	1000	0.808	0.066	0.797	0.230	0.804	0.085
	2000	0.803	0.044	0.799	0.156	0.801	0.061
$\gamma_3 = 0.200$	500	0.203	0.023	0.204	0.016	0.207	0.113
	1000	0.202	0.016	0.201	0.010	0.198	0.076
	2000	0.201	0.011	0.201	0.007	0.200	0.053
$\beta_1 = 1.000$	500	1.044	0.184	1.002	0.121	1.093	0.272
	1000	1.015	0.120	0.996	0.089	1.035	0.175
	2000	1.001	0.043	1.000	0.060	1.026	0.144
$\beta_2 = 0.700$	500	0.678	0.112	0.697	0.181	0.653	0.163
	1000	0.692	0.073	0.705	0.137	0.680	0.106
	2000	0.698	0.038	0.696	0.096	0.687	0.084
$\beta_3 = 1.100$	500	1.095	0.028	1.100	0.008	1.087	0.085
	1000	1.098	0.018	1.100	0.006	1.096	0.057
	2000	1.100	0.009	1.100	0.004	1.096	0.042
$\phi = 1.200$	500	1.225	0.162	1.220	0.114	1.223	0.169
	1000	1.213	0.104	1.212	0.081	1.219	0.127
	2000	1.206	0.078	1.206	0.053	1.208	0.087
$\rho = 0.700$	500	0.609	0.341	0.691	0.122	0.519	0.498
	1000	0.665	0.212	0.694	0.076	0.623	0.302
	2000	0.694	0.077	0.699	0.053	0.646	0.260

4.2. Size and power of the likelihood ratio test

We here investigate the size and power of the likelihood ratio (LR) test for checking the presence of sample selection bias through a small Monte Carlo simulation. In order works, we are interested in testing the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_1 : \rho \neq 0$. We consider Scenario 1 described in the previous subsection under three different configurations. For Configuration 1, we consider

$$\{(Y_{1i}^*, Y_{2i}^*)\}_{i=1}^n \stackrel{iid}{\sim} \text{BS}(\mu_{1i}, \mu_{2i}, \phi, 1, \rho)$$

and compute $Y_i = Y_{1i}^* U_i$, with $U_i = I\{Y_{2i}^* > 1\}$. In Configuration 2, we generate

$$\begin{pmatrix} Y_{1i}^* \\ Y_{2i}^* \end{pmatrix} \stackrel{iid}{\sim} \mathcal{N} \left(\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \begin{pmatrix} \phi^2 & \rho\phi \\ \rho\phi & 1 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

and compute $Y_i = Y_{1i}^* U_i$, with $U_i = I\{Y_{2i}^* > 0\}$. In order to keep 30% of censoring approximately, we set $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, $j = 2, 3$ and $x_{4i} \stackrel{iid}{\sim} \mathcal{N}(0, 4)$. Finally, in configuration

$$\{(Y_{1i}^*, Y_{2i}^*)\}_{i=1}^n \stackrel{iid}{\sim} \text{SN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$$

and compute $Y_i = Y_{1i}^* U_i$, where $U_i = I\{Y_{2i}^* > 0\}$ and $\text{SN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ denotes the bivariate skew-normal distribution with location vector $\boldsymbol{\mu} = (\mu_{1i}, \mu_{2i})$, variance-covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \phi^2 & \rho\phi \\ \rho\phi & 1 \end{pmatrix}$ and skewness vector $\boldsymbol{\lambda} = (1, 0)$. Again, to keep 30% of censoring, we consider $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, $j = 2, 3$ and $x_{4i} \stackrel{iid}{\sim} \mathcal{N}(0, 4)$.

We here consider, for each setting, the likelihood ratio test based on the BS, classical Heckman, and skew-normal sample selection models. We do not present the results related to the Student-t sample selection model, since an instability was observed, yielding strange results. A deeper investigation is needed in this case and this is out of the scope of this paper. We set the significance level at $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$, and sample sizes $n = 500, 1000, 2000$. The number of Monte Carlo replicas was 1000.

Tables 3, 4 and 5 show the rejection proportion of the null hypothesis ($H_0 : \rho = 0$) based on the likelihood ratio test under BS, classical Heckman and skew-normal sample selection models, with $\rho = 0$. Tables 6, 7 and 8 give us the proportion of rejections of the null hypothesis of the LR test under BS, classic Heckman and skew-normal models as the data generator having sample selection bias, respectively, for some values of ρ .

Table 3. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - without sample selection bias - under the BS, Heckman and skew-normal sample selection models. Data generated with Birnbaum-Saunders distribution and approximate 30% censorship.

α	n	BS	Normal	SN
1%	500	0.7	1.1	1.2
	1000	1.3	1.3	1.3
	2000	1.3	1.1	1.1
5%	500	4.7	4.4	4.4
	1000	6.7	5.3	5.5
	2000	5.0	5.0	5.0
10%	500	9.8	8.9	9.1
	1000	11.2	10.7	10.8
	2000	9.9	9.9	9.8

Table 4. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - without sample selection bias - under the BS, Heckman and skew-normal sample selection models. Data generated with Normal distribution and approximate 30% censorship.

α	n	BS	Normal	SN
1%	500	1.7	1.6	1.5
	1000	2.5	0.6	0.9
	2000	3.7	1.9	2.0
5%	500	8.9	6.8	6.8
	1000	9.6	6.0	6.5
	2000	12.8	6.9	7.2
10%	500	16.5	12.2	12.1
	1000	18.1	12.1	12.4
	2000	23.2	12.6	12.3

Table 5. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - without sample selection bias - under the BS, Heckman and skew-normal sample selection models. Data generated with skew-normal distribution and approximate 30% censorship.

α	n	BS	Normal	SN
1%	500	1.0	0.9	0.9
	1000	2.0	1.1	1.5
	2000	2.3	0.8	0.8
5%	500	6.4	4.6	4.7
	1000	7.1	6.4	7.0
	2000	8.2	5.1	5.2
10%	500	12.2	11.0	11.1
	1000	13.8	12.2	13.1
	2000	14.4	11.3	11.0

Results in Table 3 indicate that the likelihood ratio test covers well the nominal significance levels under the BS, classical Heckman, and skew-normal models. We observe a slight advantage of the BS model in some cases. In general, all models performed well for detecting the absence of sample selection bias. In the case of data generated using the classic Heckman and skew-normal models (Tables 7 and 8), we observed higher rejection rates for the BS model. When data are generated by Heckman or skew-normal models, we apply an exponential transform to the data to fit the BS model, since its support is on \mathbb{R}^+ .

Looking at the results from the Table 6, we see that the likelihood ratio test based on the BS model is more powerful than the test based on the Heckman and skew-normal sample selection models, which is expected since we are generating the data under the BS assumption. When data are generated from the Heckman and skew-normal sample selection models (Tables 7 and 8), we observed similar power among the three models.

In summary, we observe a satisfactory finite-sample behavior of the likelihood ratio test based on the BS model for checking the presence of sample selection bias.

Table 6. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - with sample selection bias (Scenario 1 with $\rho = \pm 0.3, \pm 0.7$) - under the BS, Heckman and skew-normal sample selection models. Data generated with Birnbaum-Saunders distribution and approximate 30% censorship.

ρ	n	BS	Normal	SN
-0.7	500	90.6	84.1	86.1
	1000	99.5	99.3	99.3
	2000	100	100	100
-0.3	500	11.4	8.5	9.1
	1000	29.7	20.7	21.2
	2000	62	46.8	47.2
0.3	500	14.2	8.0	7.8
	1000	26.9	18.8	19.0
	2000	61.8	47.2	46.0
0.7	500	90.6	85.2	84.5
	1000	99.2	98.8	98.8
	2000	100	100	100

Table 7. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - with sample selection bias (Scenario 1 with $\rho = \pm 0.3, \pm 0.7$) - under the BS, Heckman and skew-normal sample selection models. Data generated with normal distribution and approximate 30% censorship.

ρ	n	BS	Normal	SN
-0.7	500	97.5	98.8	98.8
	1000	99.9	100	100
	2000	100	100	100
-0.3	500	24.4	24.8	24.7
	1000	42.2	45.0	45.1
	2000	80.1	84.7	86.0
0.3	500	22.4	23.4	23.3
	1000	43.8	46.9	47.0
	2000	81.4	84.2	84.5
0.7	500	97.3	99.2	99.2
	1000	100	100	100
	2000	100	100	100

Table 8. Rejection rate (in percentage) of the null hypothesis based on the likelihood ratio test - with sample selection bias (Scenario 1 with $\rho = \pm 0.3, \pm 0.7$) - under the BS, Heckman and skew-normal sample selection models. Data generated with skew-normal distribution and approximate 30% censorship.

ρ	n	BS	Normal	SN
-0.7	500	96.4	94.4	94.4
	1000	100	99.9	99.9
	2000	100	100	100
-0.3	500	21.1	16.0	15.9
	1000	41.2	33.8	33.0
	2000	77.8	67.4	67.6
0.3	500	11.1	16.4	16.4
	1000	20.8	29.9	31.2
	2000	46.4	64.3	64.0
0.7	500	88.1	92.1	92.2
	1000	99.6	99.8	99.8
	2000	100	100	100

5. BS model under misspecification

We now investigate the Birnbaum-Saunders sample selection model under misspecification. For this, we construct a sample selection model based on a bivariate gamma distribution introduced by [8], which has gamma marginals. More specifically, we consider $\{(Y_{1i}^*, Y_{2i}^*)\}_{i=1}^n$ following the Cherian bivariate gamma distribution with correlation parameter

$$0 < \rho < \frac{\min(\kappa_1, \kappa_2)}{\sqrt{\kappa_1 \kappa_2}},$$

with Y_{1i}^* has gamma distribution with mean $\mu_{1i} > 0$ and shape parameter $\kappa_1 > 0$, and Y_{2i}^* has gamma distribution with mean $\mu_{2i} > 0$ and shape parameter $\kappa_2 > 0$, for $i = 1, \dots, n$. We also assume independence among the pairs. Hence, the gamma sample selection model is defined by previous assumptions and $Y_i = U_i Y_{1i}^*$, where

$U_i = I\{Y_{2i}^* > 1\}$, for $i = 1, \dots, n$. For this model, the absence of sample selection bias occurs as a limiting case when $\rho \rightarrow 0^+$.

Table 9. Empirical mean of the maximum likelihood estimates of the parameters with their respective root mean square error (RMSE) based on the BS, classical Heckman, skew-normal and Student-t sample selection models. Sample size equal to 1000.

Presence of the Exclusion Restriction								
Parameters	Birnbaum-Saunders		Heckman		skew-normal		Student-t	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
$\gamma_1 = 2.5$	2.248	0.315	1.441	1.067	1.717	0.797	1.682	0.836
$\gamma_2 = 1.0$	0.831	0.210	0.742	0.273	0.681	0.336	0.852	0.179
$\gamma_3 = 2.5$	2.098	0.465	1.870	0.648	1.716	0.807	2.149	0.398
$\gamma_4 = 2.0$	1.642	0.410	1.480	0.536	1.358	0.663	1.709	0.334
$\beta_1 = -0.5$	-0.545	0.060	-0.772	0.275	0.132	0.650	-0.716	0.222
$\beta_2 = 1.5$	1.506	0.041	1.500	0.031	1.499	0.040	1.496	0.031
$\beta_3 = 2.0$	2.016	0.052	2.000	0.038	2.003	0.048	1.989	0.040
$\alpha_1 = 2.0$	2.547	0.622	0.789	1.211	1.235	0.783	0.703	1.297
$\rho = 0.4$	0.431	0.113	0.346	0.134	0.531	0.212	0.284	0.186

Absence of the Exclusion Restriction								
Parameters	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
$\gamma_1 = 2.5$	2.260	0.294	1.444	1.064	1.641	0.880	1.676	0.839
$\gamma_2 = 1.0$	0.803	0.231	0.729	0.284	0.685	0.327	0.842	0.188
$\gamma_3 = 2.5$	2.055	0.497	1.846	0.670	1.736	0.782	2.121	0.422
$\beta_1 = -0.5$	-0.550	0.071	-0.766	0.271	-0.070	0.591	-0.688	0.200
$\beta_2 = 1.5$	1.508	0.044	1.498	0.031	1.502	0.029	1.488	0.035
$\beta_3 = 2.0$	2.024	0.063	1.994	0.046	2.003	0.039	1.964	0.068
$\alpha_1 = 2.0$	2.534	0.609	0.787	1.213	1.131	0.890	0.698	1.302
$\rho = 0.4$	0.431	0.140	0.314	0.174	0.472	0.192	0.169	0.330

We run a small Monte Carlo simulation with 1000 replicas. In each loop, we generate a gamma sample selection model, as described above. We set the outcome equation by $\mu_{1i} = \exp(-0.5 + 1.5x_{2i} + 2x_{3i})$ and the selection equation we consider two scenarios:

- (i) exclusion restriction given by $\mu_{2i} = \exp(2.5 + x_{2i} + 2.5x_{3i} + 2x_{4i})$;
- (ii) absence of exclusion restriction by assuming $\mu_{2i} = \exp(2.5 + x_{2i} + 2.5x_{3i})$, where $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, $j = 2, 3, 4$, $\forall i$, are explanatory variables kept fixed along the simulation, for $i = 1, \dots, 1000$. Further, we set $\kappa_1 = 2$, $\kappa_2 = 1$ and $\rho = 0.4$. Under these settings, we obtain about 30% of censoring.

Table 9 presents the empirical mean of the maximum likelihood estimates of the parameters with their respective root mean square error (RMSE) based on the BS, classical Heckman, skew-normal and Student-t sample selection models, under presence and absence of exclusion restriction.

From the results given in Table 9, we observe that the Birnbaum-Saunders model provides satisfactory results in both cases (with and without exclusion restriction), in contrast with the Heckman, skew-normal and Student-t models. We call the attention that the BS model is fitted without transformation of the original data, while a logarithmic transformation is used for the other models.

6. MEPS data analysis

In this section we analyze the data on ambulatory expenditures (outcomes of interest, here denoted by Y) from the 2001 Medical Expenditure Panel Survey (MEPS), which can be obtained from the R package *ssmrob* by [36]. The data are composed of 3328 observations, where 526 (15.8%) of them correspond to zero expense amounts. Furthermore, the data consist of the following explanatory variables:

- x_2 (age): age measured in tens of years;
- x_3 (fem): gender (female);
- x_4 (educ): represents years of schooling;
- x_5 (blhisp): ethnicity;
- x_6 (totcr): number of chronic diseases;
- x_7 (ins): insurance status;
- x_8 (inc): income.

These data were also analyzed by [7], [23], [29] and [37]. Here, the indicator variable U represents the decision to spend or not with medical care. According to these authors, the decision to spend is probably related to the value of the expense and, therefore, it is natural to consider a sample selection model for these data. Further, previous analyses show that the normality assumption for these data is not suitable. For instance, according to [29], in the presence of deviations from normality due to asymmetry, the classical Heckman model can erroneously attribute null sample selection bias. We call the attention that the distribution of expenditures is highly skewed, as mentioned by [23]. In this sense, it is very natural to consider our proposed BS sample selection model for analyzing the MEPS data. Furthermore, our model does not require any transformation of the data and, therefore, the interpretation of the parameters can be done in the original scale, in contrast with the existent models in the literature which require some transformation of the data.

We consider the following outcome and selection equations, respectively:

$$\begin{aligned}\log \mu_{1i} &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7}, \\ \log \mu_{2i} &= \gamma_1 + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i4} + \gamma_5 x_{i5} + \gamma_6 x_{i6} + \gamma_7 x_{i7} + \gamma_8 x_{i8},\end{aligned}$$

for $i = 1, 2, \dots, 3328$. The income explanatory variable is included in the selection equation but not in the outcome equation. As mentioned by [23], the inclusion of this variable is debatable. In this way, the exclusion restriction is imposed. In Table 10 we present parameter estimates with respective p-values of the classic Heckman model, the robust version of the two-step method proposed by [37], the skew-normal and Heckman-t models, in addition to our proposed BS sample selection model.

Table 10. Estimates of the parameters with their respective p-value under the classic Heckman, robust two-step method, Birnbaum-Saunders (BS), skew-normal (SN) and Heckman-t (St) sample selection models.

Selection Equation										
Parameters	BS	p-value	Heckman	p-value	Robust	p-value	SN	p-value	St	p-value
(Intercept)	0.117	0.618	-0.676	0.000	-0.700	0.000	-0.899	0.000	-0.748	0.000
age (x_2)	0.090	0.005	0.088	0.001	0.095	0.001	0.087	0.001	0.099	0.001
fem (x_3)	0.722	0.000	0.663	0.000	0.704	0.000	0.645	0.000	0.725	0.000
educ (x_4)	0.068	0.000	0.062	0.000	0.062	0.000	0.060	0.000	0.065	0.000
blhisp (x_5)	-0.399	0.000	-0.364	0.000	-0.389	0.000	-0.351	0.000	-0.394	0.000
totcr (x_6)	0.806	0.000	0.797	0.000	0.834	0.000	0.775	0.000	0.890	0.000
ins (x_7)	0.179	0.014	0.170	0.007	0.173	0.007	0.170	0.005	0.180	0.008
inc (x_8)	0.003	0.017	0.003	0.040	0.003	0.059	0.003	0.038	0.003	0.040
Outcome Equation										
Parameters	BS	p-value	Heckman	p-value	Robust	p-value	SN	p-value	St	p-value
(Intercept)	5.779	0.000	5.044	0.000	5.409	0.000	6.452	0.000	5.206	0.000
age (x_2)	0.246	0.000	0.212	0.000	0.200	0.000	0.202	0.000	0.207	0.000
fem (x_3)	0.411	0.000	0.348	0.000	0.252	0.000	0.287	0.000	0.307	0.000
Educ (x_4)	-0.007	0.503	0.019	0.076	0.013	0.255	0.013	0.185	0.017	0.091
blhisp (x_5)	-0.215	0.000	-0.219	0.000	-0.153	0.018	-0.201	0.000	-0.193	0.001
totcr (x_6)	0.589	0.000	0.540	0.000	0.480	0.000	0.492	0.000	0.513	0.000
ins (x_7)	-0.061	0.216	-0.030	0.557	-0.068	0.187	-0.080	0.106	-0.053	0.298
$\sigma \setminus \phi$	0.703	--	1.271	--	1.320	--	1.736	--	1.195	--
ρ	0.273	0.000	-0.131	0.375	-0.523	0.007	-0.339	0.016	-0.322	0.008

Our main interest here is to test the hypotheses $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$. We use the Wald test to do this. Under the classical Heckman model, we do not reject the null hypothesis H_0 by using any usual significance level (p-value > 0.1). Conversely, under the Heckman-BS model, the robust version of the two-step method, and the Heckman-t model, we reject the null hypothesis H_0 (p-value < 0.001) by using any significance level in favor of the alternative hypothesis that states that there is sample selection bias. Under the skew-normal sample selection model, we reject the null hypothesis for any significance level greater than 1.6%. We obtain the same conclusions by using the likelihood ratio test. The inference based on the BS sample selection model agrees with those obtained by robust models by [23] and [37]. This empirical illustration shows that the Birnbaum-Saunders sample selection model can be robust and has strong potential for practical situations, besides its relative mathematical simplicity and computational speed experienced by us along with simulations and application.

The estimated value of the skewness parameter of the skew-normal model was $\lambda = -1.66$ (p-value < 0.001), which suggests that MEPS data are skewed. This reinforces the importance of considering sample selection models having this feature like our proposed BS model.

Figure 1 displays the quantile versus quantile (QQ) plots with the simulated envelope of the GCS and Normalized Quantile residuals for the Heckman-BS model discussed at the end of Section 3. From this figure we can observe a good performance of the Birnbaum-Saunders sample selection model for fitting the MEPS data.

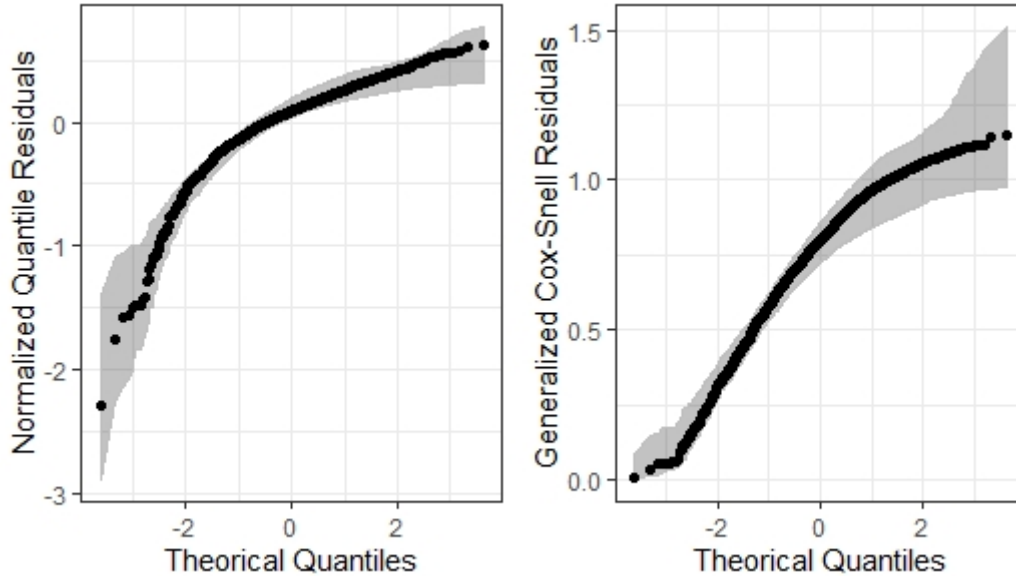


Figure 1. QQ plots and their simulated envelopes for the Normalized Quantile residual (left) and Generalized Cox-Snell residual (right) for the Medical Expenditure Panel Survey data under BS assumption.

7. Concluding remarks

We introduced a sample selection model based on the bivariate Birnbaum-Saunders distribution. Monte Carlo simulations were performed to evaluate the finite-sample performance of the MLEs of our model. We observed satisfactory results in terms of bias and consistency, with the exception in case of absence of exclusion restriction. Therefore, our recommendation is to consider this restriction, which is in agreement with the literature on sample selection models.

Under the specification of the Heckman-BS model, our simulations of power and level of empirical significance, to test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$, using the likelihood ratio, indicated that all models used maintained the correct nominal level. Our model, however, presented greater power for all the levels of significance used especially for the RV test. The skew-normal and Heckman classic models had similar results for the power of both tests. We expected higher results for the skew-normal model since the data are skewed.

Our model was also evaluated for lack of specification for skewed and positive data with bivariate distribution of Cherian, and again, the estimates obtained under the Heckman-BS model presented a lower bias and lower RMSE when compared with the other models.

We suggest consider the Heckman-BS model instead of other parametric models whenever the response variable is skewed and positive. Our model provides greater flexibility for the modeling of positive skewness data when compared to the classic Heckman model, maintaining the same amount of parameters of the classic model, which is an advantage when compared also to the skew-normal and Heckman-t models. In addition, when one is dealing with positive outcomes, our model facilitates the interpretation of the parameters, since there is no need to transform the response variable into its fit.

If data can be negative-valued, one could apply an exponential transformation to

the outcomes so that our proposed Birnbaum-Saunders model could be considered in this case, similarly as done when Heckman, Student-t and skew-normal models are applied so that data are log-transformed. We also would like to grab attention that positive outcomes are very frequent in the applications where sample selection models are used but not negative outcomes. Furthermore, one of the main messages of our paper is not to transform data. Therefore, our recommendation for negative outcomes is to consider existing models such as Heckman, Student-t, and skew-normal.

One point that deserves further investigation is to obtain conditions ensuring the consistency and asymptotic normality for the maximum likelihood estimators of the BS sample selection model. These asymptotic results have not been addressed even for the existing parametric sample selection models. We are working on this point and hope to report the results as soon as possible.

Acknowledgments

We thank the Associate Editor and two anonymous Referees for their important comments and suggestions which lead to an improvement of this paper. This work is part of the Ph.D. thesis by Fernando de Souza Bastos realized at the Department of Statistics from the Universidade Federal de Minas Gerais. Wagner Barreto-Souza thanks the agencies CNPq-Brazil (grant 305543/2018-0), FAPEMIG-Brazil (grant APQ-00786-14), and KAUST Research Fund NIH 1R01EB028753-01 for the financial support.

References

- [1] H. Ahn and J.L. Powell, *Semiparametric Estimation of Censored Selection Models With a Nonparametric Selection Mechanism*, Journal of Econometrics 58 (1993), pp. 3–29.
- [2] R. Arellano-Valle, G. del Pino, and E.S. Martin, *Definition and Probabilistic Properties of Skew-Distributions*, Statistics & Probability Letters 58 (2002), pp. 111–121.
- [3] R.B. Arellano-Valle, M.D. Branco, and M.G. Genton, *A Unified View on Skewed Distributions Arising From Selections*, Canadian Journal of Statistics 34 (2006), pp. 581–601.
- [4] B.C. Arnold and R.J. Beaver, *Hidden Truncation Models*, Sankhyā: The Indian Journal of Statistics 62 (2000), pp. 23–35.
- [5] A. Azzalini, *The Skew-normal Distribution and Related Multivariate Families*, Scandinavian Journal of Statistics 32 (2005), pp. 159–188.
- [6] M. Barros, M. Galea, M. González, and V. Leiva, *Influence diagnostics in the tobit censored response model*, Statistical Methods & Applications 19 (2010), pp. 379–397.
- [7] C.A. Cameron and P.K. Trivedi, *Microeconometrics Using Stata*, revised ed., TX: Stata press, College Station, 2009.
- [8] K.C. Cherian, *A bi-variate correlated gamma type distribution function*, Journal of the Indian Mathematical Society 5 (1941), pp. 133–144.
- [9] M. Das, W.K. Newey, and F. Vella, *Nonparametric Estimation of Sample Selection Models*, The Review of Economic Studies 70 (2003), pp. 33–58.
- [10] P. Ding, *Bayesian Robust Inference of Sample Selection Using Selection-t Models*, Journal of Multivariate Analysis 124 (2014), pp. 451–464.
- [11] P.K. Dunn and G.K. Smyth, *Randomized quantile residuals*, Journal of Computational and Graphical Statistics 5 (1996), pp. 236–244.
- [12] F. Hampel, *The influence curve and its role in robust estimation*, Journal of the American Statistical Association 69 (1974), pp. 383–393.
- [13] R.S. Hartman, *A Monte Carlo Analysis of Alternative Estimators in Models Involving Selectivity*, Journal of Business & Economic Statistics 9 (1991), pp. 41–49.

- [14] J.J. Heckman, *The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models*, *Annals of Economic and Social Measurement* 5 (1976), pp. 475–492.
- [15] J.J. Heckman, *Sample Selection Bias as a Specification Error*, *Econometrica* 47 (1979), pp. 153–161.
- [16] P. Huber and E. Ronchetti, *Robust Statistics*, Wiley, New York, 2009.
- [17] A. Jamalizadeh and D. Kundu, *A Multivariate Birnbaum-Saunders Distribution Based on the Multivariate Skew Normal Distribution*, *Journal of the Japan Statistical Society* 45 (2015), pp. 1–20.
- [18] D. Kundu, N. Balakrishnan, and A. Jamalizadeh, *Bivariate Birnbaum-Saunders Distribution and Associated Inference*, *Journal of Multivariate Analysis* 101 (2010), pp. 113–125.
- [19] L.F. Lee, *Generalized Econometric Models With Selectivity*, *Econometrica* 51 (1983), pp. 507–512.
- [20] S.F. Leung and S. Yu, *On the Choice Between Sample Selection and TwoPart Models*, *Journal of Econometrics* 72 (1996), pp. 197–229.
- [21] S.F. Leung and S. Yu, *Collinearity and Two-Step Estimation of Sample Selection Models: Problems, Origins, and Remedies*, *Computational Economics* 15 (2000), pp. 173–199.
- [22] W. Manning, N. Duan, and W. Rogers, *Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models*, *Journal of Econometrics* 35 (1987), pp. 59 – 82.
- [23] Y.V. Marchenko and M.G. Genton, *A Heckman Selection-t Model*, *Journal of the American Statistical Association* 107 (2012), pp. 304–317.
- [24] W. Miao, P. Ding, and Z. Geng, *Identifiability of Normal and Normal Mixture Models with Nonignorable Missing Data*, *Journal of the American Statistical Association* 111 (2016), pp. 1673–1683.
- [25] K. Nawata, *A Note on the Estimation of Models With Sample-Selection Biases*, *Economics Letters* 42 (1993), pp. 15 – 24.
- [26] K. Nawata, *Estimation of Sample Selection Bias Models By the Maximum Likelihood Estimator and Heckman’s Two-Step Estimator*, *Economics Letters* 45 (1994), pp. 33 – 40.
- [27] F.D. Nelson, *Efficiency of The Two-Step Estimator for Models With Endogenous Sample Selection*, *Journal of Econometrics* 24 (1984), pp. 181 – 196.
- [28] W.K. Newey, *Two-Step Series Estimation of Sample Selection Models*, *Econometrics Journal* 12 (2009), pp. 217–229.
- [29] E.O. Ogundimu and J.L. Hutton, *A Sample Selection Model With Skew-Normal Distribution*, *Scandinavian Journal of Statistics* 43 (2016), pp. 172–190.
- [30] H.J. Paarsch, *A Monte Carlo Comparison of Estimators for Censored Regression Models*, *Journal of Econometrics* 24 (1984), pp. 197–213.
- [31] J.L. Powell, *Estimation of Semiparametric Models*, *Handbook of Econometrics* 4 (1994), pp. 2443–2521.
- [32] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Programming*, 2nd ed., Cambridge University Press, 1992.
- [33] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2016). Available at <https://www.R-project.org/>.
- [34] D. Rubin, *Inference and missing data*, *Biometrika* 63 (1976), pp. 581–592.
- [35] R.M. Stolzenberg and D.A. Relles, *Theory Testing in a World of Constrained Research Design: The Significance of Heckman’s Censored Sampling Bias Correction for Nonexperimental Research*, *Sociological Methods & Research* 18 (1990), pp. 395–415.
- [36] M. Zhelonkin, M.G. Genton, and E. Ronchetti, *R Package ssmrob: Robust Estimation and Inference in Sample Selection Models*, CRAN (2014). Available at <https://CRAN.R-project.org/package=ssmrob>, R package version 0.4.
- [37] M. Zhelonkin, M.G. Genton, and E. Ronchetti, *Robust Inference in Sample Selection Models*, *Journal of the Royal Statistical Society - Series B* 78 (2016), pp. 805–827.