Statistica Sinica Preprint No: SS-2016-0300R2	
Title	Generalization of Heckman selection model to nonignorable nonresponse using call-back information
Manuscript ID	SS-2016.0300
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0300
Complete List of Authors	Baojiang Chen Pengfei Li and Jing Qin
Corresponding Author	Baojiang Chen
E-mail	baojiang.chen@uth.tmc.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0300

function, 1 if A is true and 0 otherwise. Furthermore, in the Heckman model, it is typically assumed that $Corr(\epsilon_{1i}, \epsilon_{2i}) = \rho_{12}$ and $(\epsilon_{1i}, \epsilon_{2i})^\tau$ follows a bivariate normal distribution.

Note that

$$\begin{aligned} P(R_i = 1|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) &= P(Z_i > 0|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= P(\epsilon_{2i} > -\mathbf{X}_{2i}^\tau \boldsymbol{\gamma}|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= \Phi\Big(\frac{\mathbf{X}_{2i}^\tau \boldsymbol{\gamma} + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\Big), \end{aligned}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal random variable. This means that the Heckman model leads to a nonignorably missing mechanism when $\rho_{12} \neq 0$, since the missing probability depends on y_i.

Heckman (1979) introduced a two-step procedure to estimate the coefficients in the response and missing-data models (1.1) and (1.2). Alternatively, one can estimate the coefficients using a likelihood-based method. The likelihood function of the unknown parameters is

$$L_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}) = \prod_{i=1}^{n} \Big[\{P(R_i = 1, Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i})\}^{R_i} \{P(R_i = 0|\mathbf{X}_{1i}, \mathbf{X}_{2i})\}^{1-R_i}\Big],$$

where

$$\begin{aligned} P(R_i = 1, Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}) &= P(R_i = 1|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) P(Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= \Phi\Big(\frac{\mathbf{X}_{2i}^\tau \boldsymbol{\gamma} + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\Big) \\ &\quad \times \sigma^{-1} \phi\Big(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1}{\sigma}\Big), \end{aligned} \tag{1.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0300

$$P(R_i = 0|\mathbf{X}_{1i}, \mathbf{X}_{2i}) = P(\epsilon_{2i} < -\mathbf{X}_{2i}^\tau \gamma | \mathbf{X}_{2i}) = \Phi(-\mathbf{X}_{2i}^\tau \gamma).$$

Here $\phi(x)$ is the probability density function of the standard normal random variable. Consequently, the log-likelihood of the unknown parameters is

$$\begin{aligned} & \ell_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}) \\ = & \sum_{i=1}^{n} \Big[R_i \log \Big\{ \Phi\Big(\frac{\mathbf{X}_{2i}^\tau \gamma + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}} \Big) \Big\} - R_i \log(\sigma) \\ & + R_i \log \Big\{ \phi\Big(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1}{\sigma} \Big) \Big\} + (1 - R_i) \log\{\Phi(-\mathbf{X}_{2i}^\tau \gamma)\} \Big]. \end{aligned} \tag{1.4}$$

Maximizing (1.4) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ, and ρ_{12}, we obtain the maximum likelihood estimators of the unknown parameters:

$$(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \tilde{\sigma}, \tilde{\rho}_{12}) = \arg \max_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}} \ell_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}).$$

4 Incorporating call-back information by generalizing the Heckman selection model

In this section, we discuss how to incorporate call-back information by generalizing the Heckman selection model. We further study the consistency of the estimator of $\boldsymbol{\beta}_1$ in (1.1) under model misspecification. For convenience of presentation, we assume that there is a single call-back. For multiple call-backs, see the Supplementary Material.