# Note on the Uniqueness of the Maximum Likelihood Estimator for a Heckman's Simultaneous Equations Model

**Kentaro Akashi*** **Tetsushi Horie**†

Revised version, February 2022

**Abstract**

It is shown that the likelihood function of a Heckman's simultaneous equations model is identified by complementing the approach of parameter transformation. Therefore, the expectation of the log-likelihood function has a single maximum. Thus, the maximum likelihood estimator becomes asymptotically consistent without an initial consistent estimator. Additionally, the approach can show the uniqueness of the log-likelihood functions for the simultaneous Tobit, sample selection (Type 2 Tobit), and simultaneous generalized selectivity models.

**Keywords:** Heckman's simultaneous equations model, Full information maximum likelihood estimator, Exclusion restriction, Reparametrization method, Unique maximum.
**JEL Classification:** C35, C62

## 1 Introduction

We consider two structural equations of the following form

$$\begin{aligned} y_{1i} &= \beta_1 y_{2i} + \boldsymbol{\gamma}_1' \mathbf{z}_i + u_{1i} \text{ , and} && (1.1) \\ y_{2i} &= \mathbb{1}\{\beta_2 y_{1i} + \beta_3 y_{2i} + \boldsymbol{\gamma}_2' \mathbf{z}_i + u_{2i} \geq 0\} \text{ ,} && (1.2) \end{aligned}$$

where $y_{1i}$ and $y_{2i}$ are endogenous variables, and $\mathbf{z}_i$ is the $K$-variate exogenous variables independent of the error terms. The error terms $(u_{1i},\ u_{2i})$ follow a bivariate normal distribution with zero mean, i.e., $\mathcal{E}[u_{1i}^2] = \sigma^2 > 0$ and $\mathcal{E}[u_{1i}u_{2i}] = \sigma_{12}$ $(\sigma_{12}^2 < \sigma^2)$. Without loss of generality, we use the normalization of $\mathcal{E}[u_{2i}^2] = 1$. The indicator function $\mathbb{1}\{.\}$ takes the value of 1 if the argument is true; otherwise, it is 0.

These structural equations constitute a simultaneous equations model investigated in a seminal paper by Heckman (1978). We show that the expectation of the log-likelihood function has a unique maximum using the reparametrization method proposed by Olsen (1978) to prove the uniqueness of the maximum likelihood estimator of the Tobit model.

*Corresponding author, Faculty of Economics, Gakushuin University, Tokyo, Japan: Email address kentaro.akashi@gakushuin.ac.jp (K.Akashi)
†Graduate School of Economics, Hitotsubashi University, Tokyo, Japan.

The motivations of this study are as follows. First, the transformation from the maximum likelihood estimators of the reduced form to the structural parameters of interest is complex in this model. Therefore, we consider the full information maximum likelihood (FIML) estimator that directly estimates the structural parameters. Second, an initial consistent estimator is useful if multiple maximization points exist. Although Blundell and Smith (1994) proposed the conditional maximum likelihood estimator that can simply estimate the structural parameters, its second structural parameters of (1.2) are normalized, and an exclusion restriction for the first structural equation of (1.1) is used. Thus, an initial consistent estimator for our FIML estimator seems to have been unavailable. Although we present an initial estimator, the procedure is a three-stage estimation. Third, if the uniqueness of the maximum likelihood estimator is guaranteed, the efficient estimator can be obtained directly without the initial estimation procedure. Additionally, structural analysis can be performed under less assumptions if the first structural equation does not require a zero constraint. Thus, our approach makes the Heckman model easier for practitioners to use.

The remainder of the paper is organized as follows. Section 2 explains the log-likelihood function and reparametrization method. Section 3 presents the conclusions.

## 2 The log-likelihood function and result

Following Heckman (1978) and Blundell and Smith (1994) for deriving an estimator, we assume that $\beta_1\beta_2 + \beta_3 = 0$, which is called the principal assumption or coherency condition. This condition contributes to the uniqueness of solutions of the nonlinear simultaneous equations model. The substitution of (1.1) into (1.2) yields

$$y_{2i} = \mathbb{1}\{(\beta_1\beta_2 + \beta_3)y_{2i} + y_{2i}^* \geq 0\} \ , \tag{2.3}$$

where $y_{2i}^* = \boldsymbol{\pi}_2'\mathbf{z}_i + v_{2i}$, $\boldsymbol{\pi}_2'\mathbf{z}_i = (\beta_2\boldsymbol{\gamma}_1' + \boldsymbol{\gamma}_2')\mathbf{z}_i$, and $v_{2i} = \beta_2 u_{1i} + u_{2i}$. Then, the reduced form $y_{2i} = \mathbb{1}\{y_{2i}^* \geq 0\}$ is uniquely determined under the coherency condition. We can also construct a maximum likelihood estimator.

Heckman (1978) proposed the maximum likelihood estimators of reduced and structural forms. This study investigates the FIML estimator that directly estimates the structural parameters of the first and second structural equations. The likelihood becomes a function of the parameters $\boldsymbol{\theta}$ under the coherency condition and an exclusion restriction, where $\boldsymbol{\theta} = (\beta_1,\ \beta_2,\ \boldsymbol{\gamma}_1',\ \boldsymbol{\gamma}_{21}',\ \sigma,\ \sigma_{12})'$ and $\boldsymbol{\gamma}_{21} = (\gamma_{21},\ \gamma_{22},\ \cdots,\ \gamma_{2(K-1)})'$. For $i = 1,\ \cdots,\ n$, the contribution to the likelihood function by observation $\#i$ is $\ell_i(\boldsymbol{\theta}) = \ell_{1i}^{y_{2i}}\ell_{0i}^{1-y_{2i}}$, where

$$\ell_{1i} = \int_0^\infty f(y_{1i},\ y_{2i}^*)dy_{2i}^* \ , \qquad \ell_{0i} \ = \ \int_{-\infty}^0 f(y_{1i},\ y_{2i}^*)dy_{2i}^* \ , \tag{2.4}$$

and $f(y_{1i},\ y_{2i}^*)$ stands for the joint-density function.

**Assumption** *(i)* $\mathcal{E}[\mathbf{z}_{\phi_i}\mathbf{z}_{\phi_i}']$ *is nonsingular, where* $\mathbf{z}_{\phi_i} = (\mathcal{E}[y_{2i}|\mathbf{z}_i],\ \mathbf{z}_i')'$. *(ii) For* $k =$

*$K$, $\gamma_{2k} = 0$ and $\gamma_{1k} \neq 0$ . (iii) $\mathcal{E}[z_{ik}^2]$ is bounded. (iv) The parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^{2K+3}$ is compact. (v) $\{\mathbf{z}_i,\ u_{1i},\ u_{2i}\}_{i=1}^n$ are independently and identically distributed.*

The first assumption indicates the absence of multicollinearity. The second one indicates the exclusion restriction to identify the structural parameters. It means that the last variable, $z_{iK}$, of the second structural equation is the excluded variable without loss of generality. Thus, we obtain the relation $\boldsymbol{\gamma}_2'\mathbf{z}_i = (\boldsymbol{\gamma}_{21}',\ 0)\mathbf{z}_i = \boldsymbol{\gamma}_{21}'\mathbf{z}_{2i}$. The third and fourth assumptions are used for the existence of the expectation of Hessian. Theorem derives the results for the contribution to the likelihood function by observation $\#i$, so that the fifth assumption is not used in the proof. However, the assumption is necessary for our statements related to the consistency or limit of the log-likelihood function $l_n(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^n l_i(\boldsymbol{\theta})$, where $l_i(\boldsymbol{\theta})$ is defined below. This is because these asymptotic results are obtained by $l_n(\boldsymbol{\theta}) \xrightarrow{p} \mathcal{E}[\log \ell_i(\boldsymbol{\theta})]$ as $n \to \infty$ under the fifth assumption.

As derived in the appendix, the contribution to the log-likelihood function by observation $\#i$ is represented by

$$l_i(\boldsymbol{\theta}) = \log\left(2\pi\sigma^2\right)^{-\frac{1}{2}} - \frac{u_{1i}^2}{2\sigma^2} + y_{2i}\log\Phi\left(\frac{-\beta_1\beta_2 + u_i}{\sigma_3}\right) + (1 - y_{2i})\log\Phi\left(\frac{-u_i}{\sigma_3}\right)\ , \quad (2.5)$$

$$u_{1i} = y_{1i} - \beta_1 y_{2i} - \boldsymbol{\gamma}_1'\mathbf{z}_i\ , \quad u_i = \beta_2 y_{i1} + \boldsymbol{\gamma}_{21}'\mathbf{z}_{2i} + \frac{\sigma_{12}}{\sigma^2}u_{1i}\ , \text{ and } \quad \sigma_3^2 = 1 - \frac{\sigma_{12}^2}{\sigma^2}\ , \quad (2.6)$$

where $\Phi$ is the cumulative normal distribution function, $\sigma_3^2 = \mathcal{E}[u_{3i}^2]$, and $u_{3i} = (\sigma_{12}/\sigma^2)u_{1i} - u_{2i}$. The FIML estimator is obtained by maximizing the log-likelihood function $l_n(\boldsymbol{\theta})$. The contribution of the log-likelihood function (2.5) becomes highly nonlinear in the parameters. Thus, Heckman (1978) suggested constructing a second-round estimator. If multiple maximization points exist, the second-round estimator is constructed using a consistent estimator for the initial value to obtain consistency and efficiency (cf. Amemiya, 1985).

Olsen (1978) proposed parameter transformation such that the log-likelihood function of the Tobit model becomes quasilinear in parameters. For instance, $(y_{i1} - \boldsymbol{\gamma}_1'\mathbf{z}_i)/\sigma$ becomes $\omega y_{i1} - \boldsymbol{\gamma}_{1\omega}'\mathbf{z}_i$ under $\omega = 1/\sigma$ and $\boldsymbol{\gamma}_{1\omega} = \boldsymbol{\gamma}_1/\sigma$; thus, it is linear in the transformed parameters. He established the global concavity of the log-likelihood function for the transformed parameters. Meanwhile, the likelihood function of Heckman's model is partially similar to that of the Type 2 Tobit model, which also has a correlation parameter between equations, such as $\sigma_{12}$. These log-likelihood functions are not globally concave using a parameter transformation. In Heckman's model, the log-likelihood function has the term $\pi_\omega(\omega y_{i1} - \boldsymbol{\gamma}_{1\omega}'\mathbf{z}_i)$ in $\Phi$ after some parameter transformation, where the parameter $\pi_\omega$ relates to $\sigma_{12}$. It means that the products of parameters remain, such as $\pi_\omega\boldsymbol{\gamma}_{1\omega}$; thus, it is not linear in the parameters. However, Olsen (1982) and Zuehlke (2021) stated that the Type 2 Tobit model is globally concave conditional on a correlation parameter. It is also true for Heckman's model, given $\pi_\omega$. To show that $\pi_\omega$ can be given as the true value, we introduce a higher-order derivative for an identity of the density function in the proof by contradiction. This differential calculus complements the reparametrization approach.

The proof of the theorem is given in the appendix.

**Theorem** *Let Assumptions (i)-(iv) hold. Then,*
*(i)* $\boldsymbol{\theta}$ *is identified without exclusion restrictions for the first structural equation;*
*(ii)* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *implies* $\log \ell_i(\boldsymbol{\theta}) \neq \log \ell_i(\boldsymbol{\theta}_0)$*; thus,* $\mathcal{E}[\log \ell_i(\boldsymbol{\theta})]$ *has a unique maximum at the true value* $\boldsymbol{\theta}_0$*.*

The advantage of the first result is that we do not have to use the instrumental variables for the first structural equation. The second result describes the identification of the density function. Thus, the limit of the log-likelihood function attains a unique global maximum at $\boldsymbol{\theta}_0$. The advantage of this result is that the FIML estimator is asymptotically consistent without an initial consistent estimator.

Blundell and Smith (1994) discussed the simultaneous probit, simultaneous Tobit, and simultaneous generalized selectivity models and proposed the conditional maximum likelihood estimator for the structural parameters.

First, their simultaneous probit model is identical to our Heckman's model; therefore, we can apply the Theorem. Their consistent two-stage estimator of the simultaneous probit model is based on the following:

$$\begin{aligned} \tilde{\Pr}(y_{2i} = 1 \mid \mathbf{z}_i,\ \tilde{u}_{1i}) &= \Phi\left(\frac{\beta_2}{\sigma_3}\tilde{y}_{12i} + \frac{\boldsymbol{\gamma}_{21}'}{\sigma_3}\mathbf{z}_{2i} + \frac{\sigma_{12}}{\sigma_3\sigma^2}\tilde{u}_{1i}\right) && (2.7)\\ &\simeq \Phi\left(\frac{1}{\sigma_{3\hat{\sigma}}}\left\{\beta_2\hat{y}_{12i} + \boldsymbol{\gamma}_{21}'\mathbf{z}_{2i} + \frac{\sigma_{12}}{\hat{\sigma}^2}\hat{u}_{1i}\right\}\right)\ , && (2.8) \end{aligned}$$

where $\tilde{y}_{12i} = y_{1i} - \tilde{\beta}_1 y_{2i}$ and $(\tilde{\beta}_1,\ \tilde{u}_{1i})$ is the estimate of the instrumental variable estimator and its residual from the first structural equation, respectively. Thus, the normalized parameters from $\beta_2/\sigma_3$ to $\sigma_{12}/(\sigma_3\sigma^2)$ of (2.7) can be obtained using the probit estimation. To obtain an initial estimator for efficient estimation without an instrumental variable, we slightly modify their estimator (2.8), where $\hat{y}_{12i} = y_{1i} - \hat{\beta}_1 y_{2i}$, $\sigma_{3\hat{\sigma}} = (1 - \sigma_{12}^2/\hat{\sigma}^2)^{1/2}$, and $(\hat{\beta}_1,\ \hat{\sigma},\ \hat{u}_{1i})$ is the estimates of a nonlinear two-stage least squares estimator and its residual. The nonlinear two-stage least squares estimator and consistency including $\sigma_{12}$ of the modified estimator are discussed in the appendix. Although we can construct the initial consistent estimator for the FIML estimator, the procedure requires three-stage estimation and nonlinear probit estimation of (2.8).

Second, the simultaneous Tobit model under the condition $\beta_1\beta_2 + \beta_3 = 0$ is the model with the second structural equation of (2.3) replaced by the following Tobit model

$$y_{2i} = \mathbb{1}\{y_{2i}^* \geq 0\} y_{2i}^*\ . \tag{2.9}$$

This simultaneous equations model becomes a simple case of that proposed by Amemiya (1974). Our approach can derive the same results of Theorem for the simultaneous Tobit model, as shown in the appendix.

Third, the simultaneous generalized selectivity model is obtained by replacing the second structural equation of (2.3) with the following Type 2 Tobit model

$$y_{2i} = \mathbb{1}\{y^*_{3i} \geq 0\} y^*_{2i} \ . \tag{2.10}$$

This simultaneous equations model constructs an extended model of Heckman (1979). The proof method also clarifies the identification of the density function of the Type 2 Tobit model. The simultaneous generalized selectivity model has three endogenous variables, increasing correlation parameters. The proof method can be generalized to the three simultaneous equations model using a cross-partial derivative to obtain the same results of Theorem, as shown in the appendix.

The above results are summarized as a corollary of Theorem.

**Corollary** *Let Assumptions (i)-(iv) hold. Then, for the simultaneous Tobit and generalized selectivity models,*
*(i) The structural parameters are identified without exclusion restrictions for the first structural equation;*
*(ii) The log of the density function is identified; thus, its expectation has a unique maximum at the true value.*

# 3 Conclusions

This note demonstrated that the uniqueness of the objective function of the FIML estimator for a Heckman's model using the reparametrization method. Therefore, we can obtain the consistent and efficient estimator by maximizing the log-likelihood function directly. Furthermore, we proposed a complemented reparametrization method using some differential calculus. The proposed method can also be used for similar models to show the uniqueness of the maximum likelihood estimator.

### Appendix

**Proof of Theorem :** First, we present the contribution to the likelihood function by observation $i$. Under $\beta_3 = -\beta_1\beta_2$ ,

$$\begin{aligned}
\ell_{1i} &= \int_{\beta_1\beta_2-\beta_2 y_{1i}-\boldsymbol{\gamma}'_{21}\mathbf{z}_{2i}}^{\infty} g(u_{1i},\ u_{2i}) du_{2i} \\
&= \frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right) \int_{-\infty}^{-\beta_1\beta_2+u_i} \phi\left(\frac{u_{3i}}{\sigma_3}\right) \frac{1}{\sigma_3} du_{3i} \\
&= \frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right) p(u_{1i}) \ , \ \text{and} \qquad \text{(A.1)} \\
\ell_{0i} &= \frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right) (1 - p(u_{1i})) \ , \qquad \text{(A.2)}
\end{aligned}$$

where $g$ and $\phi$, are the joint and normal density function, respectively. The second equality arises from the variable transformation to $u_{3i} = (\sigma_{12}/\sigma^2)u_{1i} - u_{2i}$, and $y_{1i} = \beta_1 y_{2i} + \boldsymbol{\gamma}_1'\mathbf{z}_i + u_{1i}$ is substituted at the third equality, with $p(u_{1i})$ defined by (A.5). Given $\mathbf{z}_i$, for any $\boldsymbol{\theta}$ the following holds,

$$\sum_{y_{2i}=0,1} \int_{-\infty}^{\infty} \ell_i(\boldsymbol{\theta}) du_{1i} = 1 \ . \tag{A.3}$$

**Proof of (i)** The outline is as follows. (i-1) shows the identification of the structural parameters of the first structural equation without an exclusion restriction using the nonlinearity of $\mathcal{E}[y_{2i}|\mathbf{z}_i] = \Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i)$ in $\mathbf{z}_i$. (i-2) clarifies that $\boldsymbol{\theta}$ has a one-to-one correspondence with a transformed parameter $\boldsymbol{\theta}_1$, which is identifiable.

**(i-1)** The first structural equation is given by

$$y_{1i} \;\; = \;\; \beta_1\Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i) + \boldsymbol{\gamma}_1'\mathbf{z}_{1i} + (u_{1i} + \beta_2 y_{2i} - \beta_2\Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i)) \ , \tag{A.4}$$

where $\boldsymbol{\pi}_{22} = \boldsymbol{\pi}_2/\omega_2, \omega_2^2 = \mathcal{E}[v_{2i}^2]$ and $\omega_2$ is positive by $\sigma_{12}^2 < \sigma^2$. $\boldsymbol{\pi}_{22}$ is identified by the expectation of the log of the probability function of the probit model, $\Pr(y_{2i} = 1) = \Pr(y_{2i}^*/\omega_2 \geq 0)$. This is because of its global concavity and Assumption $(i)$. Then, $(\beta_1,\ \boldsymbol{\gamma}_1')$ is identified from a nonlinear two-stage least squares estimator on the population, i.e., $(\beta_1, \boldsymbol{\gamma}_1')' = \mathcal{E}[\mathbf{z}_{\phi i}\mathbf{z}_{\phi i}']^{-1}\mathcal{E}[\mathbf{z}_{\phi i}y_{1i}]$ where $\mathcal{E}[\mathbf{z}_{\phi i}\mathbf{z}_{\phi i}']$ is nonsingular from Assumption $(i)$. Then, $u_{1i}$ is identified as the residual, and $\sigma$ is determined by $\mathcal{E}[u_{1i}^2]$. Therefore, we do not use an exclusion restriction for the first structural equation in the above derivation.

**(i-2)** Consider the three-stage probit maximum likelihood estimator for identifying the parameters of the second equation. Given $u_{1i}$ by the residual, the conditional probability $\mathcal{E}[y_{2i}|\mathbf{z}_i,\ u_{1i}]$ is given by

$$p(u_{1i}) = \Phi\left(\frac{1}{\sigma_3}\left\{\boldsymbol{\pi}_2'\mathbf{z}_i + \beta_2 u_{1i} + \frac{\sigma_{12}}{\sigma^2}u_{1i}\right\}\right) \ , \tag{A.5}$$

since $y_{2i} = \mathbb{1}\{\boldsymbol{\pi}_2'\mathbf{z}_i + \beta_2 u_{1i} + (\sigma_{12}/\sigma^2)u_{1i} \geq u_{3i}\}$ and $u_{3i}$ is independent of $u_{1i}$ and $\mathbf{z}_i$. The probit model identifies the coefficients again by the nonsingularity of $\mathcal{E}[(\mathbf{z}_i',\ u_{1i})'(\mathbf{z}_i',\ u_{1i})]$. The reduced form coefficients $(\boldsymbol{\pi}_3',\ \pi_\rho)$ are identified and have the following relation: $(\boldsymbol{\pi}_3',\ \pi_\rho) = (\boldsymbol{\pi}_2'/\sigma_3,\ (\beta_2 + \sigma_{12}/\sigma^2)/\sigma_3)$ .

Using $(\beta_1,\ \boldsymbol{\gamma}_1',\ \sigma)$ and $(\boldsymbol{\pi}_3',\ \pi_\rho)$, we show that $(\beta_2,\ \boldsymbol{\gamma}_{21}',\ \sigma_{12})$ is identified. From Assumption $(ii)$, $\beta_2/\sigma_3$ is identified by $\pi_{3K}/\gamma_{1K}$. Moreover, $\boldsymbol{\gamma}_{21}'/\sigma_3 = \boldsymbol{\pi}_3' - (\beta_2/\sigma_3)\boldsymbol{\gamma}_1'$, except for $k = K$. For $\sigma_{12}$ , $\sigma_{12}/\sigma_3 = \sigma^2(\pi_\rho - \beta_2/\sigma_3) = \pi_{12}$ , i.e., $\sigma_{12} = \pi_{12}(1 - \sigma_{12}^2/\sigma^2)^{1/2}$. By solving it for $\sigma_{12}$ , we obtain that

$$\sigma_{12} = \pm\sqrt{\frac{\pi_{12}^2}{1 + \pi_{12}^2/\sigma^2}} \ , \tag{A.6}$$

where the sign is identified by $\mathrm{sgn}(\pi_{12})$. Then, $\sigma_3 = (1 - \sigma_{12}^2/\sigma^2)^{1/2}$ is determined. Therefore, $\beta_2$ is identified by $(\beta_2/\sigma_3)\sigma_3$, which is also true for $\boldsymbol{\gamma}_{21}$.

Thus, $\boldsymbol{\theta}$ is identified; it has a one-to-one correspondence with $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(\boldsymbol{\theta}) = (\beta_1,\ \boldsymbol{\gamma}_1',\ \sigma,\ \boldsymbol{\pi}_3',\ \pi_\rho)'$ .
■

**Proof of (ii)** The outline is as follows. Our strategy is that if the density function is not identified, then the result contradicts Jensen's inequality (ii-1, ii-4). Thus, we consider the global concavity of $l_i(\boldsymbol{\theta}_1)$, which is generally concave if the arguments of $\phi$ and $\Phi$ become linear in the parameters (e.g., Olsen, 1978; Pratt, 1981). However, the transformation by $\boldsymbol{\theta}_1$ is insufficient for linearization. Moreover, (ii-2) is the essential part and derives that using partial derivatives, $1/\sigma$ and $\pi_\rho$ can be fixed at the true values. Thus, we obtain linearization for the remaining parameters. Compared with Jensen's inequality, (ii-3) checks the concavity in the sense that $\mathcal{E}[l_i(\boldsymbol{\theta}_1)]$ is globally concave given $1/\sigma$ and $\pi_\omega = \sigma\pi_\rho$. Although linearization is sufficient by fixing only $\pi_\omega$ with the transformation $\omega = 1/\sigma$, we fix $\sigma$ because the third-order partial derivative can only identify up to $\pi_\rho$.

**(ii-1)** We use the transformed parameter $\boldsymbol{\theta}_1$ defined above, and put $\boldsymbol{\theta}_{01}$ to distinguish the true value. Suppose that some $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{01}$ exists such that $l_i(\boldsymbol{\theta}_1) = l_i(\boldsymbol{\theta}_{01})$ for any given $(y_{1i},\ y_{2i},\ \mathbf{z}_i')$, i.e., these values are the same as the density function. Through the representation of (A.1) and (A.2), we obtain the identity

$$\begin{aligned} l_i(\boldsymbol{\theta}_1) &= \log \frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right)\Phi\left(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i}\right)^{y_{2i}}\left(1-\Phi(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\right)^{1-y_{2i}} \\ &= \log \frac{1}{\sigma_0}\phi\left(\frac{u_{01i}}{\sigma_0}\right)\Phi\left(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0} u_{01i}\right)^{y_{2i}}\left(1-\Phi(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0} u_{01i})\right)^{1-y_{2i}} \end{aligned} \quad (A.7)$$

where $u_{1i} = y_{1i} - \beta_2 y_{2i} - \boldsymbol{\gamma}_1'\mathbf{z}_i$, and the representative subscript 0 is evaluated at the true value $\boldsymbol{\theta}_{01}$.

**(ii-2)** Partial differentiation is possible with respect to $y_{1i}$ by changing only the error term $u_{01i}$. The third-order partial derivative for $y_{1i}$ is also the identity given by

$$\begin{aligned} & y_{i2}\,\ddot{r}(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\pi_\rho^3 - (1-y_{2i})\,\ddot{h}(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\pi_\rho^3 \\ = \; & y_{i2}\,\ddot{r}(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i})\pi_{\rho 0}^3 - (1-y_{2i})\,\ddot{h}(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i})\pi_{\rho 0}^3\ , \end{aligned} \quad (A.8)$$

where $r(x) = \phi(x)/\Phi(x)$, $h(x) = \phi(x)/(1-\Phi(x))$, $\ddot{r} = \partial^2 r/\partial x^2$, and $\ddot{h}(x) = \partial^2 h/\partial x^2$ is the second derivative of the hazard rate, which is $\ddot{h}(x) = h(x)[(2h(x)-x)(h(x)-x)-1]$ .

Suppose $\pi_\rho \neq \pi_{\rho 0}$. Then, take $y_{2i} = 0$ and $y_{1i}$ as $y_{1i} = 1/(\pi_\rho - \pi_{\rho 0})[(\boldsymbol{\pi}_{03} - \boldsymbol{\pi}_3)'\mathbf{z}_i + (\pi_\rho\gamma_1 - \pi_{\rho 0}\gamma_{01})'\mathbf{z}_i]$, respectively. It is the same that the error term $u_{01i}$ is evaluated at $y_{1i} - \gamma_{01}'\mathbf{z}_i$ and $u_{03i}$ satisfies the inequality $\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i} < u_{03i}$ at (A.5), where $\mathbf{z}_i$ is not restricted. Then, $\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i} = \boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i}$ , so that $\ddot{h}(x)$ is equal on both sides:

$$\ddot{h}(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\pi_\rho^3 = \ddot{h}(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\pi_{\rho 0}^3\ . \quad (A.9)$$

The convexity of $h(x)$ is known, i.e., $\ddot{h}(x) > 0$ for any $x$, so that we obtain $\pi_\rho^3 = \pi_{\rho 0}^3$ dividing by it. However, it contradicts the assumption that $\pi_\rho \neq \pi_{\rho 0}$ . Therefore, $\pi_\rho = \pi_{\rho 0}$ is a necessary condition for the identity of (A.7).

Similarly, the second partial derivative of $y_{1i}$ under $y_{2i} = 0$ yields the following:

$$-\frac{1}{\sigma^2} + \dot{h}(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_{\rho 0}u_{1i})\pi_{\rho 0}^2 = -\frac{1}{\sigma_0^2} + \dot{h}(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i})\pi_{\rho 0}^2 \ , \tag{A.10}$$

where $\dot{h}(x) = \partial h(x)/\partial x = h(x)(h(x) - x)$. We can take $\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_{\rho 0}u_{1i} = \boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i}$ under $\pi_{\rho 0} \neq 0$, then, $\sigma^2 = \sigma_0^2$. If $\pi_{\rho 0} = 0$, then $\sigma^2 = \sigma_0^2$ also holds. Therefore, the arguments of $\phi$ and $\Phi$ of $l_i(\boldsymbol{\theta}_1)$ become linear in the remaining parameters given $(\pi_\rho,\ \sigma) = (\pi_{\rho 0},\ \sigma_0)$.

**(ii-3)** Let $\boldsymbol{\theta}_2 = (\omega,\ \beta_{1\omega},\ \boldsymbol{\gamma}_{1\omega}',\ \boldsymbol{\pi}_3',\ \pi_\omega)'$ be the parameter transformation such that $\omega = 1/\sigma$, $\beta_{1\omega} = \beta_1/\sigma$, $\boldsymbol{\gamma}_{1\omega} = \boldsymbol{\gamma}_1/\sigma$, and $\pi_\omega = \sigma\pi_\rho$. Then, there exists a relation $\pi_\rho u_{1i} = \pi_\omega(\omega y_{1i} - \beta_{1\omega}y_{2i} - \boldsymbol{\gamma}_{1\omega}'\mathbf{z}_i)$. We show the global concavity of $\mathcal{E}[l_i(\boldsymbol{\theta}_2)] = \mathcal{E}[l_i(\boldsymbol{\psi}_2)]$ with respect to the vector $\boldsymbol{\psi}_2 = (\beta_{1\omega},\ \boldsymbol{\gamma}_{1\omega}',\ \boldsymbol{\pi}_3')'$, given $\omega = \omega_0 = 1/\sigma_0$ and $\pi_\omega = \sigma_0\pi_{\rho 0}$ . For any nonzero vector $\mathbf{t} = (\mathbf{t}_1',\ \mathbf{t}_2')'$, the quadratic form of the Hessian is given by

$$\mathcal{E}\left[\mathbf{t}'\frac{\partial^2 l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2'}\mathbf{t}\right] = \mathcal{E}\left[-(\mathbf{t}_1'\mathbf{z}_{y_i})^2 + \dot{g}(z_{\pi i})(-\sigma_0\pi_{\rho 0}\mathbf{t}_1'\mathbf{z}_{y_i} + \mathbf{t}_2'\mathbf{z}_i)^2\right] \ , \tag{A.11}$$

where $\mathbf{z}_{yi} = (y_{2i},\ \mathbf{z}_i')'$, $z_{\pi i} = \boldsymbol{\pi}_3'\mathbf{z}_i + \pi_{\omega 0}u_{1i}$, $\dot{r}(x) = \partial r/\partial x$, and $\dot{g}(z_{\pi i}) = y_{2i}\dot{r}(z_{\pi i}) - (1 - y_{2i})\dot{h}(z_{\pi i})$ .

For $\mathbf{t}_1 \neq \mathbf{0}$, if $\mathcal{E}[(\mathbf{t}_1'\mathbf{z}_{y_i})^2] > 0$, then (A.11) is negative since $\dot{g}(z_{\pi i}) < 0$, where $\dot{r}(x) < 0$ and $\dot{h}(x) > 0$ are known for any $x$. Suppose $\mathcal{E}[(\mathbf{t}_1'\mathbf{z}_{yi})^2] = 0$ for some $\mathbf{t}_1 \neq 0$. Then, $\mathbf{t}_1'\mathbf{z}_{yi}$ is degenerated, $\mathbf{t}_1'\mathbf{z}_{yi} = 0$. Take the conditional expectation, $t_{11}\Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i) + \mathbf{t}_{12}'\mathbf{z}_i = 0$ , where $(t_{11},\ \mathbf{t}_{12}')' = \mathbf{t}_1$. Then, we have the relation $\mathbf{t}_1'\mathcal{E}[\mathbf{z}_{\phi i}\mathbf{z}_{\phi i}'] = \mathbf{0}'$ or $\mathbf{t}_1 = \mathbf{0}$ by Assumption $(i)$, which contradicts $\mathbf{t}_1 \neq \mathbf{0}$. Therefore, (A.11) is negative.

For $\mathbf{t}_1 = \mathbf{0}$, it holds that $\mathbf{t}_2 \neq \mathbf{0}$ by the definition of $\mathbf{t}$. Then, (A.11) equals $\mathcal{E}[\dot{g}(z_{\pi i})(\mathbf{t}_2'\mathbf{z}_i)^2]$. Suppose $\mathbf{t}_2 \neq \mathbf{0}$ exists such that $\mathcal{E}[y_{2i}\dot{h}(z_{\pi i})(\mathbf{z}_i'\mathbf{t}_2)^2] = 0$. Then, $y_{2i}\dot{h}(z_{\pi i})(\mathbf{t}_2'\mathbf{z}_i)^2 = 0$, since it is a nonnegative random variable. Moreover, $y_{2i}(\mathbf{t}_2'\mathbf{z}_i)^2 = 0$ by $\dot{h}(z_{\pi i}) > 0$ and $(\mathbf{t}_2'\mathbf{z}_i)^2 = 0$ by $\mathcal{E}[y_{2i}|\mathbf{z}_i] = \Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i) > 0$. Since $\mathcal{E}[\mathbf{z}_i\mathbf{z}_i']$ is nonsingular by Assumption $(i)$, $\mathbf{t}_2 = \mathbf{0}$, which is a contradiction. Thus, using similar arguments, we obtain that $\mathcal{E}[y_{2i}\dot{r}(z_{\pi i})(\mathbf{t}_2'\mathbf{z}_i)^2] < 0$ and $\mathcal{E}[(1 - y_{2i})\dot{h}(z_{\pi i})(\mathbf{t}_2'\mathbf{z}_i)^2] > 0$. Therefore, $\mathcal{E}[\dot{g}(z_{\pi i})(\mathbf{t}_2'\mathbf{z}_i)^2] < 0$ for any $\mathbf{t}_2 \neq \mathbf{0}$, meaning that $\mathcal{E}[\partial^2 l_i(\boldsymbol{\psi}_2)/\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2']$ is negative definite and indeed exists, as shown below.

We show that

$$\frac{\partial^2\mathcal{E}[l_i(\boldsymbol{\psi}_2)]}{\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2'} = \mathcal{E}\left[\frac{\partial^2 l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2'}\right] \ . \tag{A.12}$$

To verify this condition, we consider the interchanges of the derivative and expectation:

$$\frac{\partial\mathcal{E}[l_i(\boldsymbol{\psi}_2)]}{\partial\boldsymbol{\psi}_2'} = \mathcal{E}\left[\frac{\partial l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2'}\right] \ , \text{ and } \quad \frac{\partial}{\partial\boldsymbol{\psi}_2}\mathcal{E}\left[\frac{\partial l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2'}\right] = \mathcal{E}\left[\frac{\partial^2 l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2'}\right] \ . \tag{A.13}$$

Let $\mathbf{e}_k' = (0,\ \cdots,\ 1,\ \cdots,0)$ whose $k$-th element is only unity. For the second equality,

$$\left|\mathbf{e}_k'\frac{\partial^2 l_i(\boldsymbol{\psi}_2)}{\partial\boldsymbol{\psi}_2\partial\boldsymbol{\psi}_2'}\mathbf{e}_k\right| \leq (\mathbf{e}_{k1}'\mathbf{z}_{y_i})^2 + |\dot{g}(z_{\pi i})|(-\sigma_0\pi_{\rho 0}\mathbf{e}_{k1}'\mathbf{z}_{y_i} + \mathbf{e}_{k2}'\mathbf{z}_i)^2 \ , \tag{A.14}$$

where $\mathbf{e}_k' = (\mathbf{e}_{k1}',\ \mathbf{e}_{k2}')$. Using the relations that $r(x) = h(-x)$ and $\dot{h}(x) = h(x)(h(x)-x)$ ,

$$\begin{aligned} |\dot{g}(z_{\pi i})| &\leq y_{2i}|\dot{h}(-z_{\pi i})| + (1-y_{2i})|\dot{h}(z_{\pi i})| \\ &\leq 1 , \end{aligned} \tag{A.15}$$

where the last inequality is from $0 \leq \dot{h}(x) \leq 1$ since $0 \leq 1 + h(x)x - h(x)^2 \leq 1$ is given by Heckman (1979). From Assumption (*iv*), the parameters are bounded, and there exist positive constants $c_k$ such that

$$\left| \mathbf{e}_k' \frac{\partial^2 l_i(\boldsymbol{\psi}_2)}{\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\psi}_2'} \mathbf{e}_k \right| \leq (c_0 + \sum_{k=1}^{K} c_k |z_{ik}|)^2 + (c_0 + \sum_{k=1}^{K} c_k |z_{ik}|)^2 \quad a.s. . \tag{A.16}$$

The expectation of the right-hand side is finite under Assumption (*iii*). Under similar arguments, the case holds for the off-diagonal elements of (A.13). Therefore, the moment of the Hessian of (A.13) exists and the second equality is valid according to Lebesgue's dominated convergence theorem. Similarly, the first equality of (A.13) can be shown. Thus, we conclude that $\partial^2 \mathcal{E}[l_i(\boldsymbol{\psi}_2)]/\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\psi}_2'$ is negative definite, i.e., $\mathcal{E}[l_i(\boldsymbol{\psi}_2)]$ is globally concave.

**(ii-4)** Now, $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{01}$ implies $\boldsymbol{\psi}_2 \neq \boldsymbol{\psi}_{02}$, since $\pi_\rho$ and $\sigma$ are fixed at $\pi_{\rho 0}$ and $\sigma_0$, respectively. Therefore, for some $\lambda$ such that $0 < \lambda < 1$ , we obtain

$$\begin{aligned} \mathcal{E}[l_i(\boldsymbol{\theta}_{02})] &= \lambda \mathcal{E}[l_i(\boldsymbol{\psi}_2)] + (1-\lambda)\mathcal{E}[l_i(\boldsymbol{\psi}_{02})] \\ &< \mathcal{E}[l_i(\lambda \boldsymbol{\psi}_2 + (1-\lambda)\boldsymbol{\psi}_{02})] = \mathcal{E}[l_i(\bar{\boldsymbol{\theta}}_2)] , \end{aligned} \tag{A.17}$$

where $\bar{\boldsymbol{\theta}}_2 = (\omega_0,\ (\lambda \boldsymbol{\psi}_2 + (1-\lambda)\boldsymbol{\psi}_{02})',\ \pi_{\omega 0})'$. The first equality is from the identity. Meanwhile, we have $\mathcal{E}[l_i(\bar{\boldsymbol{\theta}}_2)] \leq \mathcal{E}[l_i(\boldsymbol{\theta}_{02})]$ by (A.3) and Jensen's inequality, which is a contradiction.

Therefore, the first assumption $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{01}$ is false, and $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{01}$ is necessary for the identity. Its contraposition is that $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{01}$ implies $l_i(\boldsymbol{\theta}_1) \neq l_i(\boldsymbol{\theta}_{01})$. Then, Jensen's inequality strictly holds, $\mathcal{E}[l_i(\boldsymbol{\theta}_1)] < \mathcal{E}[l_i(\boldsymbol{\theta}_{01})]$ . Under Assumption (*ii*), $\boldsymbol{\theta}$ has a one-to-one correspondence with $\boldsymbol{\theta}_1$. Therefore, the true value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$ is a unique maximum point. Thus, we obtain the desired result. ■

Using the strategy of the above proof under Assumption, we consider the consistency of the modified initial estimator and clarify the essential part of the proof to identify the simultaneous Tobit and generalized selectivity models.

**Initial estimation :** This part considers the consistency of the modified initial estimator of (2.8). We can assume that the true parameters of the first structural equation are given. This is because these can be consistently estimated in advance from the nonlinear two-stage least squares estimator on the sample space, as discussed in Proof of (i). Then, the representation of (2.8) becomes $p(u_{1i})$ of (A.5), since $y_{12i} = y_{1i} - \beta_1 y_{2i} = \boldsymbol{\gamma}_1' \mathbf{z}_i + u_{1i}$.

Hence, the conditional probability given $u_{01i}$ is given as follows:

$$\Phi\left(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{01i}\right)^{y_{2i}}\left(1 - \Phi(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{01i})\right)^{1-y_{2i}}, \tag{A.18}$$

where $\sigma = \sigma_0$ is given in $\boldsymbol{\pi}_3$ and $\pi_\rho$. Since $u_{01i}$ is evaluated at the true value, the remaining parameters $(\beta_2,\ \boldsymbol{\gamma}_{21}',\ \sigma_3,\ \sigma_{12})$ of (2.8) are identified by the same arguments of Proof of (i). Then, the consistency of the modified estimator is shown. ■

**Simultaneous Tobit :** In this part, we identify the structural parameters and density function for the simultaneous Tobit model.

**(i)** $\mathcal{E}[y_{2i}|\mathbf{z}_i] = [\boldsymbol{\pi}_2'\mathbf{z}_i + \omega_2\phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i)/\Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i)]\Phi(\boldsymbol{\pi}_{22}'\mathbf{z}_i)$ is nonlinear in $\mathbf{z}_i$, and its parameters is identified by the expectation of the log of the density function of standard Tobit model according to the result of Olsen (1978). Hence, similar to Proof of (i), the structural parameters of the first structural equation can be identified by a nonlinear two-stage least squares estimator on the population without an exclusion restriction. We redefine $\sigma_3^2 = \sigma_2^2 - \sigma_{12}^2/\sigma^2$, where $\sigma_2^2 = \mathcal{E}[u_{2i}^2]$ is the structural parameter to be added instead of $\sigma_2^2 = 1$. The standard Tobit model given $u_{1i}$ of (A.5) can identify $\sigma_3$, and $\sigma_{12} = \pi_{12}\sigma_3$ holds since the Tobit model contains the probability function of the probit model. The additional parameter $\sigma_2^2$ is determined by $\sigma_3^2 + \sigma_{12}^2/\sigma^2$. Therefore, $\boldsymbol{\theta}_\tau = (\boldsymbol{\theta}',\ \sigma_2)'$ is identified. ■

**(ii)** Using the representation $y_{2i} = \boldsymbol{\pi}_2'\mathbf{z}_i + (\beta_2 + \sigma_{12}/\sigma^2)u_{1i} - u_{3i}$, the Jacobian of the variable transformation from $(y_{1i}, y_{2i})$ to $(u_{1i}, u_{3i})$ becomes unity. Then, for the density function of the simultaneous Tobit model, we suppose the following identity:

$$\begin{aligned} l_i(\boldsymbol{\theta}_{1\tau}) &= \log\frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right)\left(\omega_3\phi\left(\omega_3 y_{2i} - \boldsymbol{\pi}_3'\mathbf{z}_i - \pi_\rho u_{1i}\right)\right)^{w_{2i}}\left(1 - \Phi(\boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i})\right)^{1-w_{2i}} \\ &= \log\frac{1}{\sigma_0}\phi\left(\frac{u_{01i}}{\sigma_0}\right)\left(\omega_{03}\phi\left(\omega_{03}y_{2i} - \boldsymbol{\pi}_{03}'\mathbf{z}_i - \pi_{\rho 0}u_{01i}\right)\right)^{w_{2i}}\left(1 - \Phi(\boldsymbol{\pi}_{03}'\mathbf{z}_i + \pi_{\rho 0}u_{01i})\right)^{1-w_{2i}}, \end{aligned} \tag{A.19}$$

where $w_{2i} = \mathbb{I}\{y_{2i}^* \geq 0\}$ and $\omega_3 = 1/\sigma_3$. Then, the transformed parameters become $\boldsymbol{\theta}_{1\tau} = (\boldsymbol{\theta}_1',\ \omega_3)'$. Considering the third-order partial derivative of $y_{1i}$ under $w_{2i} = 0$, we obtain the conditions $(\pi_\rho,\ \sigma) = (\pi_{\rho 0},\ \sigma_0)$ from the result of Proof of (ii). The arguments of $\phi$ and $\Phi$ become linear in the parameters given the conditions $(1/\sigma,\ \pi_\rho) = (1/\sigma_0,\ \pi_{\rho 0})$ and transformation $\omega_3 = 1/\sigma_3$. The remaining part of the proof is similar to Proof of (ii). Thus, the identification of $l_i(\boldsymbol{\theta}_{1\tau})$ is obtained. Moreover, $\boldsymbol{\theta}_{1\tau}$ has a one-to-one correspondence with $\boldsymbol{\theta}_\tau$. Then, $\boldsymbol{\theta}_{1\tau} \neq \boldsymbol{\theta}_{1\tau 0}$ implies that $\boldsymbol{\theta}_\tau \neq \boldsymbol{\theta}_{\tau 0}$ and $l_i(\boldsymbol{\theta}_\tau) = l_i(\boldsymbol{\theta}_{1\tau}) \neq l_i(\boldsymbol{\theta}_{1\tau 0}) = l_i(\boldsymbol{\theta}_{\tau 0})$. Therefore, the density function with $\boldsymbol{\theta}_\tau$ of the simultaneous Tobit model is also identified. ■

**Generalized selectivity model :** In this part, we identify parameters and density function for the simultaneous generalized selectivity model. The third reduced form is given by $y_{3i}^* = \boldsymbol{\pi}_{3*}'\mathbf{z}_i + v_{3i}$, where $\mathbf{u}_i = (u_{1i},\ u_{2i},\ v_{3i})'$ follows a normal distribution with zero mean and $\mathcal{E}[v_{3i}^2] = 1$, and $\mathcal{E}[\mathbf{u}_i\mathbf{u}_i']$ is positive definite.

**(i)** The reduced form parameters of $\mathcal{E}[y_{2i}|\mathbf{z}_i] = [\boldsymbol{\pi}_2'\mathbf{z}_i + \omega_{23}\phi(\boldsymbol{\pi}_{3*}'\mathbf{z}_i)/\Phi(\boldsymbol{\pi}_{3*}'\mathbf{z}_i)]\Phi(\boldsymbol{\pi}_{3*}'\mathbf{z}_i)$ where $\omega_{23} = \mathcal{E}[v_{2i}v_{3i}]$ are identified by the density function of the Type 2 Tobit model, as shown below.

**Type 2 Tobit :** Similar to Proof of (ii), we suppose the following identity, where the original parameters of the density function are $\boldsymbol{\theta}_\pi = (\boldsymbol{\pi}_2',\ \omega_2,\ \omega_{23},\ \boldsymbol{\pi}_{3*}')'$:

$$\begin{aligned} l_i(\boldsymbol{\theta}_{1\pi}) &= \log\left(\frac{1}{\omega_2}\phi\left(\frac{v_{2i}}{\omega_2}\right)\Phi\left(\frac{1}{\omega_4}\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\tau v_{2i}\right)\right)^{y_{3i}}\left(1-\Phi(\boldsymbol{\pi}_{3*}'\mathbf{z}_i)\right)^{1-y_{3i}} \\ &= \log\left(\frac{1}{\omega_{02}}\phi\left(\frac{v_{02i}}{\omega_{02}}\right)\Phi\left(\frac{1}{\omega_{04}}\boldsymbol{\pi}_{03*}'\mathbf{z}_i + \pi_{\tau0} v_{02i}\right)\right)^{y_{3i}}\left(1-\Phi(\boldsymbol{\pi}_{03*}'\mathbf{z}_i)\right)^{1-y_{3i}} \end{aligned} \tag{A.20}$$

where $y_{3i} = \mathbb{I}\{y_{3i}^* \geq 0\}$, $\pi_\tau = \omega_{23}/(\omega_2^2\omega_4)$, and $\omega_4 = (1-\omega_{23}^2/\omega_2^2)^{1/2} = (1-\omega_2^2\pi_\tau^2/(1+\omega_2^2\pi_\tau^2))^{1/2}$. Therefore, the transformed parameters $\boldsymbol{\theta}_{1\pi} = (\boldsymbol{\pi}_2',\ \omega_2,\ \pi_\tau,\ \boldsymbol{\pi}_{3*}')'$ do not include $\omega_4$. The representative subscript 0 is evaluated at the true value. The third-order partial derivative of $y_{2i}$ under $y_{3i} = 1$ yields the following:

$$\dddot{r}((1/\omega_4)\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\tau v_{2i})\pi_\tau^3 = \dddot{r}((1/\omega_{04})\boldsymbol{\pi}_{03*}'\mathbf{z}_i + \pi_{\tau0} v_{02i})\pi_{\tau0}^3\ . \tag{A.21}$$

Suppose $\pi_\tau \neq \pi_{\tau0}$ and take $y_{2i} = 1/(\pi_\tau - \pi_{\tau0})[(\boldsymbol{\pi}_{03*}/\omega_{04} - \boldsymbol{\pi}_{3*}/\omega_4)'\mathbf{z}_i + (\pi_\tau\boldsymbol{\pi}_2 - \pi_{\tau0}\boldsymbol{\pi}_{02})'\mathbf{z}_i]$. Then, $\dddot{r}((1/\omega_4)\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\tau v_{2i})\pi_\tau^3 = \dddot{r}((1/\omega_4)\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\tau v_{2i})\pi_{\tau0}^3$, or $\pi_\tau = \pi_{\tau0}$ is the necessary condition, since $\dddot{r}(x) = \ddot{h}(-x) > 0$. From the similar arguments of Proof (ii), $\omega_2 = \omega_{02}$ holds. Therefore, $\omega_4$ can also be fixed at the true value. The arguments of $\phi$ and $\Phi$ become linear in the parameters given $(1/\omega_2,\ 1/\omega_4,\ \pi_\tau) = (1/\omega_{02},\ 1/\omega_{04},\ \pi_{\tau0})$. Thus, the identification of $l_i(\boldsymbol{\theta}_{1\pi})$ is shown. Moreover, $\boldsymbol{\theta}_{1\pi}$ has a one-to-one correspondence with $\boldsymbol{\theta}_\pi$ because $w_{23} = \pm(\omega_2^4\pi_\tau^2/(1+\omega_2^2\pi_\tau^2))^{1/2}$, where the sign is determined by $\mathrm{sgn}(\pi_\tau)$. Therefore, the density function with $\boldsymbol{\theta}_\pi$ of the Type 2 Tobit model is identified. ■

From the above result, a nonlinear two-stage least squares estimator based on $\mathcal{E}[y_{2i}|\mathbf{z}_i]$ can identify the structural parameters of the first structural equation without an exclusion restriction. The structural and reduced form parameters of the simultaneous generalized selectivity model become $\boldsymbol{\theta}_* = (\boldsymbol{\theta}_\tau',\ \sigma_{13},\ \sigma_{23},\ \boldsymbol{\pi}_{3*}')'$, where $\sigma_{13} = \mathcal{E}[u_{1i}v_{3i}]$ and $\sigma_{23} = \mathcal{E}[u_{2i}v_{3i}]$. Meanwhile, the transformed parameters $\boldsymbol{\theta}_{1*} = (\boldsymbol{\theta}_{1\tau}',\ \pi_\sigma,\ \pi_\lambda,\ \boldsymbol{\pi}_{3*}')'$ defined by (A.22) can be identified. Consequently, $\boldsymbol{\theta}_*$ has a one-to-one correspondence with $\boldsymbol{\theta}_{1*}$ because $\sigma_{13} = \pi_\sigma\sigma^2$, $\sigma_{23} = \mathcal{E}[((\sigma_{12}/\sigma^2)u_{1i} - u_{3i})v_{3i}] = \pi_\sigma\sigma_{12} - \pi_\lambda\sigma_3^2$ and $\boldsymbol{\theta}_{1\tau}$ has a one-to-one correspondence with $\boldsymbol{\theta}_\tau$. Therefore, $\boldsymbol{\theta}_*$ is also identified. ■

**(ii)** We use the representation $v_{3i} = (\sigma_{13}/\sigma^2)u_{1i} + (\sigma_{33}/\sigma_3^2)u_{3i} + v_{4i}$, where $\sigma_{13} = \mathcal{E}[u_{1i}v_{3i}]$ and $\sigma_{33} = \mathcal{E}[u_{3i}v_{3i}]$. For the density function of the simultaneous generalized selectivity model, we suppose the following identity:

$$\begin{aligned} l_i(\boldsymbol{\theta}_{1*}) &= \log \frac{1}{\sigma}\phi\left(\frac{u_{1i}}{\sigma}\right)\left[\frac{1}{\sigma_3}\phi\left(\frac{y_{2i}}{\sigma_3} - \boldsymbol{\pi}_3'\mathbf{z}_i - \pi_\rho u_{1i}\right)\Phi\left(\frac{1}{\sigma_4}\left\{\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\sigma u_{1i} + \pi_\lambda u_{3i}\right\}\right)\right]^{y_{3i}} \\ &\quad \times \left[1 - \Phi\left(\frac{1}{\sigma_5}\left\{\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\sigma u_{1i}\right\}\right)\right]^{1-y_{3i}} \\ &= \log \frac{1}{\sigma}\phi\left(\frac{u_{01i}}{\sigma_0}\right)\left[\frac{1}{\sigma_{03}}\phi\left(\frac{y_{2i}}{\sigma_{03}} - \boldsymbol{\pi}_{03}'\mathbf{z}_i - \pi_{\rho 0} u_{01i}\right)\Phi\left(\frac{1}{\sigma_{04}}\left\{\boldsymbol{\pi}_{03*}'\mathbf{z}_i + \pi_{\sigma 0} u_{01i} + \pi_{\lambda 0} u_{03i}\right\}\right)\right]^{y_{3i}} \\ &\quad \times \left[1 - \Phi\left(\frac{1}{\sigma_{05}}\left\{\boldsymbol{\pi}_{03*}'\mathbf{z}_i + \pi_{\sigma 0} u_{01i}\right\}\right)\right]^{1-y_{3i}}, \end{aligned} \tag{A.22}$$

where $\pi_\sigma = \sigma_{13}/\sigma^2$, $\pi_\lambda = \sigma_{33}/\sigma_3^2$, $\sigma_4 = (1-\pi_\sigma^2\sigma^2 - \pi_\lambda^2\sigma_3^2)^{1/2}$, and $\sigma_5 = (1-\sigma_{13}^2/\sigma^2)^{1/2} = (1-\pi_\sigma^2\sigma^2)^{1/2}$. Therefore, the transformed parameters $\boldsymbol{\theta}_{1*} = (\boldsymbol{\theta}_{1\tau}',\ \pi_\sigma,\ \pi_\lambda,\ \boldsymbol{\pi}_{3*}')'$ do not include $\sigma_4$ and $\sigma_5$. The representative subscript 0 is evaluated at the true value. Partial differentiation is possible with respect to $y_{1i}$ by fixing $v_{02i}$ and $v_{03i}$, which is also true for $y_{2i}$. Consider the third-order partial derivative of $y_{1i}$ and taking $y_{3i} = 0$, the conditions $\pi_\sigma/\sigma_5 = \pi_{\sigma 0}/\sigma_{05}$ and $\sigma = \sigma_0$ are obtained from the result of Proof of (ii). Then, we have $\sigma_{13} = \pm[\sigma_0^4(\pi_{\sigma 0}/\sigma_{05})^2/(1+\sigma_0^2(\pi_{\sigma 0}/\sigma_{05})^2)]^{1/2}$ where the sign of $\sigma_{13}$ is determined by $\mathrm{sgn}(\pi_{\sigma 0})$. Hence, $\pi_\sigma = \pi_{\sigma 0}$ and $\sigma_5 = \sigma_{05}$ can be fixed.

The first partial derivative of $y_{2i}$ becomes

$$\frac{u_{1i}}{\sigma^2}\beta_1 + y_{3i}\left[\frac{u_{3i}}{\sigma_3}\left(\frac{1}{\sigma_3} + \pi_\rho\beta_1\right) + r(g(u_i))\pi_\beta\right] = \frac{u_{01i}}{\sigma^2}\beta_{01} + y_{3i}\left[\frac{u_{03i}}{\sigma_{03}}\left(\frac{1}{\sigma_{03}} + \pi_{\rho 0}\beta_{01}\right) + r(g_0(u_{0i}))\pi_{\beta 0}\right], \tag{A.23}$$

where $g(u_i) = (1/\sigma_4)(\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\sigma u_{1i} + \pi_\lambda u_{3i})$ and $g_0(u_{0i}) = (1/\sigma_{04})(\boldsymbol{\pi}_{03*}'\mathbf{z}_i + \pi_{\sigma 0}u_{01i} + \pi_{\lambda 0}u_{03i})$, and $\pi_\beta$ and $\pi_{\beta 0}$ are defined by (A.25). (A.23) is obtained since $u_{3i}/\sigma_3 = -y_{2i}/\sigma_3 + \boldsymbol{\pi}_3'\mathbf{z}_i + \pi_\rho u_{1i}$ and $y_{2i} = 0$ is constant in $(1-y_{3i})\log[1-\Phi((1/\sigma_5)\{\boldsymbol{\pi}_{3*}'\mathbf{z}_i + \pi_\sigma u_{1i}\})]$. The cross-partial derivative is given by the partial derivative of (A.23) with respect to $y_{1i}$, and taking $y_{3i} = 0$, it follows that $\beta_1/\sigma^2 = \beta_{01}/\sigma^2$. Therefore, we obtain the condition $\beta_1 = \beta_{01}$.

We have the following third-order partial derivative of $y_{1i}$ and $y_{2i}$ under $y_{3i} = 1$, respectively:

$$\ddot{r}(g(u_i))\left(\frac{\pi_\sigma + \pi_\lambda\pi_\rho\sigma_3}{\sigma_4}\right)^3 = \ddot{r}(g_0(u_{0i}))\left(\frac{\pi_{\sigma 0} + \pi_{\lambda 0}\pi_{\rho 0}\sigma_{03}}{\sigma_{04}}\right)^3, \text{ and} \tag{A.24}$$

$$\ddot{r}(g(u_i))\left(\frac{-\beta_1\pi_\sigma - \pi_\lambda(1+\pi_\rho\sigma_3\beta_1)}{\sigma_4}\right)^3 = \ddot{r}(g_0(u_{0i}))\left(\frac{-\beta_{01}\pi_{\sigma 0} - \pi_{\lambda 0}(1+\pi_{\rho 0}\sigma_{03}\beta_{01})}{\sigma_{04}}\right)^3. \tag{A.25}$$

Taking $g(u_i) = g_0(u_{0i})$ is possible for some realized values of $u_{01i}$ and $u_{03i}$. Thus, we obtain $\pi_\alpha = \pi_{\alpha 0}$ and $\pi_\beta = \pi_{\beta 0}$ since $\ddot{r}(x) > 0$, where $\pi_\alpha = (\pi_\sigma + \pi_\lambda\pi_\rho\sigma_3)/\omega_4$, $\pi_{\alpha 0} = (\pi_{\sigma 0} + \pi_{\lambda 0}\pi_{\rho 0}\sigma_{03})/\omega_{04}$, $\pi_\beta = (-\beta_1\pi_\sigma - \pi_\lambda(1+\pi_\rho\sigma_3\beta_1))/\sigma_4$, and $\pi_{\beta 0} = (-\beta_{01}\pi_{\sigma 0} - \pi_{\lambda 0}(1+\pi_{\rho 0}\sigma_{03}\beta_{01}))/\sigma_{04}$.

Then, we obtain the second partial derivatives of $y_{1i}$ and $y_{2i}$ under $y_{3i} = 1$ as follows, respectively:

$$-\frac{1}{\sigma^2} - \pi_\rho^2 - \dot{r}(g(u_i))\pi_\alpha^2 = -\frac{1}{\sigma^2} - \pi_{\rho 0}^2 + \dot{r}(g_0(u_{0i}))\pi_\alpha^2 \text{ , and} \tag{A.26}$$

$$-\frac{\beta_1^2}{\sigma^2} - \left(\frac{1}{\sigma_3} + \pi_\rho\beta_1\right)^2 + \dot{r}(g(u_i))\pi_\beta^2 = -\frac{\beta_1^2}{\sigma^2} - \left(\frac{1}{\sigma_{03}} + \pi_{\rho 0}\beta_1\right)^2 + \dot{r}(g_0(u_{0i}))\pi_\beta^2 \text{ .} \tag{A.27}$$

The cross-partial derivative under $y_{3i} = 1$ is given by the partial derivative of (A.23) with respect to $y_{1i}$:

$$\frac{\beta_1}{\sigma^2} + \frac{\pi_\rho}{\sigma_3}\left(\frac{1}{\sigma_3} + \pi_\rho\beta_1\right) + \dot{r}(g(u_i))\pi_\alpha\pi_\beta = \frac{\beta_1}{\sigma^2} + \frac{\pi_{\rho 0}}{\sigma_{03}}\left(\frac{1}{\sigma_{03}} + \pi_{\rho 0}\beta_1\right) + \dot{r}(g_0(u_{0i}))\pi_\alpha\pi_\beta \text{ .} \tag{A.28}$$

Taking $g(u_i) = g_0(u_{0i})$, we obtain the following nonlinear equations with respect to $\pi_\rho$ and $\sigma_3$ given $\beta_1 = \beta_{01}$:

$$\pi_\rho^2 = \pi_{\rho 0}^2 \text{ ,} \tag{A.29}$$

$$\left(\frac{1}{\sigma_3} + \pi_\rho\beta_1\right)^2 = \left(\frac{1}{\sigma_{03}} + \pi_{\rho 0}\beta_1\right)^2 \text{ , and} \tag{A.30}$$

$$\frac{\pi_\rho}{\sigma_3}\left(\frac{1}{\sigma_3} + \pi_\rho\beta_1\right) = \frac{\pi_{\rho 0}}{\sigma_{03}}\left(\frac{1}{\sigma_{03}} + \pi_{\rho 0}\beta_1\right) \text{ .} \tag{A.31}$$

If $\pi_{\rho 0} = 0$ or $\beta_{01} = 0$, then $\sigma_3 = \sigma_{03}$ by (A.30). Squaring (A.31), we have $\pi_\rho^2/\sigma_3^2 = \pi_{\rho 0}^2/\sigma_{03}^2$, if $1/\sigma_{03} \neq -\pi_{\rho 0}\beta_{01}$ and $\pi_{\rho 0}\beta_{01} \neq 0$. Thus, $\sigma_3 = \sigma_{03}$ holds. When $1/\sigma_{03} = -\pi_{\rho 0}\beta_{01}$ and $\pi_{\rho 0}\beta_{01} \neq 0$, it follows that $1/\sigma_3^2 = \pi_\rho^2\beta_1^2 = 1/\sigma_{03}^2$. Therefore, we obtain $\sigma_3 = \sigma_{03}$. From $\pi_\rho = \pm\pi_{\rho 0}$ and (A.30), it follows that $\pi_\rho = \pi_{\rho 0}$ under $\beta_{01} \neq 0$. If $\beta_{01} = 0$, then $\pi_\rho = \pi_{\rho 0}$ holds by (A.31). Thus, we obtain $\pi_\rho = \pi_{\rho 0}$.

Using above results, $\pi_\beta = \pi_{\beta 0}$, and $\beta_{01}\pi_\alpha = \beta_{01}\pi_{\alpha 0}$ of (A.24), we have the relation that $\pi_\lambda(1 + \pi_{\rho_0}\sigma_{03}\beta_{01} - \pi_{\rho_0}\sigma_{03}\beta_{01}) = \pi_\lambda = -(\pi_{\beta 0} + \beta_{01}\pi_{\alpha 0})\sigma_4$. Therefore, by solving it for $\pi_\lambda$, we obtain

$$\pi_\lambda = \pm\sqrt{\frac{(\pi_{\beta 0} + \beta_{01}\pi_{\alpha 0})^2(1 - \pi_{\sigma 0}^2\sigma_0^2)}{1 + (\pi_{\beta 0} + \beta_{01}\pi_{\alpha 0})^2\sigma_{03}^2}} \text{ ,} \tag{A.32}$$

where the sign is determined by sgn$(-(\pi_{\beta 0} + \beta_{01}\pi_{\alpha 0}))$=sgn$(\pi_{\lambda 0})$. Thus, we obtain $\pi_\lambda = \pi_{\lambda 0}$ since $\pi_{\lambda 0}$ also satisfies the relation $\pi_{\lambda 0} = -(\pi_{\beta 0} + \beta_{01}\pi_{\alpha 0})\sigma_{04}$. From (A.32), we obtain that $\sigma_4 = \sigma_{04}$.

Therefore, we have the following conditions:

$$(1/\sigma,\ 1/\sigma_3,\ 1/\sigma_4,\ 1/\sigma_5,\ \pi_\rho,\ \pi_\sigma,\ \pi_\lambda) = (1/\sigma_0,\ 1/\sigma_{03},\ 1/\sigma_{04},\ 1/\sigma_{05},\ \pi_{\rho 0},\ \pi_{\sigma 0},\ \pi_{\lambda 0}) \text{ .} \tag{A.33}$$

Thus, the arguments of $\phi$ and $\Phi$ become linear in the parameters given these conditions. Additionally, the identification of $l_i(\boldsymbol{\theta}_{1*})$ is shown similar to Proof of (ii). The density function with $\boldsymbol{\theta}_*$ of the simultaneous generalized selectivity model is also identified since $\boldsymbol{\theta}_{1*}$ has a one-to-one correspondence with $\boldsymbol{\theta}_*$. ■

**Acknowledgments** The authors are grateful to Erricos J. Kontoghiorghes, the Associate Editor, and two anonymous reviewers for constructive comments as well as the participants of the JSS 2021. All remaining errors are ours.

## References

[1] Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables Are Truncated Normal," *Econometrica*, Vol. 42, 999-1012.

[2] Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

[3] Blundell, R. and R. J. Smith (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, Vol. 64, 355-373.

[4] Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, Vol. 46, 931-959.

[5] Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47, 153-161.

[6] Olsen, R. J. (1978), "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," *Econometrica*, Vol. 46, 1211-1215.

[7] Olsen, R. J. (1982), " Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator," *International Economic Review*, Vol. 23, 223-240.

[8] Pratt, J. W. (1981), "Concavity of the Log Likelihood," *Journal of the American Statistical Association*, Vol. 76, 103-106.

[9] Zuehlke, T. W. (2021), "Estimation of a type 2 Tobit Model with Generalized Box-Cox Transformation," *Applied Economics*, Vol. 17, 1952-1975.