Generalized vs. Standard Heckman Probit

Nicholas Renegar
Operations Research Center, Massachusetts Institute of Technology, 1 Amherst St, Cambridge, MA 02142
renegar@mit.edu

January 7, 2020

Overview This technical note shows an equivalence between the standard Heckman probit model (with variance 1 for the selection and outcome model), to the generalized version (with different variances allowed for the selection and outcome models). This demonstrates that the standard Heckman probit can be used without loss of generality.

Initial Formulation - Heckman Probit
Let $S \in \{0,1\}$ be whether someone is sampled.
Let $R \in \{0,1\}$ be the outcome if someone is sampled.
Let Z be the independent variables for the sampling model.
Let X be the independent variables for the outcome model.

Our generalized Heckman probit model has parameters γ, β, ρ and bivariate noise terms u_1 and u_2 with covariance matrix

$$\begin{pmatrix} \sigma_S^2 & \rho\sigma_S\sigma_R \\ \rho\sigma_S\sigma_R & \sigma_R^2 \end{pmatrix}$$

Where in the standard model, we just have $\sigma_S^2 = \sigma_R^2 = 1$.

The model form is:

$$\mathbb{P}(S=1|Z) = \mathbb{P}(\gamma Z + u_1 > 0)$$
$$\mathbb{P}(R=1|X) = \mathbb{P}(\beta X + u_2 > 0)$$

Cholesky Decomposition
Using the Cholesky decomposition, we can decompose u_1 and u_2 into the sum of independent standard normal random variables e_1 and e_2. We have that:

$$u_1 = \sigma_S e_1$$
$$u_2 = \rho\sigma_R e_1 + \sqrt{1-\rho^2}\sigma_R e_2$$

Standard Heckman Probit
In the standardized model, we have $\sigma_S^2 = \sigma_R^2 = 1$. So we get for the sampling model that:

$$\mathbb{P}(S=1|Z) = \mathbb{P}(\gamma Z + u_1 > 0)$$
$$\to \mathbb{P}(S=1|Z) = \mathbb{P}(\gamma Z + e_1 > 0) \tag{1}$$

And for the outcome model that:

$$\mathbb{P}(R=1|X,Z,S=1) = \mathbb{P}(\beta X + u_2 > 0|e_1 > -\gamma Z)$$
$$\to \mathbb{P}(R=1|X,Z,S=1) = \mathbb{P}(\beta X + \rho e_1 + \sqrt{1-\rho^2}e_2 > 0|e_1 > -\gamma Z) \tag{2}$$

Generalized Heckman Probit
In the generalized model, σ_S^2 and σ_R^2 are model parameters that can vary. We get for the sampling model that:

$$\mathbb{P}(S = 1|Z) = \mathbb{P}(\gamma Z + u_1 > 0)$$
$$\rightarrow \mathbb{P}(S = 1|Z) = \mathbb{P}(\gamma Z + \sigma_S e_1 > 0) \tag{3}$$

And for the outcome model that:

$$\mathbb{P}(R = 1|X, Z, S = 1) = \mathbb{P}(\beta X + u_2 > 0|e_1 > -\frac{\gamma Z}{\sigma_S})$$

$$\rightarrow \mathbb{P}(R = 1|X, Z, S = 1) = \mathbb{P}(\beta X + \rho\sigma_R e_1 + \sqrt{1-\rho^2}\sigma_R e_2 > 0|e_1 > -\frac{\gamma Z}{\sigma_S}) \tag{4}$$

Showing Equivalency

For any solution to the generalized Heckman, we have model parameters $\gamma, \beta, \rho, \sigma_S, \sigma_R$. We take:

$$\gamma' = \gamma/\sigma_S$$
$$\rho' = \rho$$
$$\beta' = \beta/\sigma_R$$

Rewriting Equation 3 we get:

$$\mathbb{P}(S = 1|Z) = \mathbb{P}(\gamma' Z + e_1 > 0) \tag{5}$$

Rewriting Equation 4 we get:

$$\mathbb{P}(R = 1|X, Z, S = 1) = \mathbb{P}(\beta' X + \rho' e_1 + \sqrt{1-\rho'^2} e_2 > 0|e_1 > -\gamma' Z) \tag{6}$$

So this is equivalent to a model formulation of the standard Heckman probit given in Equations 1 and 2. So without loss of generality, we only have to consider Heckman probit models in the standard form, with $\sigma_S = \sigma_R = 1$. $\square$