Economics Letters 125 (2014) 360–363

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A note on the identification in two equations probit model with dummy endogenous regressor[☆]

Ismael Mourifié[a], Romuald Méango[b,*]

[a] Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada

[b] Ifo Institute, Munich, Poschingerstr. 5, 81679, Munich, Germany

HIGHLIGHTS

- We study identification in two equation probit models with endogenous dummy regressor.
- Parameters are set identified without exclusion restriction.
- Numerical evidence contradict Wilde (2000).

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form
9 October 2014
Accepted 9 October 2014
Available online 22 October 2014

JEL classification:
C35

Keywords:
Probit model
Endogenous dummy regressor
Partial identification

ABSTRACT

This paper deals with the question whether exclusion restrictions on the exogenous regressors are necessary to identify two equation probit models with endogenous dummy regressor. We show that Wilde (2000)'s criterion is insufficient for (point) identification.

1. Introduction

This paper discuss identification in the following two equations probit model with endogenous dummy regressor.

$$Y_1 = I\left(x_1^T\beta_1 + u_1 > 0\right) \tag{1}$$

$$Y_2 = I\left(\delta Y_1 + x_2^T\beta_2 + u_2 > 0\right) \tag{2}$$

where

$$(u_1, u_2) \text{ follows } N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right),$$

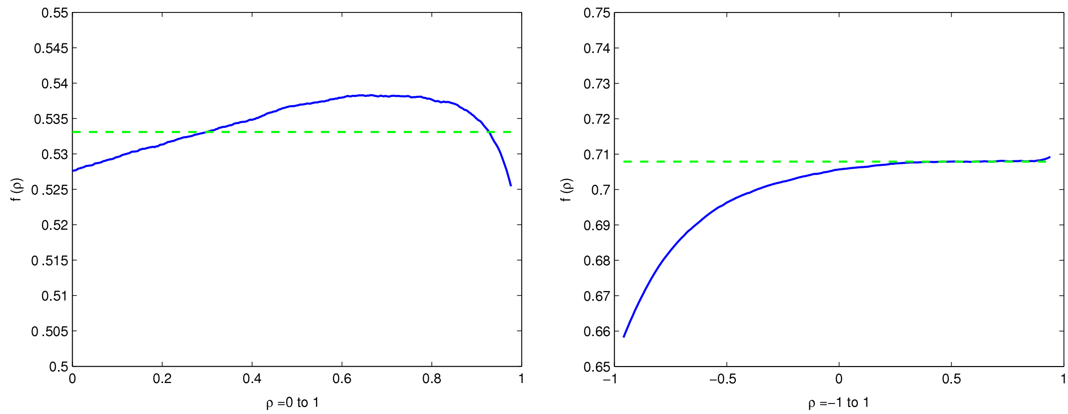
$I(A) = 1$ if A is true and zero otherwise and $\rho \in (-1, 1)$ (see Sartori, 2003, for a treatment of the case $\rho = 1$). In all the paper, we shall use the notation $\Phi_2(\cdot, \cdot; \rho)$ to denote the bivariate normal standard cumulative distribution with correlation parameter ρ and $\phi_2(\cdot, \cdot; \rho)$ the corresponding bivariate density function. We denote by $\Phi(\cdot)$ the univariate normal standard cumulative distribution and by $\phi(\cdot)$ the corresponding density function. We say that there exists an exclusion restriction when there exists a variable in x_1 that does not appear in x_2.

Two main opinions dominate the literature about identification in this model. On one hand, Maddala (1983, p. 122) claimed that an exclusion variable is necessary for identification. His argument was based on the following fact. In a case where $x_1 = x_2 = 1$, the model has four different parameters to be identified β_1, β_2, δ and ρ, while we observe only three independent probabilities. Adding an exclusion variable increases the number of observed independent probabilities, thus enabling the number of observed independent probabilities to be larger than or equal to the number of parameters to be identified. On the other hand, Wilde (2000) notes that even without an exclusion variable, the presence of only a common dichotomous covariate might result in the number of observed independent probabilities equating the number of parameters to

[☆] Financial support by the Leibniz Association (SAW-2012-ifo-3) is gratefully acknowledged. The authors are grateful to Joris Pinkse, Rachidi Kotchoni and Marc Henry and an anonymous referee for helpful discussions and comments.

[*] Corresponding author.
E-mail address: meango@ifo.de (R. Méango).

http://dx.doi.org/10.1016/j.econlet.2014.10.006
0165-1765/



(a) $(\rho, f(\rho))$ region, $\beta_{21} = -0.4$, $\rho_0 = -0.3$.

(b) $(\rho, f(\rho))$ region, $\beta_{21} = 0.4$, $\rho_0 = 0.5$.

Fig. 1. Numerical results: $f(\rho)$ (plain blue line). The straight dotted line is the observed probability P_{111}. $\beta_{11} = 0.3$, $\beta_{12} = 0.4$, $\beta_{22} = 0.5$, $\delta = 0.3$.

be identified. Therefore, following the assertion of Heckman (1978, p. 957) in a more general context, Wilde (2000) argued that only the full rank of the (regressor) data matrix is needed to identify all the model parameters.

We show that the simple criterion proposed by Wilde (2000) and the rank condition proposed by Heckman (1978) are not sufficient to ensure identification in Models (1)–(2) for the following reason: the fact that the number of unknown is larger than or equal to the number of independent probabilities does not ensure unicity of the solution since the system of equations is nonlinear in the parameters. We provide numerical evidence that contradicts the result of Wilde and suggests that the model without exclusion is usually only partially identified.

Finally, we point out that beside the fact that an exclusion variable increases the number of observed independent probabilities, its intrinsic feature to shift the selection equation (1) by keeping fix the outcome equation (2) allows us to point identify the model. All our results hold, also, for a sample selection model with binary outcome.

2. Failure of point identification

We consider the simple case where a dichotomous regressor enters both equations. In (1) and (2), let $x_1^T = x_2^T = [1, x]$ and $\beta_1 = [\beta_{11}, \beta_{12}]^T$ the associated parameters where $x \in \{0, 1\}$, a binary regressor. As noted by Wilde, we observe now 6 independent probabilities, and we have 6 parameters to identify i.e $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \delta, \rho)$. We will use the following notation:

$$P_{ijk} \equiv P(Y_1 = i, Y_2 = j | x = k) \quad \text{for all } (i, j, k) \in \{0; 1\}^3. \tag{3}$$

Wilde argued that with 6 independent equations and 6 parameters, we have now enough variation in the model to identify the parameters, unlike in the case without covariates where we had 3 independent equations with 4 parameters. Although this argument is a sensible one when the equations are linear in the parameters, it is likely to fail when linearity or monotonicity does not hold. For instance, consider the following trivial nonlinear single equation with one parameter $\rho^2 - \frac{1}{4} = 0$.

First, note that $\beta_1 = [\beta_{11}, \beta_{12}]^T$ will be identified from the usual hypothesis on a probit model with the outcome variable Y_1. Second, since the error terms are jointly normally distributed with correlation ρ, we can write: $u_1 = \rho u_2 + e$ where e follows $\mathcal{N}\left(0, 1 - \rho^2\right)$ and e is independent of u_2. Therefore, we can derive the following equalities:

$$P_{010} = \int_{-\beta_{21}}^{+\infty} \left(\Phi \left(-\frac{\beta_{11} + \rho y}{\sqrt{1 - \rho^2}} \right) \right) \phi(y)\ dy \tag{4}$$

$$P_{110} = \int_{-\beta_{21}-\delta}^{+\infty} \left(\Phi \left(\frac{\beta_{11} + \rho y}{\sqrt{1 - \rho^2}} \right) \right) \phi(y)\ dy \tag{5}$$

$$P_{011} = \int_{-\beta_{21}-\beta_{22}}^{+\infty} \left(\Phi \left(-\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1 - \rho^2}} \right) \right) \phi(y)\ dy \tag{6}$$

$$P_{111} = \int_{-\beta_{21}-\beta_{22}-\delta}^{+\infty} \left(\Phi \left(\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1 - \rho^2}} \right) \right) \phi(y)\ dy. \tag{7}$$

Note now the following: since β_{11} is identified and the integrand is always positive, once you fix a value for ρ, the right-term of Eq. (4) is strictly monotone in β_{21}. It follows that we identify a unique value for β_{21} given ρ.

By using the same recursive solving strategy applied to Eqs. (5)–(6), we find that all the parameters are identified given a value of ρ. The question is whether ρ will be identified once we consider also (7). Once we solve the first three equations for $\beta_2 = [\beta_{21}, \beta_{22}]^T$ and δ given ρ, the support of the integral on the right-hand side term (RHS) of Eq. (7) depends on ρ, and the latter is not necessary monotone with respect to ρ. The following numerical results suggest the nonmonotonicity of this function and find that several values of ρ might solve the system of equation.[1]

Denote by $f(\rho)$ the RHS of Eq. (7):

$$f(\rho) = \int_{-\beta_{21}^*(\rho)-\beta_{22}^*(\rho)-\delta^*(\rho)}^{+\infty} \left(\Phi \left(\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1 - \rho^2}} \right) \right) \phi(y)\ dy \tag{8}$$

where $\beta_{21}^*(\rho), \beta_{22}^*(\rho), \delta^*(\rho)$ solve Eqs. (4)–(6) given ρ. Fig. 1 plots $f(\cdot)$ for $\rho \in (-1, 1)$ given different values of the other parameters.[2]

Considering the first set of parameters (Fig. 1(a)), $f(\rho)$ exhibits a nonmonotonic behavior, increasing first, then decreasing after

[1] Details on the numerical are exposed in the Appendix section. The routines can be found on the following link: https://sites.google.com/site/ismaelymourifie/research-papers.

[2] Note that the endpoints, where ρ approaches -1 and 1, are trimmed for better readability. The numerical approximation behaves poorly, mainly because of the term $\sqrt{1 - \rho^2}$ in the denominator.