University of Toronto
Department of Economics

Working Paper 503

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

By Romuald Meango and Ismael Mourifie

October 14, 2013

A NOTE ON THE IDENTIFICATION IN TWO EQUATIONS PROBIT MODEL WITH DUMMY ENDOGENOUS REGRESSOR

ROMUALD MÉANGO AND ISMAEL MOURIFIÉ

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. Contradictory opinions have been exposed in the literature on the necessity of an exclusion restriction. Wilde (2000) argued that an exclusion restriction is not necessary, and proposed a simple criterion for identification in this model. We contradict his result, and show how the inherent incompleteness of the model leads to failure of (point) identification. We provide an exact identification proof when an exclusion restriction is available.

Keywords: Probit model, Endogenous dummy regressor, Partial identification
JEL subject classification: C35

Introduction

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

$$Y_1 = I\left(x_1^T\beta_1 + u_1 > 0\right) \tag{0.1}$$

$$Y_2 = I\left(\delta Y_1 + x_2^T\beta_2 + u_2 > 0\right) \tag{0.2}$$

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 $\Phi_2(.,.;\rho)$ to denote the bivariate normal standard cumulative distribution with correlation parameter ρ and $\phi_2(.,.;\rho)$ the corresponding bivariate density function. We denote $\Phi(.)$ the univariate normal standard cumulative distribution

Date: October 12, 2013. The authors are grateful to Joris Pinkse, Rachidi Kotchoni and Marc Henry for helpful discussions and comments. Correspondence address: Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada, ismael.mourifie@utoronto.ca .

and $\phi(.)$ 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 dominates 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 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 Model (0.1) - (0.2) 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.

In the following, we first complement the result of Maddala (1983) by proving that in the model without covariate the parameters of Eq.(0.2) are partially identified. We propose the full characterization of the identified set.

Second, we give numerical evidences that contradict the result of Wilde, and suggest that the model without exclusion is usually only partially identified. Finally, we show that beside the fact that an exclusion variable increases the number of observed independent probabilities, its intrinsic feature to shift the selection equation (0.1) by keeping fix the outcome equation (0.2) allows to point identify the model. All our results hold, also, for a sample selection model with binary outcome.

1. Failure of point identification: two cases

We consider first the example of Maddala (1983) where there is no covariate, following by the case where a dichotomous regressor enters both equations.

1.1. **Identification in absence of covariate.**

Example 1 (Maddala's biprobit with endogenous dummy regressor)**.** *Consider the following case*

$$\begin{aligned} Y_1 &= I\left(\beta_1 + u_1 > 0\right) && (1.1)\\ Y_2 &= I\left(\delta Y_1 + \beta_2 + u_2 > 0\right). && (1.2) \end{aligned}$$

The parameters of interest are $(\beta_1, \beta_2, \delta, \rho)$*.*

The researcher observes the following probability distribution from the data:

$$P_{ij} \equiv P\left(Y_1 = i, Y_2 = j\right) \text{ for all } i, j \in \{0,1\}^2.$$

Note that only three of these quantities are informative since the fourth one can be easily derived from the three others. Maddala's argument for failure of identification is that the researcher has four parameters of interest, but only three independent probabilities. We will show that the model as stated fails to put any restriction on the correlation parameter.

Since the error terms are jointly normally distributed with correlation ρ, we can write:

$$u_1 = \rho u_2 + e \text{ where } e \text{ follows } \mathcal{N}\left(0, 1-\rho^2\right) \tag{1.3}$$

and e is independent of u_2.

β_1 will be identified from the usual hypothesis on a probit model with the outcome variable Y_1, since

$$P_{10} + P_{11} = \Phi(\beta_1). \tag{1.4}$$

Moreover, we have:

$$\begin{aligned} P_{01} &= P\left(\frac{e}{\sqrt{1-\rho^2}} < -\frac{\beta_1 + \rho u_2}{\sqrt{1-\rho^2}}; u_2 > -\beta_2\right) \\ &= \int_{-\beta_2}^{+\infty}\left(\Phi\left(-\frac{\beta_1 + \rho y}{\sqrt{1-\rho^2}}\right)\right)\phi\left(y\right)\, dy && (1.5) \end{aligned}$$

Similarly, we can have:

$$P_{11} = \int_{-\beta_2-\delta}^{+\infty}\Phi\left(\frac{\beta_1 + \rho y}{\sqrt{1-\rho^2}}\right)\phi\left(y\right)\, dy \tag{1.6}$$

Note now the following: since β_1 is identified and the integrand is always positive, once you fix a value for ρ, the right-term of Eq.(1.5) is strictly monotone in β_2. It follows that we identify a unique value for β_2 given ρ. The same reasoning for Eq.(1.6) leads to conclude that δ is identified given a value ρ.

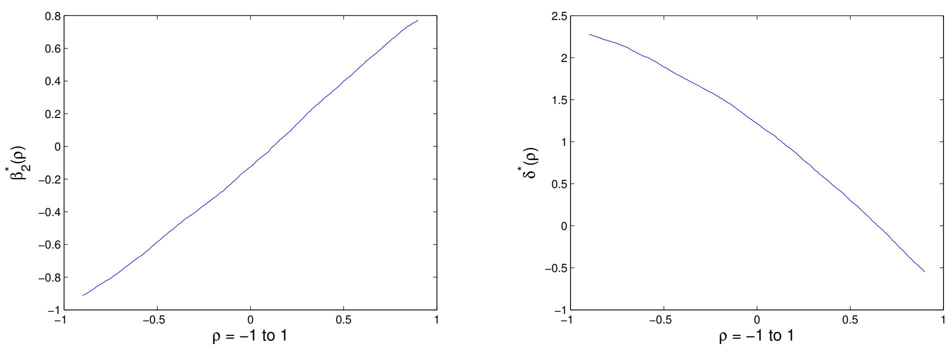
By recursively solving (1.4), (1.5), (1.6) for β_1, β_2, δ, we exhaust all the independent variations in the model and no restriction exists on the parameter ρ. That is, for $\rho \in (-1,1)$, we either find a triple $(\beta_1, \beta_2^*(\rho), \delta^*(\rho))$ that satisfies (1.4), (1.5), (1.6).

The identified set is therefore a box in $\mathbb{R}^4$ characterized in the following way:

Proposition 1. *Consider the model in 1. Denote Θ the identified set. We have:*

$$\Theta = \{(\beta_1, \beta_2^*(\rho), \delta^*(\rho), \rho) : \rho \in (-1,1) \text{ and } (\beta_1, \beta_2^*(\rho), \delta^*(\rho), \rho) \text{ satisfies (1.4), (1.5), (1.6) } \}$$

We note that β_1 is identified, ρ is completely nonidentified, but β_2 and δ are partially identified. Indeed, if $Dom(\beta_2^*(\rho)) \subset \mathbb{R}$, β_2 is partially identified, similarly for δ. In the following simulation we will show that $Dom(\beta_2^*(\rho))$ and $Dom(\delta^*(\rho))$ could be considerably small is some cases and then informative. Simulation results displayed in Figure 1 illustrate our findings.



(a) (ρ, β_2) region

(b) (ρ, δ) region

FIGURE 1. Numerical results: parameters generating observed probabilities close ($< 1e^{-8}$) to the true probabilities. $\beta_1^0 = 0.3, \beta_2^0 = 0.4, \delta^0 = 0.3, \rho_0 = 0.5$.

Fig. 1 presents projection of the identified set on different plans. The ranges of the parameters ($\beta_2 \in (-1.0921; 1.0689)$ and $\delta \in (-0.8409; 2.1867)$) are substantially reduced in comparison to their respective domain (the real line).

1.2. **Introducing a covariate.** In (0.1) and (0.2), 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:

$$P_{ijk} \equiv P\left(Y_1 = i, Y_2 = j | x = k\right) \text{ for all } (i,j,k) \in \{0;1\}^3. \tag{1.7}$$

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 $\rho^2 - \frac{1}{4} = 0$.

As before, $\beta_1 = [\beta_{11}, \beta_{12}]^T$ will be identified from the usual hypothesis on a probit model with the outcome variable Y_1. Similarly, to the previous section, we can derive the following:

$$P_{010} = \int_{-\beta_{21}}^{+\infty} \left(\Phi\left(-\frac{\beta_{11} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi(y)\ dy \tag{1.8}$$

$$P_{110} = \int_{-\beta_{21}-\delta}^{+\infty} \left(\Phi\left(\frac{\beta_{11} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi(y)\ dy \tag{1.9}$$

$$P_{011} = \int_{-\beta_{21}-\beta_{22}}^{+\infty} \left(\Phi\left(-\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi(y)\ dy \tag{1.10}$$

$$P_{111} = \int_{-\beta_{21}-\beta_{22}-\delta}^{+\infty} \left(\Phi\left(\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi(y)\ dy. \tag{1.11}$$

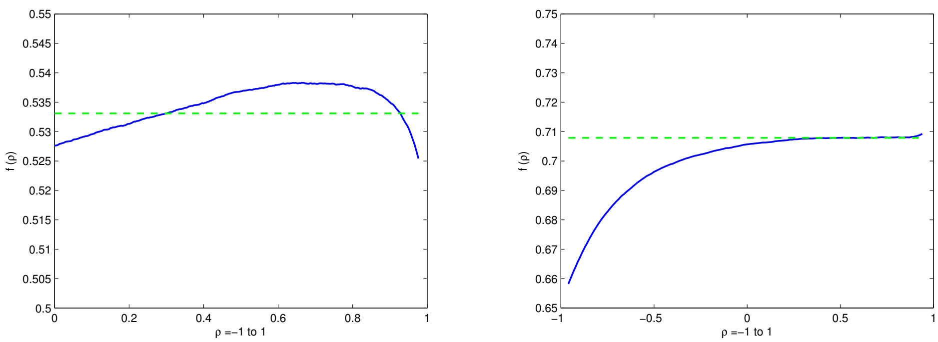
By using the earlier recursive solving strategy applied to Eq. (1.8) - (1.10), we find all the parameters are identified given a value of ρ. The question is whether ρ will be identified once we consider also (1.11). Once we solve the first three equations for β_2 and δ given ρ, the support of the integral in right-hand side term (RHS) of Eq. (1.11) 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.

Denote $f(\rho)$, the RHS (1.11):

$$f(\rho) = \int_{-\beta_{21}^*(\rho)-\beta_{22}^*(\rho)-\delta^*(\rho)}^{+\infty} \left(\Phi\left(\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi(y)\ dy \tag{1.12}$$

where $\beta_{21}^*(\rho), \beta_{22}^*(\rho), \delta^*(\rho)$ solve Eq. (1.8) - (1.10) given ρ. Fig. 2 plots $f(.)$ for $\rho \in (-1, 1)$ given different values of the other parameters.

Considering the first set of parameters (Fig. 2(a)), $f(\rho)$ exhibits a nonmontonic behavior, increasing first, then decreasing after reaching a maximum. The identified set consists of two singletons. Considering the second set of parameters (Fig. 2(b)), we observe that $f(\rho)$ is (relatively) flat in the neighborhood of δ, suggesting weak or set identification. Several values of ρ contained in the interval $[0.483; 0.596]$ deliver probability values close to the value observed ($|f(\rho) - P_{111}| < 1e-4$).



(a) $(\rho, f(\rho))$ region, $\beta_{21} = -0.4$, $\rho_0 = -0.3$ (b) (ρ, δ) region, $\beta_{21} = 0.4$, $\rho_0 = 0.5$

FIGURE 2. Numerical results: $f(\rho)$ (plain blue line). The straight dotted line is the observed probability P_{111}. $\beta_{11} = 0.3, \beta_{12} = 0.4, \beta_{22} = 0.5, \delta = 0.3$.

Remark 1. *The full rank condition suggested by Heckman (1978, p 957) holds for all random variables X, Y_1 such that $\mathbb{E}[X^2 Y_1^2] \neq 0$. This implies that the full rank condition is not sufficient to point identified the model.*

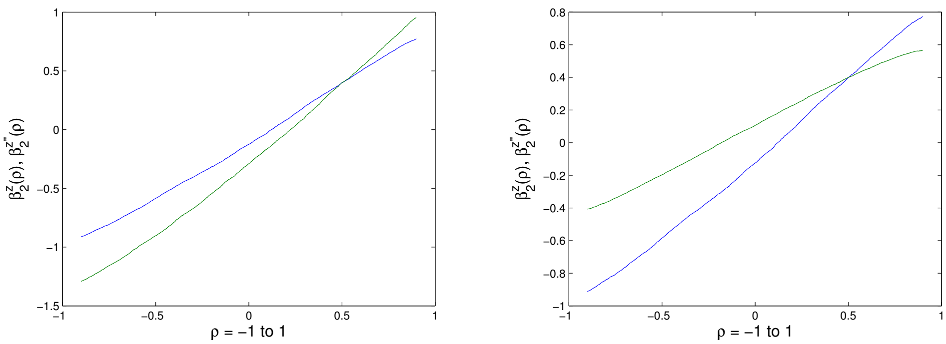
One, might think that the identified set will shrink to a point as soon as the covariate is non-binary. In fact, if $x_1^T = x_2^T \in \{0, 1, 2\}$ we have 9 independent probabilities and 6 parameters to be identified. We would think that there is an overidentication, but there is no necessary identification due to the nonlinearity of the system. Indeed, in addition to (1.8), (1.10), we have the following equation:

$$P_{012} \quad = \quad \int_{-\beta_{21}-2\beta_{22}}^{+\infty} \left(\Phi \left(-\frac{\beta_{11} + 2\beta_{12} + \rho y}{\sqrt{1-\rho^2}} \right) \right) \phi\left(y\right)\ dy. \tag{1.13}$$

As previously we can invert the three equations and get $\beta_{21}^*(\rho) = \Psi_0(\rho, P_{010})$, $(\beta_{21} + \beta_{22})^*(\rho) = \Psi_1(\rho, P_{011})$ and $(\beta_{21} + 2\beta_{22})^*(\rho) = \Psi_2(\rho, P_{012})$. By solving this simple system we get the following equation in ρ: $g(\rho) = \Psi_1(\rho, P_{011}) - \frac{1}{2}\Psi_2(\rho, P_{012}) - \frac{1}{2}\Psi_0(\rho, P_{010}) = 0$. Since $g(\rho)$ is not necessarily monotone we might have multiple solutions and then set identification.

2. Introducing an exclusion variable

The major insight from Example 1 is that all parameters are identified given the correlation parameter. Point identification requirement translates then in the existence of restrictions that will pin down the value of ρ. An exclusion restriction on a binary variable, will provide just the right restriction. Applied to Example 1, the main idea of the proof is that the exclusion restriction will provide two set of values $P(Y_1 = 0, Y_2 = 1|z)$ and $P(Y_1 = 0, Y_2 = 1|z')$ related respectively to two functions $\beta_2^z(\rho)$ and $\beta_2^{z'}(\rho)$. In the Proposition 3 below, we prove that they are single-crossing. ρ is therefore uniquely identified at the crossing point of the two functions (see Fig 2).



(a) (ρ, β_{22}) region, $\beta_{12} = 0.4$

(b) (ρ, δ) region, $\beta_{12} = -0.6$

FIGURE 3. Numerical results: $\beta_2^z(\rho)$ and $\beta_2^{z'}(\rho)$.

$\beta_{11}^0 = 0.3, \beta_2^0 = 0.4, \beta_{22}^0 = 0.0, \delta^0 = 0.3, \rho_0 = 0.5$. In Fig 3(a) (Fig 3(b)) $\beta_2^z(\rho)$ the lowest slope (highest slope) and $\beta_2^{z'}(\rho)$ the highest slope (lowest slope) are increasing and intersect each other at a single point.

In (0.1) and (0.2), let $x_1^T = [1, Z]$, $x_2^T = 1$ and $\beta_1 = [\beta_{11}, \beta_{12}]^T$ the associated parameters where $Z \in \{0, 1\}$, a binary regressor. To avoid confusion with the previous setup, let define:

$$\begin{aligned} P_{ijz} \equiv & \quad P\left(Y_1 = i, Y_2 = j|z = 0\right) \text{ for all } i, j \in \{0,1\}^2 \\ P_{lkz'} \equiv & \quad P\left(Y_1 = l, Y_2 = k|z = 1\right) \text{ for all } l, k \in \{0,1\}^2. \end{aligned}$$

Compared to the previous setup, we restrict here $\beta_{22} = 0$. With a slight abuse of notation, we wish to identify the parameters $(\beta_{11}, \beta_{12}, \beta_2, \delta, \rho)$. β_{12} is the coefficient pertaining to the excluded variable. Given this exclusion restriction, consider now the following two equations (compare with Eq. (1.8) and Eq. (1.10)):

$$P_{01z} = \int_{-\beta_2}^{+\infty} \left(\Phi\left(-\frac{\beta_{11} + \rho y}{\sqrt{1-\rho^2}}\right)\right)\phi\left(y\right)\, dy \tag{2.1}$$

$$P_{01z'} = \int_{-\beta_2}^{+\infty} \left(\Phi\left(-\frac{\beta_{11} + \beta_{12} + \rho y}{\sqrt{1-\rho^2}}\right)\right)\phi\left(y\right)\, dy \tag{2.2}$$

Given $z = 0$ (resp. $z = 1$) we know by Proposition 1 that the set of solutions to Eq. (2.1) (resp. Eq. (2.2)) is characterized by a continuous function of ρ, call it $\beta_2^z(\rho)$ (resp. $\beta_2^{z'}(\rho)$).

We show first that $\beta_2^z(.)$ and $\beta_2^{z'}(.)$ are increasing functions and provide an expression of their derivatives.

Proposition 2. *Fix P_{01z} (resp. $P_{01z'}$). $\beta_2^z(\rho)$ (resp. $\beta_2^{z'}(\rho)$) is strictly increasing in ρ.*

The derivative of the function $\beta_2^z(\rho)$ with respect to ρ is proportional to the hazard rate of a univariate standard normal variable. Namely:

$$\frac{\partial \beta_2^z(\rho)}{\partial \rho} = \frac{1}{\sqrt{1-\rho^2}} r\left(\frac{\beta_{11} - \rho\beta_2^z(\rho)}{\sqrt{1-\rho^2}}\right) \tag{2.3}$$

$$\frac{\partial \beta_2^{z'}(\rho)}{\partial \rho} = \frac{1}{\sqrt{1-\rho^2}} r\left(\frac{\beta_{11} + \beta_{12} - \rho\beta_2^{z'}(\rho)}{\sqrt{1-\rho^2}}\right) \tag{2.4}$$

where $r(t)$ is the hazard rate of a standard normal at the point t.

Proof. $\beta_2^z(\rho)$ is implicitly defined by Eq. (2.1) that we rewrite in the form

$$f_0(\beta_2^z, \rho) = 0. \tag{2.5}$$

Using the implicit function theorem, we have that the derivative of $\beta_2^z(\rho)$ is given by the following expression

$$\frac{\partial \beta_2^z(\rho)}{\partial \rho} = -\frac{\partial f_0(\beta_2, \rho)}{\partial \rho} / \frac{\partial f_0(\beta_2, \rho)}{\partial \beta_2},$$

where we use the argument β_2 instead of $\beta_2^z(\rho)$ to lighten up the notations. Both terms of this ratio are positive. Indeed,

$$\frac{\partial f_0(\beta_2, \rho)}{\partial \beta_2} = \Phi\left(-\frac{\beta_{11} - \rho\beta_2}{\sqrt{1-\rho^2}}\right)\phi(-\beta_2) > 0.$$

Note that :

$$\int_{-\beta_2}^{+\infty}\left(\Phi\left(-\frac{\beta_{11}+\rho y}{\sqrt{1-\rho^2}}\right)\right)\phi(y)\ dy = (P_{000}+P_{010}) - \int_{-\infty}^{-\beta_2}\left(\Phi\left(-\frac{\beta_{11}+\rho y}{\sqrt{1-\rho^2}}\right)\right)\phi(y)\ dy$$

Using the result of Tallis (1962), we have:

$$\left(-\frac{\partial f_0(\beta_2, \rho)}{\partial \rho}\right) = \phi_2(-\beta_2, -\beta_{11}; \rho) > 0.$$

It follows that:

$$\begin{aligned}\frac{\partial \beta_2^z(\rho)}{\partial \rho} &= \frac{\phi_2(-\beta_2, -\beta_{11}; \rho)}{\Phi(-\frac{\beta_{11}-\rho\beta_2}{\sqrt{1-\rho^2}})\phi(-\beta_2)} \\ &= \frac{1}{\sqrt{1-\rho^2}} r\left(\frac{\beta_{11}-\rho\beta_2}{\sqrt{1-\rho^2}}\right),\end{aligned}$$

The proof for the derivative of $\beta_2^{z'}$ follows by replacing β_{11} by $\beta_{11} + \beta_{12}$. $\square$

We have now all the elements in hand to state the main result of this section.

Proposition 3. *There exists only one scalar* $\rho_0 \in (-1, 1)$ *such that* $\beta_2^z(\rho_0) = \beta_2^{z'}(\rho_0) = \beta_2^0$.

Proof. Suppose $(\beta_{11}^0, \beta_{12}^0, \beta_2^0, \delta^0, \rho_0)$ are the true parameters we wish to identify in Model (0.1)-(0.2), then we must have $\beta_2^z(\rho_0) = \beta_2^{z'}(\rho_0) = \beta_2^0$. Again, $\beta_{11}^0, \beta_{12}^0$ are readily identified. By definition of the model, the two functions will cross at the values of the true parameters (β_2^0, ρ_0). Using Eq. (2.3) we have:

$$\frac{\partial \beta_2^z(\rho_0)}{\partial \rho} = \frac{1}{\sqrt{1-\rho_0^2}} r\left(\frac{\beta_{11}^0 - \rho_0 \beta_2^0}{\sqrt{1-\rho_0^2}}\right) \tag{2.6}$$

and

$$\frac{\partial \beta_2^{z'}(\rho_0)}{\partial \rho} = \frac{1}{\sqrt{1-\rho_0^2}} r\left(\frac{\beta_{11}^0 + \beta_{12}^0 - \rho_0 \beta_2^0}{\sqrt{1-\rho_0^2}}\right). \tag{2.7}$$

Depending on the sign of β_{12}^0, it has to be that either:

$$\frac{\partial \beta_2^{z'}(\rho_0)}{\partial \rho} > \frac{\partial \beta_2^z(\rho_0)}{\partial \rho} \text{ or } \frac{\partial \beta_2^{z'}(\rho_0)}{\partial \rho} < \frac{\partial \beta_2^z(\rho_0)}{\partial \rho},$$

since the hazard rate of a univariate normal distribution is strictly increasing. Assume now that $\beta_{12}^0 > 0$, so that the first of these two inequalities holds. A necessary condition for existence of a second crossing point (β_2^1, ρ_1) is that:

$$\frac{\partial \beta_2^{z'}(\rho_1)}{\partial \rho} \leq \frac{\partial \beta_2^z(\rho_1)}{\partial \rho}$$

or equivalently

$$\frac{\beta_{11}^0 + \beta_{12}^0 - \rho_1 \beta_2^1}{\sqrt{1-\rho_1^2}} \leq \frac{\beta_{11}^0 - \rho_1 \beta_2^1}{\sqrt{1-\rho_1^2}}$$

This however implies that $\beta_{12}^0 \leq 0$. Hence a contradiction. The same type of contradiction arises when considering the other inequality. □

This result emphasizes the importance of both the validity and the relevance of the instrument. Since, the above argument fails in the case where $\beta_{12} = 0$ or $\beta_{22} \neq 0$.

3. Conclusion

We discussed identification in two equations probit model with endogenous dummy regressor. We contradict the identification criterion proposed by Wilde (2000), and argue that adding a regressor with enough variation allows to shrink the identified set, and may permit point identification in some cases, but in general, additional restrictions should complement the full rank condition. Therefore, we reinforce the opinion of the necessity of an exclusion restriction to ensure point identification in this model.

References

Heckman, J. J. (1978): "Dummy endogenous variables in a simultaneous equation system," .

Maddala, G. S. (1983): *Limited-dependent and qualitative variables in econometrics*, vol. 3. Cambridge University Press.

Sartori, A. E. (2003): "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions," *Political Analysis*, 11(2), 111–138.

Tallis, G. (1962): "The maximum likelihood estimation of correlation from contingency tables," *Biometrics*, 18(3), 342–353.

Wilde, J. (2000): "Identification of multiple equation probit models with endogenous dummy regressors," *Economics letters*, 69(3), 309–312.