

# THE EQUIVALENCE OF THREE LATENT CLASS MODELS AND ML ESTIMATORS

[Vidhura S. Tennekoon](#)

Department of Economics, Indiana University Purdue University Indianapolis (IUPUI), School of Liberal Arts, Cavanaugh Hall Room 524, 425 University Boulevard, Indianapolis, IN 46202, USA

## Highlights

- *We establish the equivalence of three latent class models.*
- *They are the binary Roy model, the probit model with a misclassified dependent variable and a trivariate probit model with partial observability.*
- *The probit model with measurement error is an enhanced version of existing models.*
- *A researcher working on one of these estimators may benefit from the literature and software related to others.*

## 1. Introduction

The seminal paper published by Andrew D. Roy in 1951 motivated the idea of self-selectivity, which was later developed by several others (see Maddala, 1983). Several econometric models including the switching regression model of Goldfeld and Qunadt (1973) and the switching regression model with endogenous switching discussed in Maddala (1983) fall to the board category of models now referred as Roy models. The switching regression model with endogenous switching is a hierarchal model in which there are two potential states each with an outcome that has a continuous distribution (log wage, for example) while the assignment to one of these two states is determined by a latent variable. When all three outcomes are specified as linear index functions and the errors are assumed to be potentially correlated and distributed normally, the parameters of this model can be estimated using popular

statistical packages (Lokshin and Sajaia, 2004). In binary Roy models, the two potential outcomes also are determined by latent variables as in Heckman and Vytlačil (1999).

Poirier (1980) presented and discussed a bivariate probit model with partial observability in which only two of the four potential outcomes are observed. He showed that the usual parameters of the bivariate probit model can also be identified with partial observability under certain conditions. Abowd and Farber (1982) discussed a variant of this partial observability model. These versions of the bivariate probit model are supported in popular statistical packages. Poirier (2014) extends this analysis to multivariate probit models with partial observability and *multivariate pairwise partial observability*. According to Poirier (2014), which cites applications from eleven different fields, there have been over hundred applications of these models. That paper presents the conditions for identifying multivariate probit models with partial observability and provides an example of a trivariate probit model with partial observability.

A third class of models is the probit model with misclassified dependent variables as discussed in Hausman et al. (1998). Lewbel (2000) showed that the parameters of this model can be identified even when the misclassification probabilities depend on one or more covariates. Tennekoon and Rosenman (2016) presented a model to identify the parameters of this model under parametric assumptions. The applications of this model includes Murphy et al. (2015) and Tennekoon and Rosenman (2015). These models, however, do not consider the potential correlation between the error terms. We enhance these existing measurement error models here by introducing potentially correlated errors.

The purpose of this letter is to show the equivalence of the three latent class models discussed above, the switching regression model with endogenous switching and a latent outcome (the binary Roy model), the probit model with a systematically misclassified dependent variable and a trivariate model with partial observability, under potentially correlated trivariate normal errors. Establishing this connection, we hope, will help a researcher working on one of these classes of estimators to benefit from the literature and

software related to other families. In addition, the enhanced version of the probit model with a mismeasured dependent variable we present here has not been used elsewhere.

## 2. The Three Models

### 2.1 Binary Roy model

The switching regression model with endogenous switching, widely known as the Roy model (Roy, 1951), assigns a given individual to one of two potential outcome regions using an endogenous switching mechanism. The special case that we discuss here is the binary Roy model in which the two potential outcomes also are observed as binary variables (Heckman and Vytlačil, 1999).

For each individual  $i$ , we assume two potential outcomes  $(Y_{2i}, Y_{3i})$  which corresponds to treated and untreated states, respectively. Unlike in the prototypical Roy model,  $Y_{2i}$  and  $Y_{3i}$  are not continuous variables here. They are binary indicator variables generated by the latent variables  $Y_{2i}^*$  and  $Y_{3i}^*$ , respectively. The indicator variable  $Y_{1i}$  determines the receipt or non-receipt of treatment.  $Y_{1i}$  is generated by another latent variable  $Y_{1i}^*$ .  $Y_{1i}$ ,  $Y_{2i}$  and  $Y_{3i}$  are generated as,  $Y_{ji} = 1$ . ( $Y_{ji}^* > 0$ ) and  $Y_{ji}^* = \beta_j X_{ji} + \varepsilon_{ji}$  for  $j = 1, 2, 3$ . The error terms are assumed to be jointly trivariate normally distributed with potentially correlated errors as,

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \right).$$

Following the potential outcome model of Rubin (1978), we can write,  $Y_i = (1 - Y_{1i}) \cdot Y_{2i} + Y_{1i} \cdot Y_{3i}$  where  $Y_i$  is the observed outcome of individual  $i$ . Since  $E[Y_{ji}] = \Pr[Y_{ji} = 1]$  for  $j = 2, 3$ ,

$$E[Y_i] = \Pr[Y_i = 1] = \Pr[Y_{1i} = 0 \& Y_{2i} = 1] + \Pr[Y_{1i} = 1 \& Y_{3i} = 1]$$

Therefore,

$$(1) E[Y_i | X_{1i}, X_{2i}, X_{3i}] = Pr[Y = 1_i | X_{1i}, X_{2i}, X_{3i}] = \Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) + \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13})$$

## 2.2. Binary choice model with misclassification

The binary indicator variable is misclassified when some of the true ‘1’s are recorded as ‘0’s and vice versa. Hausman et al (1998) show that the usual parameters of the binary choice model can be identified consistently together with the two types of misclassification probabilities using MLE if the dependent variable is misclassified randomly. They assume a latent relationship that generates the true indicator variable,  $Y_{1i} = 1$ . ( $Y_{1i}^* > 0$ ) where  $Y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i}$ ,  $\varepsilon_{1i} \sim N(0,1)$  and two types of (constant) misclassification probabilities,  $\alpha_0 = Pr[(Y_i = 1)|(Y_{1i} = 0)]$  and  $\alpha_1 = Pr[(Y_i = 0)|(Y_{1i} = 1)]$  that generates the observed indicator variable,  $Y_i$ . As Hausman and colleagues show,

$$(2) E[Y_i | X_{1i}] = Pr[Y = 1_i | X_{1i}] = \alpha_0 + (1 - \alpha_0 - \alpha_1)\Phi(\beta_1 X_{1i}).$$

If we slightly change the notation of Hausman et al (1998) without changing its structure by defining  $\alpha_0 = Pr[(Y_i = 1)|(Y_{1i} = 0)]$  and  $\alpha_1 = Pr[(Y_i = 1)|(Y_{1i} = 1)]$ ,

$$(3) E[Y_i | X_{1i}] = Pr[Y = 1 | X_{1i}] = \alpha_0 + (\alpha_1 - \alpha_0)\Phi(\beta_1 X_{1i}).$$

When two types of misclassification probabilities are covariant-dependent as in Lewbel (2000) and Tennekoon and Rosenman (2016), using a normal link function we can write,

$$\begin{aligned} (4) E[Y_i | X_{1i}, X_{2i}, X_{3i}] &= Pr[Y = 1_i | X_{1i}, X_{2i}, X_{3i}] = \Phi(\beta_2 X_{2i}) + (\Phi(\beta_3 X_{3i}) - \Phi(\beta_2 X_{2i}))\Phi(\beta_1 X_{1i}) \\ &= \Phi(\beta_2 X_{2i}) + \Phi(\beta_3 X_{3i})\Phi(\beta_1 X_{1i}) - \Phi(\beta_2 X_{2i})\Phi(\beta_1 X_{1i}) \\ &= \Phi(\beta_2 X_{2i})\Phi(-\beta_1 X_{1i}) + \Phi(\beta_1 X_{1i})\Phi(\beta_3 X_{3i}) \end{aligned}$$

The two misclassification probabilities in this model can be thought to be generated by a latent process where  $\alpha_0 = \Pr[(Y_i = 1)|(Y_{1i} = 0)] = \Pr(\beta_2 X_{2i} + \varepsilon_{2i} > 0)$  and  $\alpha_1 = \Pr[(Y_i = 1)|(Y_{1i} = 1)] = \Pr(\beta_3 X_{3i} + \varepsilon_{3i} > 0)$ . Allowing the two error terms  $\varepsilon_{2i}$  and  $\varepsilon_{3i}$  to be correlated,

$$(5) E[Y_i | X_{1i}, X_{2i}, X_{3i}] = \Pr[Y = 1_i | X_{1i}, X_{2i}, X_{3i}] = \Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) + \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13})$$

Note that the Lewbel (2000) and Tennekoon and Rosenman (2016) models do not assume correlated errors as in above model. Each bivariate CDF term in (5) breaks to the product of two univariate terms in Tennekoon and Rosenman model. Here, we have enhanced that model allowing for the potential correlation between the error terms.

### 2.3 Trivariate probit models with partial observability

The bivariate probit model with full observability can be specified as,  $Y_{ji} = 1 \cdot (Y_{ji}^* > 0)$  and  $Y_{ji}^* = \beta_j X_{ji} + \varepsilon_{ji}$  for  $j = 1, 2$ . The error terms are assumed to be jointly bivariate normally distributed with potentially correlated errors as,  $\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \rho_{12} & 1 \\ 1 & \rho_{12} \end{bmatrix}\right)$ . There are four potential outcomes given by  $(Y_{1i}, Y_{2i}) = (0, 0), (0, 1), (1, 0), (1, 1)$ .

When one or more of above outcomes are not observed we have a bivariate probit model with partial observability. When the outcome  $(Y_{1i}, Y_{2i}) = (0, 1)$  is not observed (or not possible), we have the bivariate selection model. When both outcomes  $(Y_{1i}, Y_{2i}) = (0, 1)$  and  $(Y_{1i}, Y_{2i}) = (1, 0)$  are not observed, we have the partial observability model discussed in Poirior (1980). In a similar manner, we can define trivariate probit models with partial observability.

The triavariate probit model with full observability can be specified as,  $Y_{ji} = 1 \cdot (Y_{ji}^* > 0)$  and  $Y_{ji}^* = \beta_j X_{ji} + \varepsilon_{ji}$  for  $j = 1, 2, 3$ . The error terms are assumed to be jointly trivariate normally distributed with

potentially correlated errors as,  $\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}\right)$ . There are eight potential outcomes as

given in Table 1. As before, when one or more of above outcomes are not observed we have a trivariate probit model with partial observability. In a trivariate selection model, only four out of these eight potential outcomes (1, 5, 7 and 8) are observed.

**Table 1: Trivariate probit potential outcomes**

Outcome	$(Y_{1i}, Y_{2i}, Y_{3i})$	$Y_i$
1	(0,0,0)	0
2	(0,0,1)	0
3	(0,1,0)	1
4	(0,1,1)	1
5	(1,0,0)	0
6	(1,0,1)	1
7	(1,1,0)	0
8	(1,1,1)	1

The partial observability model we consider here can be viewed as a univariate mapping of a trivariate probit model which only includes two potential outcomes as in the standard probit model. The relationship between the potential trivariate outcomes and the observed binary variable is given in Table 1. Note that when  $Y_{1i} = 0$ ,  $Y_i$  doesn't depend on  $Y_{3i}$ . Similarly, when  $Y_{1i} = 1$ ,  $Y_i$  doesn't depend on  $Y_{2i}$ . Therefore, we can write,  $Y_i = (1 - Y_{1i})Y_{2i} + Y_{1i} Y_{3i}$  and

$$(6) E[Y_i | X_{1i}, X_{2i}, X_{3i}] = \Pr[Y_i = 1 | X_{1i}, X_{2i}, X_{3i}] = \Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) + \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13})$$

## 2.4 Estimation and Identification

The equations (1), (5) and (6) are identical. Therefore, all three models discussed above are observationally equivalent. The parameters of this model can be estimated by maximizing the following

log-likelihood function. Note that the correlation coefficient between  $X_{2i}$  and  $X_{3i}$ ,  $\rho_{23}$ , does not appear in the log-likelihood function and therefore is never identified.

$$(7) \mathcal{L}(\beta_1, \beta_2, \beta_3, \rho_{12}, \rho_{13}, X_{1i}, X_{2i}, X_{3i}) = \frac{1}{n} \sum_{i=1}^n (Y_i \ln(\Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) + \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13})) + (1 - Y_i) \ln(\Phi_2(-\beta_1 X_{1i}, -\beta_2 X_{2i}, \rho_{12}) + \Phi_2(\beta_1 X_{1i}, -\beta_3 X_{3i}, -\rho_{13})))$$

The first order necessary conditions and the information matrix of (4) can be expressed as below.

$$(8) \mathbf{g} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \beta_1} \\ \frac{\partial \mathcal{L}}{\partial \beta_2} \\ \frac{\partial \mathcal{L}}{\partial \beta_3} \\ \frac{\partial \mathcal{L}}{\partial \rho_{12}} \\ \frac{\partial \mathcal{L}}{\partial \rho_{13}} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{Y_i}{P_i} - \frac{(1-Y_i)}{(1-P_i)} \right] C_i = 0$$

$$(9) \mathbf{I} = -E \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \beta_2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \beta_3} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \rho_{13}} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \beta_2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \beta_3} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \rho_{13}} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \beta_3} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \beta_3} & \frac{\partial^2 \mathcal{L}}{\partial \beta_3^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_3 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_3 \rho_{13}} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_3 \rho_{12}} & \frac{\partial^2 \mathcal{L}}{\partial \rho_{12}^2} & \frac{\partial^2 \mathcal{L}}{\partial \rho_{12} \rho_{13}} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \rho_{13}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \rho_{13}} & \frac{\partial^2 \mathcal{L}}{\partial \beta_3 \rho_{13}} & \frac{\partial^2 \mathcal{L}}{\partial \rho_{12} \rho_{13}} & \frac{\partial^2 \mathcal{L}}{\partial \rho_{13}^2} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{P_i(1-P_i)} \right] C_i C_i'$$

where  $P_i = \Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) + \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13})$  and

$$C_i = \begin{bmatrix} \left\{ \Phi \left( \frac{\beta_3 X_{3i} - \rho_{13} \beta_1 X_{1i}}{\sqrt{1-(\rho_{13})^2}} \right) - \Phi \left( \frac{\beta_2 X_{2i} - \rho_{12} \beta_1 X_{1i}}{\sqrt{1-(\rho_{12})^2}} \right) \right\} \Phi(\beta_1 X_{1i}) X_{1i} \\ \Phi \left( \frac{-\beta_1 X_{1i} + \rho_{12} \beta_2 X_{2i}}{\sqrt{1-(\rho_{12})^2}} \right) \Phi(\beta_2 X_{2i}) X_{2i} \\ \Phi \left( \frac{\beta_1 X_{1i} - \rho_{13} \beta_3 X_{3i}}{\sqrt{1-(\rho_{13})^2}} \right) \Phi(\beta_3 X_{3i}) X_{3i} \\ -\Phi_2(-\beta_1 X_{1i}, \beta_2 X_{2i}, -\rho_{12}) \\ \Phi_2(\beta_1 X_{1i}, \beta_3 X_{3i}, \rho_{13}) \end{bmatrix}.$$

If the matrix  $C_i C_i'$  is non-singular, the parameters are uniquely identified under MLE regulatory conditions. When  $X_{2i} = X_{3i}$  the two parameter vectors  $[\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\rho}_{12}, \hat{\rho}_{13}]$  and  $[-\hat{\beta}_1, \hat{\beta}_3, \hat{\beta}_2, -\hat{\rho}_{12}, -\hat{\rho}_{13}]$  cannot be identified separately. Therefore, we need  $X_{2i} \neq X_{3i}$  to identify the parameter vector uniquely.

However, we do not require the conditions  $X_{1i} \neq X_{2i}$  or  $X_{1i} \neq X_{3i}$  for identifying the parameters of this model. Additional technical details are presented in Poirier (2014).

### 3. Conclusions

In this letter, we established the equivalence of three maximum likelihood estimators that belong to different classes. The switching regression models and the multivariate probit models with partial observability have been used in a wide range of applications for several decades. Probit models with potentially misclassified dependent variables, in contrast, are relatively new and used infrequently. Moreover, this third class of models is not yet supported in popular statistical packages to our knowledge. As we have shown here, the three models are algebraically equivalent even though the underlying data generating processes are different. Therefore, any application software developed to estimate one model can also be used to estimate the parameters of other models.

The model we presented here is a relatively generalized one which nests several other restricted cases. For example, the models with uncorrelated errors including Goldfeld and Qunadt (1973), Abowd and Farber (1982) and Tennekoon and Rosenman (2016), models with one-sided misclassification, bivariate model with partial observability first proposed in Poirier (1980) are all nested in this model. However, this equivalence can be easily extended to more complex higher order multivariate probit models if needed.

### Acknowledgments

The author thanks Zhuang Hao and an anonymous referee for helpful comments.

### References

- Abowd, JM, and Farber HS. (1982). Job queues and the union status of workers. *Industrial and Labor Relations Review* 35: 354–367.
- Goldfeld, S, and Quandt R. (1973). "The estimation of structural shifts by switching regressions." *Annals of Economic and Social Measurement* 2(4): 475-485.



- Hausman JA, Abrevaya J, Scott-Morton FM. (1998). "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87: 239–269.
- Heckman, JJ and Vytlacil EJ. (1999). "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the national Academy of Sciences*, 96(8): 4730-4734.
- Lewbel A. (2000). "Identification of the Binary Choice Model with Misclassification." *Econometric Theory* 16(4): 603–609.
- Lokshin, M and Sajaia Z. (2004). "Maximum likelihood estimation of endogenous switching regression models." *Stata Journal* (2004): 282-289.
- Maddala, G S (1983). *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press)
- Murphy, S.M., Friesner, D., & Rosenman, R. (2015). "Opioid misuse among adolescents: New evidence from a misclassification analysis." *Applied Health Economics and Health Policy* 13(2): 181-192.  
doi:10.1007/s40258-015-0151-z.
- Poirier, DJ. (1980). "Partial observability in bivariate probit models." *Journal of Econometrics* 12: 209–217.
- Poirier, DJ. (2014). "Identification in multivariate partial observability probit." *International Journal of Mathematical Modelling and Numerical Optimisation* 5.1-2 : 45-63.
- Roy, A. D. (1951). "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2): 135-146.
- Rubin, DB. (1978) "Bayesian inference for causal effects: The role of randomization." *The Annals of statistics* 6 (1): 34-58.
- Tennekoon, V., & Rosenman, R. (2015). "The Pot Calling the Kettle Black? A Comparison of Measures of Current Tobacco Use." *Applied Economics* 47(5): 431-448.  
DOI: <http://dx.doi.org/10.1080/00036846.2014.972546>.
- Tennekoon, V., & Rosenman, R. (2016). "Systematically Misclassified Binary Dependent Variables." *Communications in Statistics - Theory and Methods* DOI:10.1080/03610926.2014.887105.