

# Properties of Estimators of Count Data Model with Endogenous Switching

K. Oya

Graduate School of Economics, Osaka University, Machikaneyama, Toyonaka, 560-0043, Japan.  
(kosuke@econ.osaka-u.ac.jp)

**Abstract:** We examine properties of estimators of count data model with endogenous switching. The estimation of the count data model that accommodates endogenous switching can be accomplished by Full Information Maximum Likelihood (FIML). However, FIML estimation requires fully and correctly specified model and is computationally burdensome. Alternative estimation methods do not require fully specified model have been proposed. The typical methods are Two Stage Method of moments (TSM) and Nonlinear Weighted Least Squares (NWLS). The properties of these estimators have never been studied so far. In this paper, we compared the finite sample properties of these estimators under correct and incorrect model specifications using Monte Carlo experiments. We find that FIML estimator has the smallest standard deviation and TSM estimator has the largest when the model is correctly specified. This property also holds under incorrect model specification. An important point to emphasize is the large standard deviation of the estimate of endogeneity suggests that TSM and NWLS estimations result in an anomalous estimate of endogeneity very frequent even under the correct specification.

**Keywords:** *Count data; endogenous switching; Full maximum information; Two stage method of moment; Nonlinear weighted least squares; Monte Carlo experiment*

## 1. INTRODUCTION

Count data modeling has been widely used in empirical analysis. General surveys of count data analysis are given in Cameron and Trivedi (1998) and Winkelmann (2000). The standard model for count data is based on a Poisson regression model. Several extensions of the model have been proposed to deal with overdispersion and sample selectivity. Terza (1998) proposed a count data model with sample selection. The proposed estimation methods are Full Information Maximum Likelihood (FIML), Two Stage Method of moment (TSM) and Nonlinear Weighted Least Squares (NWLS). On theoretical point of view, the FIML estimator is most efficient among three estimators when the model is correctly specified. However, count data often exhibits non-Poisson features. The properties of these estimators have never been studied so far. In this paper we examine these estimators under correct and incorrect model specifications by Monte Carlo experiments.

The outline of this paper is as follows. Section 2 details the count data model. In section 3, we summarize the estimation of the count data model. In section 4, some Monte Carlo experiments are conducted to compare the estimators. Section 5 contains concluding remarks.

## 2. MODEL

### 2.1. Poisson regression model

A Poisson regression model with heterogeneity is used for count data that exhibits non-Poisson feature such as overdispersion. The probability distribution function of count data  $y_i$  given a  $(k \times 1)$  vector of explanatory variable  $x_i$  is

$$f(y_i | x_i, \varepsilon_i) = \Pr(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad (1)$$

where  $\lambda_i$  is a conditional mean function

$$\lambda_i = E[Y_i | x_i, \varepsilon_i] = \exp\{x_i' \alpha + \varepsilon_i\}. \quad (2)$$

The error term  $\varepsilon_i$  represents the heterogeneity and  $\alpha$  is a  $(k \times 1)$  vector of unknown parameter to be estimated. (1) and (2) are jointly define the Poisson regression model with heterogeneity.

### 2.2. Poisson model with heterogeneity and endogenous switching

Terza (1998) proposed the count model with endogenous switching. The model consists of a conditional Poisson probability distribution and a switching variable  $c_i$ . The conditional probability distribution function of  $y_i$  is

$$f(y_i | x_i, \varepsilon_i, c_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad i = 1, \dots, n \quad (3)$$

and the conditional mean function is

$$\lambda_i = \exp\{x_i' \alpha + c_i \beta + \varepsilon_i\} \quad (4)$$

where  $\beta$  is an unknown parameter. The switching variable  $c_i$  is characterized as

$$c_i = \begin{cases} 1, & \text{if } c_i^* = z_i' \gamma + v_i \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where  $c_i^*$  is latent random variable and  $v_i$  is an error term.  $z_i$  and  $\gamma$  are a  $(l \times 1)$  vector of observable explanatory variables and their unknown coefficient vector, respectively. The potential endogeneity of  $c_i$  is represented using a correlation coefficient  $\rho$  between two error terms  $\varepsilon_i$  and  $v_i$ . The joint distribution of  $\varepsilon_i$  and  $v_i$  is assumed to be normal with mean vector zero and variance-covariance matrix

$$\begin{pmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{pmatrix}. \quad (6)$$

The variance of  $v_i$  is normalized to one since the switching equation (5) identifies  $\gamma$  only up to a scale factor. Under the joint normality assumption, we have the conditional joint probability distribution function of  $y_i$  and  $c_i$  as follow

$$f(y_i, c_i | x_i, z_i) = \int_{-\infty}^{\infty} f(y_i | x_i, z_i, c_i, \varepsilon_i) \times \{c_i \Phi^*(\varepsilon_i) + (1 - c_i)(1 - \Phi^*(\varepsilon_i))\} \phi(\varepsilon_i) d\varepsilon_i \quad (7)$$

where  $\Phi^*(\varepsilon_i) = \Phi(\{z_i' \gamma + (\rho/\sigma)\varepsilon_i\} / \sqrt{1 - \rho^2})$ ,  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the probability density function and the cumulative distribution function of standard normal distribution, respectively.

To obtain a maximum likelihood estimator of unknown parameters in the model, we have to maximize a likelihood function. Unfortunately the joint probability density function (7) cannot evaluate in closed form. A typical procedure to deal with such difficulty is approximation of (7) using Gauss-Hermite quadrature. The details of the procedure are given in Winkelmann (2000). After change of variable  $\zeta_i = \varepsilon_i / (\sqrt{2}\sigma)$ , we obtain

$$f(y_i, c_i | x_i, z_i) \approx \frac{1}{\sqrt{\pi}} \sum_{h=1}^H w_h \frac{\lambda_i(u_h)^{y_i} e^{-\lambda_i(u_h)}}{y_i!}$$

$$\times \{c_i \Phi_i^{**}(u_h) + (1 - c_i)(1 - \Phi_i^{**}(u_h))\} \quad (8)$$

where  $w_h$  and  $u_h$  are the weight and the evaluation point of Gauss-Hermite approximation,

$$\lambda_i(u_h) = \exp\{x_i' \alpha + c_i \beta + \sqrt{2}\sigma u_h\} \quad (9)$$

and

$$\Phi_i^{**}(u_h) = \Phi\left(\frac{z_i' \gamma + \sqrt{2}\rho u_h}{\sqrt{1 - \rho^2}}\right). \quad (10)$$

### 3. ESTIMATION

In this section, we summarize Full Information Maximum Likelihood (FIML), Two Stage Method of moments (TSM) and Nonlinear Weighted Least Squares (NWLS) estimators proposed in Terza (1998).

#### 3.1. FIML

FIML estimation is based on the likelihood, which is defined by the approximated joint probability function (8). We can derive the first and second partial derivatives of unknown parameters in (8). The maximization of the likelihood function is conducted by numerical optimization procedure.

#### 3.2. TSM

For TSM estimation, we use a moment condition

$$E[y_i | x_i, z_i, c_i, \varepsilon_i] = \exp\{x_i' \alpha + c_i \beta + \varepsilon_i\} \quad (11)$$

which is much weaker assumption of distribution of the error term in previous section. Based on (11), we can represent

$$y_i = \exp\{x_i' \alpha + c_i \beta + \varepsilon_i\} + \eta_i \quad (12)$$

where  $\eta_i$  is a random variable such that

$$E[\eta_i | x_i, z_i, c_i, \varepsilon_i] = 0. \quad (13)$$

The conditional moment of  $y_i$  is given as

$$\begin{aligned} E[y_i | x_i, z_i, c_i] &= E_\varepsilon[E[y_i | x_i, z_i, c_i, \varepsilon_i]] \\ &= E_\varepsilon[\exp\{x_i' \alpha + c_i \beta + \varepsilon_i\}] \\ &= \exp\{x_i' \alpha + c_i \beta\} E[\exp(\varepsilon_i) | x_i, z_i, c_i]. \end{aligned} \quad (14)$$

After tedious calculations, we obtain

$$E[y_i | x_i, z_i, c_i] = \exp\{J_i^* \theta^*\} \psi_i(\tau, \gamma) \quad (15)$$

where  $J_i^* = (1 \quad x_i' \quad c_i)$ ,  $\theta^* = (0.5\sigma^2 \quad \alpha' \quad \beta)'$ ,  $\tau = \rho\sigma$  and

$$\psi_i(\tau, \gamma) = c_i \frac{\Phi(\tau + z_i' \gamma)}{\Phi(z_i' \gamma)} + (1 - c_i) \frac{1 - \Phi(\tau + z_i' \gamma)}{1 - \Phi(z_i' \gamma)}.$$

(12) is rewritten using (15) as

$$y_i = \exp\{J_i^{*'} \theta^*\} \psi_i(\tau, \gamma) + e_i \quad (16)$$

where

$$e_i = \exp\{J_i^{*'} \theta^* + \varepsilon_i\} - \exp\{J_i^{*'} \theta^*\} \psi_i(\tau, \gamma) + \eta_i.$$

The estimation of unknown parameters in (16) can be conducted by two-stage technique. The unknown parameter  $\gamma$  in the switching equation (5) can be estimated by a simple probit estimation method. We define the probit estimator of  $\gamma$  as  $\hat{\gamma}$  in the first-stage. The second-stage estimator of  $(\theta^* \ \tau)'$  can be obtained by applying nonlinear least squares to

$$y_i = \exp\{J_i^{*'} \theta^*\} \psi_i(\tau, \hat{\gamma}) + e_i^0 \quad (17)$$

where  $e_i^0 = e_i + \exp\{J_i^{*'} \theta^*\} [\psi_i(\tau, \gamma) - \psi_i(\tau, \hat{\gamma})]$ .

The estimator of  $\rho$  is given by  $\hat{\rho} = \hat{\tau} / \hat{\sigma}$ . See Terza (1998) for details.

### 3.3. NWLS

Terza (1998) gave an exact form of conditional variance of the nonlinear least squares error term when (3) and (4) are correct specification. The conditional variance is

$$v_i = \text{var}(e_i | x_i, z_i, c_i) = \exp\{J_i^{*'} \theta^*\} \psi_i(\tau, \gamma) + (\exp\{J_i^{*'} \theta^*\})^2 \{\exp\{\sigma^2 - 2\tau^2\} \psi_i^* - \psi_i^2(\tau, \gamma)\}$$

where  $\psi_i^* = \exp\{2\tau^2\} \psi_i(2\tau, \gamma)$ .

Substituting TSM estimators into the unknown parameters in the conditional variance  $v_i$ , we obtain  $\hat{v}_i$ . NWLS estimator can be given by minimizing

$$\sum_{i=1}^n \left( \frac{y_i - \exp\{J_i^{*'} \theta^*\} \psi_i(\tau, \hat{\gamma})}{\sqrt{\hat{v}_i}} \right)^2 \quad (18)$$

with respect to  $(\theta^* \ \tau)'$ .

## 4. MONTE CARLO EXPERIMENTS

We examine the properties of FIML, TSM and NWLS estimators given in the previous section using Monte Carlo experiments. We also examine whether these estimators lead to substantially different results and whether these estimators are robust with respect to the choice of a particular distributional assumption.

### 4.1. Design

A collection of count dependent variable  $\{y_i\}_i^n$  is generated from a conditional Poisson distribution

$$f(y_i | x_i, \varepsilon_i, c_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (19)$$

and a conditional mean function

$$\lambda_i = \exp\{x_i' \alpha + c_i \beta + \varepsilon_i\}.$$

The switching variable  $c_i$  is characterized as

$$c_i = \begin{cases} 1, & \text{if } c_i^* = \gamma_1 + z_i \gamma_2 + v_i \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where parameters in the model set to be  $(\alpha, \beta, \gamma_1, \gamma_2) = (1.0, 1.0, 1.0, 2.0)$ . The variance-covariance matrix of the error terms  $\varepsilon_i$  and  $v_i$  is set as

$$\begin{pmatrix} 0.3^2 & 0.3\rho \\ 0.3\rho & 1 \end{pmatrix}$$

and the correlation coefficient  $\rho = 0.4$ . The explanatory variable  $x_i$  is generated from the uniform distribution over the interval  $[-0.5, 0.5]$  and  $z_i$  is generated from the standard normal distribution. The sample size  $n$  is set to be 500. We conduct the experiments 1000 times to see the finite sample properties of estimators.

According to the error terms  $\varepsilon_i$  and  $v_i$ , we consider the two cases. In case 1, the joint distribution  $\varepsilon_i$  and  $v_i$  is set to be normal with zero mean and the variance-covariance matrix described above. This is a correct specification. On the other hand, case 2 deals with the misspecification. We assume that the distribution of  $v_i$  is standard normal, however that of  $\varepsilon_i$  is no-normal distribution. The distribution of  $\exp\{\varepsilon_i\}$  is assumed to be Gamma.

### 4.2. Bivariate distribution

In the case 2, we have to generate two series of random variables with different distribution and correlation coefficient  $\rho$ . Lee (1983) proposed that the bivariate distribution with different marginal distributions and correlation  $\rho$ . Suppose the marginal distributions of  $\varepsilon_i$  and  $v_i$  are  $F_1(\varepsilon)$  and  $F_2(v)$ . The bivariate distribution is given by

$$H(\varepsilon, v; \rho) = B(\Phi^{-1}(F_1(\varepsilon)), \Phi^{-1}(F_2(v)); \rho)$$

where  $B(\cdot, \cdot; \rho)$  is the bivariate cumulative normal distribution function with mean zero, variance one, and correlation  $\rho$ . First, we generate random variables  $\varepsilon_i$  and  $\nu_i$  from their marginal distribution. Using two independent normal random variables  $\varepsilon^* = \Phi^{-1}(F_1(\varepsilon))$  and  $\nu^* = \Phi^{-1}(F_2(\nu))$ , we obtain bivariate normal random variables with correlation  $\rho$ . Finally, we obtain two random variables  $F_1^{-1}(\Phi(\varepsilon^*))$  and  $F_2^{-1}(\Phi(\nu^*))$ . The correlation coefficient of these random variables is  $\rho$ . The random number generation that described above is applied for the case 2.

### 4.3. Experiment results

The results of the simulations for the case 1 summarized in Table 1. In Table 1, we see that the biases and standard deviations of estimates by FIML are the smallest and those of TSM are the biggest. The properties of estimates of  $\alpha$ ,  $\beta$ ,  $\gamma_1$  and  $\gamma_2$  in the mean function and the switching equation and the estimates of the standard deviation  $\sigma$  of error term in mean function are almost same among FIML, TSM and NWLS. It is noted that the biases and standard deviations of TSM and NWLS estimates of the correlation coefficient  $\rho$  are quite large. The case of misspecification is summarized in Table 2. Result is similar to the case 1 even though the distribution supposed for model is not the same as true data distribution. One possible explanation for this tendency is that the shape of distribution of simulated data is not much different from one supposed for model specification.

**Table 1:** Means and standard deviations (in parentheses) of estimates for case 1.

	true value	FIML	TSM	NWLS
$\alpha$	1.0	1.001 (0.113)	0.998 (0.122)	0.994 (0.113)
$\beta$	1.0	1.003 (0.048)	0.983 (0.086)	0.983 (0.082)
$\gamma_1$	1.0	1.012 (0.107)	1.012 (0.107)	1.012 (0.107)
$\gamma_2$	2.0	2.019 (0.177)	2.021 (0.177)	2.021 (0.177)
$\sigma$	0.3	0.272 (0.084)	0.279 (0.215)	0.288 (0.210)
$\rho$	0.4	0.439 (0.423)	0.937 (1.416)	1.031 (1.373)

**Table 2:** Means and standard deviations (in parentheses) of estimates for case 2.

	true value	FIML	TSM	NWLS
$\alpha$	1.0	1.003 (0.114)	1.000 (0.125)	0.996 (0.114)
$\beta$	1.0	1.015 (0.051)	0.991 (0.097)	0.989 (0.091)
$\gamma_1$	1.0	1.006 (0.108)	1.007 (0.108)	1.007 (0.108)
$\gamma_2$	2.0	2.021 (0.176)	2.021 (0.176)	2.021 (0.176)
$\sigma$	0.3	0.306 (0.082)	0.317 (0.221)	0.325 (0.214)
$\rho$	0.4	0.425 (0.370)	0.794 (1.303)	0.870 (1.268)

It is noted that the estimates of  $\alpha$  and  $\beta$  in the mean function, and those of  $\gamma_1$  and  $\gamma_2$  in the switching equation are estimated well even though TSM and NWLS estimates of  $\rho$  are biased in both case 1 and case 2. Table 3 and 4 show relative standard deviations of estimator with respect to those of FIML. We see that the estimates of FIML are efficient even when the specification is not correct

**Table 3:** Relative standard deviations of estimates for case 1.

	FIML	TSM	NWLS
$\alpha$	1.00	1.08	1.00
$\beta$	1.00	1.79	1.71
$\gamma_1$	1.00	1.00	1.00
$\gamma_2$	1.00	1.00	1.00
$\sigma$	1.00	2.56	2.50
$\rho$	1.00	3.35	3.25

**Table 4:** Relative standard deviations of estimates for case 2.

	FIML	TSM	NWLS
$\alpha$	1.00	1.10	1.00
$\beta$	1.00	1.90	1.78
$\gamma_1$	1.00	1.00	1.00
$\gamma_2$	1.00	1.00	1.00
$\sigma$	1.00	2.70	2.61
$\rho$	1.00	3.52	3.43

## 5. CONCLUSIONS

This paper has examined the properties of estimators of the count data model with endogenous switching proposed in Terza (1998). The results of Monte Carlo experiments show that FIML estimator has the smallest standard deviation and TSM estimator has the largest when the model specification is correct. This property also holds under incorrect model specification. An important point to emphasize is the large standard deviation of the estimate of endogeneity  $\rho$  suggests that TSM and NWLS estimations result in an anomalous estimate of endogeneity very frequent even under correct specification.

## 6. ACKNOWLEDGEMENTS

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology: Grants-in-Aid for Scientific Research (C)(2), 13630031, 2002.

## 7. REFERENCES

- Cameron, A. C. and P. K. Trivedi, *Regression analysis of count data*, Cambridge University Press, 1998.
- Kozumi, H., A bayesian analysis of endogenous switching models for count data, *Journal of the Japan Statistical Society*, 32(2), 141-154, 2002.
- Lee, L. F., Generalize econometric models with selectivity, *Econometrica*, 51(2), 507-512, 1983.
- Terza, J.V., Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects, *Journal of Econometrics*, 84, 129-154, 1998.
- Winkelmann, R., Count data models with selectivity, *Econometric Reviews*, 17(4), 339-359, 1998.
- Winkelmann, R., *Econometric analysis of count data*, 3rd edition, Springer, 2000.