

Sample Selection Models with a Common Dummy Endogenous Regressor in Simultaneous Equations: A Simple Two-Step Estimation

Kyoo il Kim*

Department of Economics, UCLA

October 2005

Abstract

This note studies sample selection models where a common dummy endogenous regressor appears both in the selection equation and in the censored equation. We interpret this model as an endogenous switching model and develop a simple two step estimation procedure.

Keywords: Dummy Endogenous Regressor, Sample Selection, Endogenous Switching, Two-step Estimation.

JEL classification: C34, C35

1 Introduction

The limited dependent variable models such as discrete choice models and censored or truncated models have been studied for several decades. In particular, the binary (dummy) endogenous regressors problem in the limited dependent variable models has been of interest (see Heckman (1978), Angrist (2001), and Vytlacil (2002)). In the bivariate binary variables situation, imposing the joint normality assumption, Heckman (1978) refers the model to “multivariate probit model with structural shift” and develops an ML estimator. Some nonparametric extensions of this dummy endogenous variable in nonseparable models are considered in Vytlacil (2002). In a general framework of simultaneous equations models with censored endogenous regressors where the censored equation contains dummy endogenous variables, Vella (1993) introduces a simple

*I am truly grateful to Jinyong Hahn for his advice and encouragement throughout my doctoral study in UCLA. Correspondence: Kyoo il Kim, Department of Economics, UCLA, Box 951477, Los Angeles, CA 90095-1477; Tel. +1 310 430 2051; Fax. +1 310 825 9528; E-mail: kyookim@ucla.edu.

estimator using a “control function” approach and generalized residuals. Though sample selection models have been studied intensively with many variations, however, the situation where the selection equation and the censored equation contain a common endogenous dummy variable has not been dealt with yet.

Among field studies, a common binary endogenous regressor in the selection equation and the censored equation has been typically concerned in managed health care studies (from Manning et al (1984) and Welch (1984) to many recent studies), which examine the effects of managed care on health care expenditures. The managed care status such as the enrollment in a health maintenance organization (HMO) affects a person or family’s decisions on both whether or not to use a health care at all and how much to spend for a health care. Interestingly here we need to handle three simultaneous equations, where two of them contain a common endogenous binary regressor. This makes the problem intractable and hence other variations of this model have been used such as two-part models (see Duan et al (1982)) and discrete factor models (DFM) (see Goldman (1995) and Goldman et al (1998)). However, these models are too restrictive to represent the real data and can be biased or inefficient (see Kim (2005)). In labor economics, a common dummy endogenous variable is often observed in many structural studies, for example, women labor supply with endogenous participation, where the childbearing decision affects both how much to work and whether or not to participate into the labor force. Though econometric theories have been developed toward nonseparable and nonparametric (or semiparametric) models for dummy endogenous variables, unbiased and efficient estimation are still concerned for these parametric models in applied studies. In this note, we consider the model named “Endogenous Switching Type II-Tobit”, which is a hybrid of Heckman (1978)’s “multivariate probit model with structural shift” and a type II-tobit model and provide a simple two step estimator which is easy to implement and robust compared to other alternative estimators.

2 Sample Selection Model with a Common Endogenous Dummy Variable for Selection and Censored Equation

Here we are interested in the following sample selection model that contains a common endogenous dummy variable in both the selection equation and the censored equation

$$y_{1i}^* = Z'_{1i}\gamma_1 + \varepsilon_{1i}, y_{1i} = I_{y_{1i}^* > 0} \quad (1)$$

$$y_{2i}^* = Z'_{2i}\gamma_2 + y_{1i}\beta_2 + \varepsilon_{2i}, y_{2i} = I_{y_{2i}^* > 0} \quad (2)$$

$$y_{3i}^* = Z'_{3i}\gamma_3 + y_{1i}\beta_3 + \varepsilon_{3i}, y_{3i} = y_{2i}y_{3i}^* \text{ for } i = 1, \dots, n \quad (3)$$

where Z_{ji} is a $k_j \times 1$ vector of exogenous regressors, γ_j is a $k_j \times 1$ vector of parameters, β_2 and β_3 are scalar, I_A or $I(A)$ is an indicator function which has value one when A is true, $\varepsilon_i \equiv (\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i})'$ follow a multivariate normal distribution as $\varepsilon \sim N(\mathbf{0}, \Sigma)$ with $vech(\Sigma) = (1, \rho_{12}, \rho_{13}\sigma, 1, \rho_{23}\sigma, \sigma^2)'$. We assume $(y_{1i}^*, Z'_{1i}, y_{2i}^*, Z'_{2i}, y_{3i}^*, Z'_{3i})'$ are independently and identically distributed. In the model (1)-(3), we only observe the signs of (y_{1i}^*, y_{2i}^*) and observe the quantity of y_{3i}^* only when $y_{2i}^* > 0$. Therefore, this model is a hybrid of the bivariate probit and the type-II tobit model but it contains an interesting feature that (2) and

(3) contain a common endogenous binary variable y_{1i} . It implies we need to control the endogeneity problem caused by y_{1i} and the selection bias caused by the censoring indicator y_{2i} at the same time. It complicates the handling of this model.

2.1 Endogenous Switching Type II-Tobit (ESTT)

As a generalization of Heckman(1978), using a switching model we can achieve the efficient estimation for the full model of (1)-(3) under the normality assumption. For this, we extend the model (2)-(3) into two regimes respectively determined by the state of $I_{y_1^* > 0}$

$$y_{20}^* = Z_2' \gamma_2 + \varepsilon_{20}, y_{21}^* = Z_2' \gamma_2 + \beta_2 + \varepsilon_{21} \quad (4)$$

$$y_{30}^* = Z_3' \gamma_3 + \varepsilon_{30}, y_{31}^* = Z_3' \gamma_3 + \beta_3 + \varepsilon_{31} \quad (5)$$

and define accordingly $y_{20} = 1(y_{20}^*)$, $y_{21} = 1(y_{21}^*)$, $y_{30} = y_{20} y_{30}^*$, and $y_{31} = y_{21} y_{31}^*$ and hence we have $y_2 = y_1 y_{21} + (1 - y_1) y_{20}$ and $y_3 = y_1 y_{31} + (1 - y_1) y_{30}$. We could allow γ_2 and γ_3 to depend on the value of y_1 but this extension is trivial. We observe y_2 and y_3 , not the pairs of (y_{20}, y_{21}) or (y_{30}, y_{31}) but the missing data problem here is the fundamental feature of this model. Now we assume $(\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}, \varepsilon_{30}, \varepsilon_{31})$ to be jointly normally distributed with zero means with a covariance matrix

$$\Omega = \begin{pmatrix} 1 & \rho_{1,20} & \rho_{1,21} & \rho_{1,30}\sigma_{30} & \rho_{1,31}\sigma_{31} \\ & 1 & \rho_{20,21} & \rho_{20,30}\sigma_{30} & \rho_{20,31}\sigma_{31} \\ & & 1 & \rho_{21,30}\sigma_{30} & \rho_{21,31}\sigma_{31} \\ & & & \sigma_{30}^2 & \rho_{30,31}\sigma_{30}\sigma_{31} \\ & & & & \sigma_{31}^2 \end{pmatrix}. \quad (6)$$

This generalized model of (1)-(3) allows $(y_{20}, y_{21})=(1, 0)$ even when $\beta_2 > 0$ ($(y_{20}, y_{21})=(0, 1)$ when $\beta_2 < 0$) and $y_{30} > y_{31}$ even when $\beta_2 > 0$ ($y_{30} < y_{31}$ when $\beta_2 < 0$), which are relevant in many applications. If we set $\varepsilon_{20} = \varepsilon_{21}$ or $\varepsilon_{30} = \varepsilon_{31}$, those cases cannot be nested in the model. It is also noticeable that none of $\rho_{20,21}$, $\rho_{20,31}$, $\rho_{21,30}$, or $\rho_{30,31}$ are identified, since none of pairs of y_{j0} and y_{k1} , $j, k \in \{2, 3\}$ are observed at the same time. Also note that this switching model nests the model (1)-(3) with $\rho_{20,21} = 1$ and $\rho_{30,31} = 1$. This model can be estimated by ML or alternatively by the method of simulated likelihood using GHK simulator. A detailed discussion of those approaches and their limitations can be found in Kim (2005).

3 Simple Two-step Estimation

Instead of the ML estimation, we can also estimate the parameters of interest using a two step estimation procedure as an extension of Heckman's classical two step estimation to multivariate selection problems. Heckman (1979) corrects the bias caused by the sample selection using the "control function" approach, namely, the inverse Mill's ratio. Here we are dealing with two selection problems. One is the endogenous switching and the other one is the sample selection. In this perspective, we can interpret the ESTT model

as a Type V-Tobit model with bivariate selections where there exist restrictions on parameters. Here we derive the correction terms which comprises two parts where one corrects the bias due to the endogenous switching and the other corrects the sample selection bias for each state depending on y_1 and then adding these two correction terms for each state, we obtain consistent estimates for the structural equation using corresponding subsamples.

To define these correction terms, we first consider the following conditional mean

$$E[y_{3i}|Z_i, y_{1i}, y_{2i}^* > 0] = Z'_{3i}\gamma_3 + y_{1i}\beta_3 + y_{1i}E[\varepsilon_{31i}|Z_i, y_{1i}^* > 0, y_{21i}^* > 0] + (1 - y_{1i})E[\varepsilon_{30i}|Z_i, y_{1i}^* < 0, y_{20i}^* > 0]$$

where $Z_i = (Z'_{1i}, Z'_{2i}, Z'_{3i})'$ and hence we can obtain consistent estimates of γ_3 and β_3 by adding these two correction terms in the censored equation and estimate them over the subsample with $y_{2i} = 1$. To implement this procedure, we need to derive $E[\varepsilon_{31i}|Z_i, y_{21i}^* > 0, y_{1i}^* > 0]$ and $E[\varepsilon_{30i}|Z_i, y_{20i}^* > 0, y_{1i}^* < 0]$. Now we denote by $\Phi(a, b; r)$ a standardized bivariate normal cdf with mean zero and correlation r evaluated at (a, b) . We also let $\Phi(\cdot)(\phi(\cdot))$ denote the standard normal cdf (pdf). Similarly with Poirier (1980), we obtain

$$E[\varepsilon_{31i}|Z_i, y_{1i}^* > 0, y_{21i}^* > 0] = \rho_{1,31}\sigma_{31} \left[\frac{\phi(Z'_{1i}\gamma_1)\Phi((Z'_{2i}\gamma_2 + \beta_2 - \rho_{1,21}Z'_{1i}\gamma_1)/\sqrt{1 - \rho_{1,21}^2})}{\Phi(Z'_{1i}\gamma_1, Z'_{2i}\gamma_2 + \beta_2; \rho_{1,21})} \right] \quad (7)$$

$$+ \rho_{21,31}\sigma_{31} \left[\frac{\phi(Z'_{2i}\gamma_2 + \beta_2)\Phi((Z'_{1i}\gamma_1 - \rho_{1,21}(Z'_{2i}\gamma_2 + \beta_2))/\sqrt{1 - \rho_{1,21}^2})}{\Phi(Z'_{1i}\gamma_1, Z'_{2i}\gamma_2 + \beta_2; \rho_{1,21})} \right];$$

$$E[\varepsilon_{30i}|Z_i, y_{1i}^* < 0, y_{20i}^* > 0] = -\rho_{1,30}\sigma_{30} \left[\frac{\phi(Z'_{1i}\gamma_1)\Phi((Z'_{2i}\gamma_2 - \rho_{1,20}Z'_{1i}\gamma_1)/\sqrt{1 - \rho_{1,20}^2})}{\Phi(-Z'_{1i}\gamma_1, Z'_{2i}\gamma_2; -\rho_{1,20})} \right] \quad (8)$$

$$+ \rho_{20,30}\sigma_{30} \left[\frac{\phi(Z'_{2i}\gamma_2)\Phi((-Z'_{1i}\gamma_1 + \rho_{1,20}Z'_{2i}\gamma_2)/\sqrt{1 - \rho_{1,20}^2})}{\Phi(-Z'_{1i}\gamma_1, Z'_{2i}\gamma_2; -\rho_{1,20})} \right].$$

Define these correction terms as $C_{11i}(\alpha) \equiv \frac{\phi(Z'_{1i}\gamma_1)\Phi((Z'_{2i}\gamma_2 + \beta_2 - \rho_{1,21}Z'_{1i}\gamma_1)/\sqrt{1 - \rho_{1,21}^2})}{\Phi(Z'_{1i}\gamma_1, Z'_{2i}\gamma_2 + \beta_2; \rho_{1,21})}$,

$$C_{12i}(\alpha) \equiv \frac{\phi(Z'_{2i}\gamma_2 + \beta_2)\Phi((Z'_{1i}\gamma_1 - \rho_{1,21}(Z'_{2i}\gamma_2 + \beta_2))/\sqrt{1 - \rho_{1,21}^2})}{\Phi(Z'_{1i}\gamma_1, Z'_{2i}\gamma_2 + \beta_2; \rho_{1,21})}, C_{01i}(\alpha) = -\frac{\phi(Z'_{1i}\gamma_1)\Phi((Z'_{2i}\gamma_2 - \rho_{1,20}Z'_{1i}\gamma_1)/\sqrt{1 - \rho_{1,20}^2})}{\Phi(-Z'_{1i}\gamma_1, Z'_{2i}\gamma_2; -\rho_{1,20})},$$

and $C_{02i}(\alpha) = \frac{\phi(Z'_{2i}\gamma_2)\Phi((-Z'_{1i}\gamma_1 + \rho_{1,20}Z'_{2i}\gamma_2)/\sqrt{1 - \rho_{1,20}^2})}{\Phi(-Z'_{1i}\gamma_1, Z'_{2i}\gamma_2; -\rho_{1,20})}$ to simplify the notation.

Noting these correction terms are infeasible unless $\alpha \equiv (\gamma_1, \gamma_2, \beta_2, \rho_{1,20}, \rho_{1,21})'$ is known, as the first step, we estimate the selection equations

$$y_{1i} = 1(Z'_{1i}\gamma_1 + \varepsilon_{1i} > 0), y_{2i} = 1(Z'_{2i}\gamma_2 + y_{1i}\beta_2 + \varepsilon_{2i} > 0)$$

following Heckman (1978). He interprets this model as a bivariate probit with structural shift as in our framework and hence he estimates

$$y_{1i} = 1(Z'_{1i}\gamma_1 + \varepsilon_{1i} > 0), y_{20i} = 1(Z'_{2i}\gamma_2 + \varepsilon_{20i} > 0), \text{ and } y_{21i} = 1(Z'_{2i}\gamma_2 + \beta_2 + \varepsilon_{21i} > 0)$$

using the ML estimation. Following his work, we define the probabilities of four possible events as

$$\begin{aligned}
P_{00}(i) &\equiv \Pr(y_{1i} = 0, y_{2i} = 0) = \Phi(-Z'_{1i}\gamma_1, -Z'_{2i}\gamma_2; \rho_{1,20}) \\
P_{01}(i) &\equiv \Pr(y_{1i} = 0, y_{2i} = 1) = \Phi(-Z'_{1i}\gamma_1, Z'_{2i}\gamma_2; -\rho_{1,20}) \\
P_{10}(i) &\equiv \Pr(y_{1i} = 1, y_{2i} = 0) = \Phi(Z'_{1i}\gamma_1, -Z'_{2i}\gamma_2 - \beta_2; -\rho_{1,21}) \\
P_{11}(i) &\equiv \Pr(y_{1i} = 1, y_{2i} = 1) = \Phi(Z'_{1i}\gamma_1, Z'_{2i}\gamma_2 + \beta_2; \rho_{1,21}).
\end{aligned}$$

Based on these probabilities, we estimate the parameters of interest by maximizing

$$\ln L(\gamma_1, \gamma_2, \beta_2, \rho_{1,20}, \rho_{1,21}) = \ln \left(\prod_{i=1}^n [P_{00}(i)]^{(1-y_{1i})(1-y_{2i})} [P_{01}(i)]^{(1-y_{1i})y_{2i}} [P_{10}(i)]^{y_{1i}(1-y_{2i})} [P_{11}(i)]^{y_{1i}y_{2i}} \right) \quad (9)$$

Denote $\hat{\alpha} \equiv (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\beta}_2, \hat{\rho}_{1,20}, \hat{\rho}_{1,21})'$ to be the ML estimates of (9). Then as the second step, we estimate the following equation using the subsample with $y_{2i} = 1$ by OLS

$$y_{3i} = Z'_{3i}\gamma_3 + y_{1i}\beta_3 + \mu_{11}y_{1i}C_{11i}(\hat{\alpha}) + \mu_{12}y_{1i}C_{12i}(\hat{\alpha}) + \mu_{01}(1 - y_{1i})C_{01i}(\hat{\alpha}) + \mu_{02}(1 - y_{1i})C_{02i}(\hat{\alpha}) + \eta_i. \quad (10)$$

It is worthwhile to consider which parameters we can identify from this two step estimation. First, the parameters of interest, $\{\gamma_1, \gamma_2, \gamma_3, \beta_2, \beta_3\}$ are obtained from (10) and the first stage bivariate probit estimation of (9). For the auxiliary parameters, we obtain the estimates of $\rho_{1,20}$ and $\rho_{1,21}$ from the first stage estimation. From the estimated coefficients of the correction terms, we can recover other auxiliary parameters noting that $\text{Cov}[\varepsilon_1, \varepsilon_{30}] = \rho_{1,30}\sigma_{30} = \mu_{01}$, $\text{Cov}[\varepsilon_1, \varepsilon_{31}] = \rho_{1,31}\sigma_{31} = \mu_{11}$, $\text{Cov}[\varepsilon_{20}, \varepsilon_{30}] = \rho_{20,30}\sigma_{30} = \mu_{02}$, and $\text{Cov}[\varepsilon_{21}, \varepsilon_{31}] = \rho_{21,31}\sigma_{31} = \mu_{12}$. Having consistent estimators of σ_{30} and σ_{31} , we can also consistently estimate $\rho_{1,30}$, $\rho_{1,31}$, $\rho_{20,30}$, and $\rho_{21,31}$ from these covariance estimators. Consistent estimators for σ_{30} and σ_{31} can be obtained similarly with Heckman (1979). Therefore, we can recover all the auxiliary parameters.

Though these two step estimators are less efficient, it has several advantages over the efficient ML estimation. First, it is easy to implement and numerically robust. More interestingly, this approach can relax the strong normality assumption. We can construct other versions of inverse Mills' ratio for certain other distributions (possibly asymmetric ones). Semiparametric extension is also possible, which is considered in Kim (2005). However, this gain does not come without other costs. We need to correct standard errors of the second step estimators reflecting that we use pre-estimated parameters in the second step. Otherwise we may exaggerate the significances of the estimators we obtain. We could derive the adjusted asymptotic variance matrix of the two step estimator similarly with Heckman (1979) or Lee et. al. (1980) as an explicit form but here we adopt the approach suggested by Newey (1984) whose usefulness is revisited in Lewbel (2004). By formulating the two step estimation into one gigantic method of moments estimation (MM), we can easily obtain the asymptotic variance matrix of the second step estimators.

3.1 Asymptotic Distribution of Two Step Estimator

To derive the asymptotic variances of the coefficient parameters in (10), we rewrite the first stage and the second stage estimation as an MM estimation based on the following moment conditions

$$g(Z_i, \alpha) = \frac{\partial \ln L_i(\alpha)}{\partial \alpha}, \quad E[g(Z_i, \alpha)] = 0 \text{ and}$$

$$h(Z_i, \alpha, \theta) = y_{2i}W_i(\alpha) (y_{3i} - W_i(\alpha)'\theta), \quad E[h(Z_i, \alpha, \theta)] = 0$$

where $\ln L_i(\alpha)$ is a single observation likelihood of (9), $W_i(\alpha) \equiv (Z'_{3i}, y_{1i}, C_{11i}(\alpha), C_{12i}(\alpha), C_{01i}(\alpha), C_{02i}(\alpha))'$, and $\theta \equiv (\gamma'_3, \beta_3, \mu_{11}, \mu_{12}, \mu_{01}, \mu_{02})'$. The population moment conditions hold by construction. The two step estimators are obtained by solving

$$\frac{1}{n} \sum_{i=1}^n g(Z_i, \hat{\alpha}) = 0 \text{ and } \frac{1}{n} \sum_{i=1}^n h(Z_i, \hat{\alpha}, \hat{\theta}) = 0.$$

The asymptotic variance of $\hat{\alpha}$ coincides with the asymptotic variance obtained in the ML estimation of (9) by construction. We can also derive the asymptotic variance of $\hat{\theta}$ following Newey (1984) as

$$\text{AVAR}(\hat{\theta}) \equiv \Xi = H_{\theta}^{-1}V_{hh}H_{\theta}^{-1'} + H_{\theta}^{-1}H_{\alpha}[G_{\alpha}^{-1}V_{gg}G_{\alpha}^{-1'}]H'_{\alpha}H_{\theta}^{-1'} - H_{\theta}^{-1}[H_{\alpha}G_{\alpha}^{-1}V_{gh} + V_{hg}G_{\alpha}^{-1'}H'_{\alpha}]H_{\theta}^{-1'}$$

where $H_{\theta} = E \left[\frac{\partial h(Z_i, \alpha, \theta)}{\partial \theta'} \right]$, $H_{\alpha} = E \left[\frac{\partial h(Z_i, \alpha, \theta)}{\partial \alpha'} \right]$, $G_{\alpha} = E \left[\frac{\partial g(Z_i, \alpha)}{\partial \alpha'} \right]$, $V_{gg} = E [g(Z_i, \alpha)g(Z_i, \alpha)']$, $V_{gh} = E [g(Z_i, \alpha)h(Z_i, \alpha, \theta)']$, $V_{hg} = V'_{gh}$, and $V_{hh} = E [h(Z_i, \alpha, \theta)h(Z_i, \alpha, \theta)']$. This asymptotic variance can be estimated consistently with

$$\hat{\Xi} \equiv \hat{H}_{\theta}^{-1}\hat{V}_{hh}\hat{H}_{\theta}^{-1'} + \hat{H}_{\theta}^{-1}\hat{H}_{\alpha}[\hat{G}_{\alpha}^{-1}\hat{V}_{gg}\hat{G}_{\alpha}^{-1'}]\hat{H}'_{\alpha}\hat{H}_{\theta}^{-1'} - \hat{H}_{\theta}^{-1}[\hat{H}_{\alpha}\hat{G}_{\alpha}^{-1}\hat{V}_{gh} + \hat{V}_{hg}\hat{G}_{\alpha}^{-1'}\hat{H}'_{\alpha}]\hat{H}_{\theta}^{-1'}$$

where $\hat{H}_{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial h(Z_i, \hat{\alpha}, \hat{\theta})}{\partial \theta'}$, $\hat{H}_{\alpha} = \frac{1}{n} \sum_{i=1}^n \frac{\partial h(Z_i, \hat{\alpha}, \hat{\theta})}{\partial \alpha'}$, $\hat{G}_{\alpha} = \frac{1}{n} \sum_{i=1}^n \frac{\partial g(Z_i, \hat{\alpha})}{\partial \alpha'}$, $\hat{V}_{gg} = \frac{1}{n} \sum_{i=1}^n g(Z_i, \hat{\alpha})g(Z_i, \hat{\alpha})'$, $\hat{V}_{gh} = \frac{1}{n} \sum_{i=1}^n g(Z_i, \hat{\alpha})h(Z_i, \hat{\alpha}, \hat{\theta})'$, $\hat{V}_{hg} = \hat{V}'_{gh}$, and $\hat{V}_{hh} = \frac{1}{n} \sum_{i=1}^n h(Z_i, \hat{\alpha}, \hat{\theta})h(Z_i, \hat{\alpha}, \hat{\theta})'$. All of these can be evaluated analytically or numerically in a straightforward manner. Using the information equality $G_{\alpha} = -V_{gg}$, we may simplify the asymptotic variance formula a bit.

4 Summary

This note studies a sample selection model where a common dummy endogenous regressor appears both in the selection equation and in the censored equation. This model is analyzed in the framework of an endogenous switching model. A simple two-step estimator is proposed for this model, which is easy to implement and numerically robust compared to other methods.

References

- [1] Angrist, J. (2001), "Estimation of Limited-Dependent Variable Models with Binary Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*.

- [2] Duan, N, Manning, W. G., Morris, C., and Newhouse, J. P. (1982), "A Comparison of Alternative Models of the Demand for Medical Care," R-2754-HHS, RAND Corporation.
- [3] Dana P. Goldman (1995), "Managed Care as a Public Cost-Containment Mechanism," *Rand Journal of Economics*, Vol. 26, 277-295.
- [4] Goldman D. P., A. Leibowitz, and J. L. Buchanan (1998), "Cost-Containment and Adverse Selection in Medicaid HMOs," *Journal of the American Statistical Association*, Vol. 93, No.441, 54-62.
- [5] Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, Vol. 46, No.4, 931-959.
- [6] Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.
- [7] Kim, K. (2005), "Sample Selection Models with a Common Dummy Endogenous Regressor in the Simultaneous Equations", working paper, UCLA.
- [8] Lee, L.F., G.S. Maddala, and R.P. Trost (1980), "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica*, 48, 491-503.
- [9] Lewbel, A. (2004), "Simple Estimators for Hard Problems: Endogeneity in Discrete Choice Related Models," working paper, Boston College.
- [10] Manning, W., Leibowitz, A., Goldberg, G., Rogers, W., and Newhouse, J. (1984) "A Controlled Trial of the Effect of a Prepaid Group Practice on the Use of Services," *New England Journal of Medicine*, V.310, 1501-1510.
- [11] Newey, W. K. (1984), "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14, 201-206.
- [12] Poirier, Dale (1980), "Partial Observability in Bivariate Probit Models," *Journal of Econometrics*, 12(2), 290-17.
- [13] Vella, F. (1993), "A Simple Estimator for Models with Censored Endogenous Regressors," *International Economic Review*, 34, 441-57.
- [14] Vytlacil, E. (2002), "Dummy Endogenous Variables in Nonseparable Models," working paper, Stanford University.
- [15] Welch, W. P., Frank, R. G., and Diehr, P. (1984) "Health Care Costs in Health Maintenance Organizations: Correcting for Self-Selection," In L.F. Rossiter and R. M. Sheffler, *Advances in Health Economics and Health Services Research*, V.5., Greenwich, Conn.: JAI Press.