E C O N O M I C S   B U L L E T I N

Semi-nonparametric count data estimation with an endogenous binary variable

Hiroaki Masuhara
Section of Health Economics, Department of Gerontological Policy, National Center for Geriatrics and Gerontology (NCGG)

Abstract

This paper proposes a semi-nonparametric Poisson model with an endogenous binary variable, which generalizes bivariate correlated unobserved heterogeneity using Hermite polynomials, and compares this model with a parametric one. The National Health Interview Survey (NHIS) data from 1990 shows the difference between the endogenous binary variable's coefficients of the semi-nonparametric and parametric models.

I would like to thank an anonymous referee and the Associate Editor Timothy Conley for their valuable comments and suggestions that have improved the paper. This study is supported by the Research Grant for Longevity Sciences (18C-7) from the Ministry of Health, Labor, and Welfare in Japan.All remaining errors are mine.
Citation: Masuhara, Hiroaki, (2008) "Semi-nonparametric count data estimation with an endogenous binary variable." *Economics Bulletin,* Vol. 3, No. 42 pp. 1-13
Submitted: June 25, 2008. **Accepted:** August 12, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C30069A.pdf

1. Introduction

Count data models explain the behavior of discrete and non-negative dependent random variables and are used in applied econometrics such as industrial organization, health economics, and population economics. A Poisson model is one of the methods to estimate count data. Moreover, many recent studies use a negative binomial 2 (NB2) model that assumes additive separable log-gamma distributed heterogeneity.

In microeconometric applications, we often come across situations where explanatory variables (in particular, an endogenous binary variable) are simultaneously determined with the dependent variable. In this case, the Poisson and NB2 models yield biased estimates of parameters of interest because these models assume perfect exogeneity of explanatory variables. Therefore, count data models with an endogenous binary variable are required, and many studies have been conducted to analyze this problem. For example, Terza (1998) proposes a nonlinear weighted least squares (NWLS) estimator; Mullahy (1997) and Windmeijer and Santos-Silva (1997) use Generalized Method of Moments (GMM) to estimate such a model; and Kenkel and Terza (2001) analyze the endogeneity bias using Box-Cox transformation.[1] Moreover, Romeu and Vera-Hernández (2005) develop another count data model with an endogenous binary variable on the basis of the polynomial Poisson model proposed by Cameron and Johannson (1997). The main feature of their model is that it comprises a semiparametric model using a polynomial expansion by a dependent variable. However, the binary endogenous variable part is parametric, and the dependent variable does not explicitly assume heterogeneity.

This paper proposes another semiparametric model to estimate a count data variable with an endogenous binary variable. This paper considers a simple Poisson model, which has one endogenous binary variable, and the heterogeneity of both count dependent and binary variables. In this model setup, we propose a Poisson model that comprises a semi-nonparametric joint distribution using Hermite polynomials based on the discussion of Gallant and Nychka (1987), Gabler *et al.* (1993), and van der Klaauw and Koning (2003). Our model is semiparametric and includes the natural extension of a bivariate normal distribution. That is, both the count dependent and endogenous binary variables explicitly assume semiparametric heterogeneity. We investigate the difference between the endogenous binary variable's coefficients of the parametric and semi-nonparametric models using the 1990

[1] From a Bayesian point of view, Kozumi (2002), Jochmann (2003), Munkin and Trivedi (2003), and Deb *et al.* (2006) analyze the endogeneity of count data.

National Health Interview Survey (NHIS) data employed by Kenkel and Terza (2001).

The rest of the paper is organized as follows. Section 2 proposes a semi-nonparametric count data model with an endogenous binary variable and discusses an efficient maximization algorithm that contains a numerical integral. Section 3 depicts the application of the NHIS data, and Section 4 presents our concluding remarks.

2. Poisson estimation with an endogenous binary variable

We consider a count data model with an endogenous binary variable proposed by Terza (1998). Let y_i, $i = 1, \ldots, N$, denote a count dependent variable that takes a nonnegative integer value; let x_i and z_i denote explanatory variables (covariates), where x_i is a $k_1 \times 1$ vector and z_i is a $k_2 \times 1$ vector. The marginal distribution of y_i takes the following form:

$$f\left(y_i \,|\, d_i, \varepsilon_{1i}\right) = \frac{\exp\left(-\lambda_i\right)\left(\lambda_i\right)^{y_i}}{y_i!}, \qquad \lambda_i = \exp\left(\beta_d d_i + x_i'\beta_1 + \varepsilon_{1i}\right), \tag{1}$$

where β_1 and β_d denote vectors of unknown parameters, and ε_{1i} is unobserved heterogeneity. Moreover, d_i represents an endogenous binary variable and is assumed to be generated by the process $d_i = 1$ if $d_i^* = z_i'\beta_2 + \varepsilon_{2i} \geq 0$ and $d_i = 0$; otherwise, where d_i^* is a latent variable, ε_{2i} is unobserved heterogeneity, and β_2 denotes a vector of parameters.

Many studies assume that the vector $(\varepsilon_{1i}, \varepsilon_{2i})$ follows a bivariate normal distribution with zero mean and covariance matrix $(\sigma^2, \rho\sigma, 1)$. In this assumption, the joint density is easily evaluated using a numerical integral. However, this normally distributed assumption leads to a specification problem. Under a linear-exponential mean specification assumption and a set of instruments, Mullahy (1997) shows that the GMM estimators have consistency. In the GMM, to improve the efficiency of the estimators, it is necessary to use higher order moment conditions. The NWLS proposed in Terza (1998), which requires some additional distributional assumptions, has the same properties. Therefore, we require an alterative robust method for this count data model with an endogenous binary regressor.

Semiparametric estimation of this model implies approximating an unknown error term using Hermite polynomials (Gallant and Nychka, 1987; Gabler *et al.*, 1993). Following van der Klaauw and Koning (2003), the joint distribution of ε_{1i} and ε_{2i} takes the following semi-nonparametric bivariate normal density:

$$g\left(\varepsilon_{1i},\varepsilon_{2i}\right)=\frac{1}{P}\left(\sum_{j=0}^{K}\sum_{k=0}^{K}\alpha_{jk}\varepsilon_{1i}^{j}\varepsilon_{2i}^{k}\right)^{2}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times\exp\left[-\frac{1}{2\left(1-\rho^2\right)}\left\{\left(\frac{\varepsilon_{1i}}{\sigma_1}\right)^{2}-2\rho\frac{\varepsilon_{1i}}{\sigma_1}\frac{\varepsilon_{2i}}{\sigma_2}+\left(\frac{\varepsilon_{2i}}{\sigma_2}\right)^{2}\right\}\right]\equiv\frac{g^*}{P},\qquad(2)$$

where $P=\iint_{-\infty}^{\infty}g^*d\varepsilon_{1i}\,d\varepsilon_{2i}$ ensures integration to 1 by scaling the density, σ_1 and ρ are standard deviation and correlation parameters, respectively, and α_{jk} is the parameter to be estimated. To identify the parameters, we set $\alpha_{00}=1$ and $\sigma_2=1$. When $\alpha_{jk}=0$ ($\forall j\geq 1$ and $\forall k\geq 1$), this density results in a bivariate normal distribution.

Hence, the log-likelihood function of a Poisson model with a semi-nonparametric bivariate normal density takes the following form:

$$\ln f_i=\left(1-d_i\right)\ln\left[\int_{-\infty}^{-z_i'\beta_2}\int_{-\infty}^{\infty}f\left(y_i\,|\,d_i,\varepsilon_{1i}\right)g\left(\varepsilon_{1i},\varepsilon_{2i}\right)d\varepsilon_{1i}\,d\varepsilon_{2i}\right]$$
$$+\,d_i\ln\left[\int_{-z_i'\beta_2}^{\infty}\int_{-\infty}^{\infty}f\left(y_i\,|\,d_i,\varepsilon_{1i}\right)g\left(\varepsilon_{1i},\varepsilon_{2i}\right)d\varepsilon_{1i}\,d\varepsilon_{2i}\right].\qquad(3)$$

Substituting Eqs.(1) and (2) into Eq.(3) yields the full information maximum likelihood (FIML) of the semi-nonparametric Poisson model with an endogenous dummy variable.[2] This model generalizes heterogeneity and contains the FIML model with a bivariate normal distribution as a special case.

The model in Eq.(3) includes double integrals and has no analytical solution. Fortunately, we simplify the double integrals to the following single integral:

$$\ln f_i=\ln\left[\int_{-\infty}^{\infty}f\left(y_i\,|\varepsilon_{1i}\right)\frac{G_2\left(\varepsilon_{1i}\right)}{P}g_1\left(\varepsilon_{1i}\right)d\varepsilon_{1i}\right],\qquad(4)$$

where g_1 is the probability density function of a normal distribution. The term G_2 contains Hermite series and depends only on ε_1, which takes the following form:

$$G_2\left(\varepsilon_{1i}\right)=\begin{cases}\int_{-\infty}^{-z_i'\beta_2}g_2\left(\varepsilon_{2i}|\varepsilon_{1i}\right)d\varepsilon_{2i} & \text{if}\quad d_i=0\\ \int_{-z_i'\beta_2}^{\infty}g_2\left(\varepsilon_{2i}|\varepsilon_{1i}\right)d\varepsilon_{2i} & \text{if}\quad d_i=1\end{cases}.\qquad(5)$$

[2]This model has another restriction of $E\left[\varepsilon_{1i}\right]=E\left[\varepsilon_{2i}\right]=0$ (location normalization). However, this restriction is cumbersome when $K\geq 2$. Following Melenberg and van Soest (1996), we use an alternative restriction, setting the constant terms equal to those in the parametric model.

After some algebraic computation, Eq.(5) has an analytical solution.[3]

Since Eq.(4) has a single integral over $[-\infty, \infty]$, the Gauss-Hermite quadrature method is applied to evaluate the log-likelihood. However, Rabe-Hesketh *et al.* (2002, 2005) demonstrate the results of Monte Carlo simulation and conclude that the log-likelihood function approximated by this method often has a sharp peak and is poorly approximated by a low-degree polynomial. Moreover, they propose the *adaptive* Gaussian quadrature based on importance sampling and the Bayesian Markov chain method.[4] Following Rabe-Hesketh *et al.* (2002, 2005), this paper applies the adaptive Gaussian quadrature to estimate the proposed model.[5]

Let the parameter vector of this density and the mean and variance of the posterior density be $\theta = [\beta_1', \beta_2', \sigma_1, \rho, \alpha_{jk}, \ldots]'$, μ_i, and τ_i, respectively. Recall that Eq.(4) can be rewritten as follows:

$$\ln f_i = \ln \left[\sum_{q=1}^{Q} \omega_q f\left(y_i \,|\theta, u_q\right) \frac{G_2\left(u_q|\theta\right)}{P} \frac{g_1\left(u_q|\theta\right)}{h\left(u_q|\mu, \tau\right)} \frac{1}{\sqrt{\pi}} \right] \equiv \ln \left[\sum_{q=1}^{Q} \omega_q f_i\left(\theta|u_q\right) \right],$$

where ω_q is the qth weight, u_q is the qth evaluation point of the Gauss-Hermite quadrature over $[-\infty, \infty]$, Q is the number of weights, and $h(\cdot)$ is the importance function of a normal distribution with mean μ_i and variance τ_i. Further, the adaptive Gaussian quadrature obtains the parameters as follows:

1) Set the initial parameters $\theta^{(t)}$, $\mu_i^{(t)}$, $\tau_i^{(t)}$, and $t \leftarrow 0$.
2) Calculate the following posterior density based on $\mu_{i,T-1}^{(t)}$ and $\tau_{i,T-1}^{(t)}$ until convergence:

$$\mu_{i,T}^{(t)} = \frac{\sum_{q=1}^{Q} \left(\mu_{i,T-1}^{(t)} + \sqrt{2}\tau_{i,T-1}^{(t)} u_q\right) \omega_q f_i\left(\theta^{(t)}|u_q\right)}{f_i\left(\theta^{(t)}\right)},$$

$$\tau_{i,T}^{(t)} = \sqrt{\frac{\sum_{q=1}^{Q} \left(\mu_{i,T-1}^{(t)} + \sqrt{2}\tau_{i,T-1}^{(t)} u_q\right)^2 \omega_q f_i\left(\theta^{(t)}|u_q\right)}{f_i\left(\theta^{(t)}\right)} - \left(\mu_{i,T}^{(t)}\right)^2},$$

[3] See Appendix for further detail.

[4] Using the adaptive Gaussian quadrature, Miranda and Rabe-Hesketh (2006) propose the stata program (ssm.ado) of the parametric binary, ordinary, and Poisson models with an endogenous binary variable.

[5] Prior to investigating the proposed model, we estimate a parametric Poisson model with an endogenous binary variable using both the Gauss-Hermite and adaptive Gaussian quadratures. The value of the log-likelihood under the latter is higher than that under the former. Moreover, the difference between the endogenous binary variable's coefficients is not negligible (See Table 2 and Footnote 7).

where $f_i\left(\theta^{(t)}\right) = \sum_{q=1}^{Q} \omega_q f_i\left(\theta^{(t)}|u_q\right)$ and T is the number of iterations in this step.

3) Maximize the log-likelihood function with respect to $\theta^{(t)}$ given $\mu_i^{(t)}$ and $\tau_i^{(t)}$.
4) Set $t \leftarrow t+1$. Repeat steps 2 to 3 until convergence.

3. An application to drinking behavior

We present the results of the simplified application of the proposed model using a subsample of 2,467 observations from the 1990 National Health Interview Survey (NHIS) data, originally employed by Kenkel and Terza (2001).[6] All observations comprise males and current drinkers with high blood pressure. The dependent variable is the number of alcoholic beverages consumed in the last two weeks (D). The mean of this variable is 14.70 (21% of the observations are zero observations), and the minimum and maximum values of this variable are 0 and 168, respectively. Moreover, 687 of the individuals have been advised by a physician to reduce drinking (ADVICE). The explanatory variables are as follows: monthly income (EDITINC), years of schooling (EDUC), a dummy for $30 < \text{age} \leq 40$ (AGE30), $40 < \text{age} \leq 50$ (AGE40), $50 < \text{age} \leq 60$ (AGE50), $60 < \text{age} \leq 70$ (AGE60), $\text{age} > 70$ (AGEGT70), black (BLACK), non-white and non-black (OTHER), married (MARRIED), widowed (WIDOW), divorced or separated (DIVSEP), employed (EMPLOYED), unemployed (UNEMPLOY), northeastern residents (NORTHE), midwestern residents (MIDWEST), south resident (SOUTH), medicare status (MEDICARE), public insurance status (MEDICAID), military insurance status (CHAMPUS), health insurance status (HLTHINS), regional source of care (REGMED), consulting the same doctor (DRI), limits on major daily activity (MAIORLIM), limits on some daily activity (SOMELIM), having diabetes (HVDIAB), having a heart condition (HHRTCOND), and having had stroke (HADSTROKE). The entire description of the variables and summary statistics can be found in Kenkel and Terza (2001).

Table 1 shows the estimated result of the selection equation and Table 2 shows that of the drinking equation of the parametric, $K=1$, and $K=2$ models. Since the semi-nonparametric models nest the parametric model as a special case, we apply the log-likelihood ratio (LR) test to select the best model. The test statistics of normality against the semi-nonparametric

[6]The data is downloadable from the Journal of Applied Econometrics Data Archive (http://econ.queensu.ca/jae/).

Table I: Estimates of the selection equation

	parametric		semi-nonparametric			
			$K=1$		$K=2$	
EDITINC	−0.001	(0.005)	0.000	(0.005)	0.001	(0.007)
AGE30	0.206	(0.107)	0.141	(0.124)	0.237	(0.121)
AGE40	0.104	(0.109)	0.050	(0.118)	0.087	(0.119)
AGE50	0.061	(0.112)	0.003	(0.119)	0.044	(0.122)
AGE60	0.051	(0.123)	−0.068	(0.133)	0.015	(0.135)
AGEGT70	0.115	(0.151)	−0.088	(0.164)	0.019	(0.168)
EDUC	−0.028	(0.010)	−0.065	(0.026)	−0.032	(0.012)
BLACK	0.299	(0.080)	0.253	(0.126)	0.350	(0.106)
OTHER	0.262	(0.215)	0.174	(0.235)	0.327	(0.230)
MARRIED	0.147	(0.089)	0.028	(0.096)	0.122	(0.097)
WIDOW	0.244	(0.142)	0.155	(0.163)	0.286	(0.161)
DIVSEP	0.294	(0.105)	0.166	(0.128)	0.249	(0.119)
EMPLOYED	−0.005	(0.082)	−0.146	(0.103)	−0.049	(0.089)
UNEMPLOY	0.220	(0.176)	0.017	(0.187)	0.124	(0.194)
NORTHE	0.062	(0.083)	−0.033	(0.089)	0.102	(0.093)
MIDWEST	−0.052	(0.079)	−0.155	(0.100)	−0.041	(0.086)
SOUTH	−0.046	(0.079)	−0.155	(0.100)	−0.038	(0.086)
MEDICARE	−0.023	(0.081)	−0.053	(0.091)	0.001	(0.094)
MEDICAID	0.039	(0.113)	0.000	(0.125)	0.013	(0.129)
CHAMPUS	0.017	(0.082)	0.024	(0.090)	−0.034	(0.092)
HLTHINS	−0.142	(0.060)	−0.166	(0.094)	−0.179	(0.076)
REGMED	0.126	(0.090)	0.115	(0.103)	0.258	(0.107)
DRI	0.032	(0.081)	0.038	(0.088)	−0.051	(0.088)
MAJORLIM	0.148	(0.083)	0.126	(0.107)	0.072	(0.102)
SOMELIM	0.033	(0.081)	0.023	(0.087)	0.031	(0.093)
HVDIAB	0.302	(0.087)	0.337	(0.162)	0.344	(0.121)
HHRTCOND	0.183	(0.063)	0.204	(0.102)	0.181	(0.079)
HADSTROK	0.085	(0.128)	0.091	(0.145)	0.206	(0.158)
CONSTANT	−0.583	(0.182)	−0.583		−0.583	

Note: Standard errors are in parentheses.

models with $K = 1$ and $K = 2$ equal 9.408 and 226.890, respectively. This implies that we must reject the hypothesis that heterogeneity follows a bivariate normal distribution. Moreover, the test statistic of $K = 1$ against the semi-nonparametric model with $K = 2$ equals 217.482. Hence, the semi-nonparametric model with $K = 2$ is the best of the three models.

In Tables 1 and 2, we find certain features of the estimated parameters. First, both the estimated parameters and standard errors of the three models, except for the endogenous binary variable's coefficients, closely resemble each other. Second, the parameter values of the endogenous variable (ADVICE) are statistically significant at the 1% level; however, the values differ among the three models: −2.291 in the parametric model, −1.979 in the semi-nonparametric model with $K = 1$, and −1.566 in the

Table II: Estimates of the drinking equation

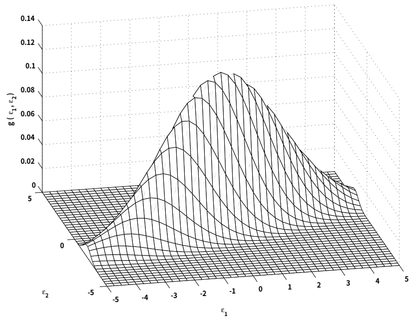
	parametric		semi-nonparametric			
			$K = 1$		$K = 2$	
ADVICE	−2.291	(0.248)	−1.979	(0.341)	−1.566	(0.213)
EDITINC	0.010	(0.011)	0.013	(0.013)	0.005	(0.012)
AGE30	0.153	(0.193)	0.010	(0.189)	0.142	(0.157)
AGE40	−0.075	(0.194)	−0.173	(0.188)	−0.005	(0.156)
AGE50	−0.243	(0.194)	−0.330	(0.188)	−0.101	(0.157)
AGE60	−0.201	(0.204)	−0.420	(0.197)	−0.066	(0.164)
AGEGT70	−0.285	(0.238)	−0.661	(0.227)	−0.280	(0.188)
EDUC	−0.027	(0.017)	−0.084	(0.019)	−0.041	(0.014)
BLACK	0.048	(0.146)	−0.102	(0.139)	−0.097	(0.118)
OTHER	−0.233	(0.400)	−0.441	(0.365)	−0.379	(0.296)
MARRIED	0.012	(0.154)	−0.207	(0.151)	−0.071	(0.125)
WIDOW	0.329	(0.245)	0.141	(0.238)	0.172	(0.200)
DIVSEP	0.403	(0.187)	0.135	(0.181)	0.335	(0.150)
EMPLOYED	0.084	(0.131)	−0.126	(0.128)	−0.035	(0.105)
UNEMPLOY	0.729	(0.304)	0.353	(0.288)	0.424	(0.241)
NORTHE	−0.063	(0.148)	−0.231	(0.142)	−0.077	(0.120)
MIDWEST	−0.272	(0.140)	−0.429	(0.137)	−0.212	(0.114)
SOUTH	−0.238	(0.139)	−0.403	(0.135)	−0.170	(0.113)
CONSTANT	2.584	(0.318)	2.584		2.584	
σ_1	2.199	(0.094)	1.843	(0.237)	1.730	(0.398)
ρ	0.835	(0.039)	0.755	(0.101)	0.784	(0.097)
α_{01}			0.326	(2.125)	−1.206	(1.205)
α_{02}					−0.643	(0.437)
α_{10}			0.051	(0.956)	0.717	(0.735)
α_{11}			0.156	(0.090)	0.805	(0.166)
α_{12}					0.183	(0.238)
α_{20}					−0.220	(0.115)
α_{21}					−0.061	(0.130)
α_{22}					−0.018	(0.011)
log-likelihood	−10,202.043		−10,197.339		−10,088.598	

Note: Standard errors are in parentheses.

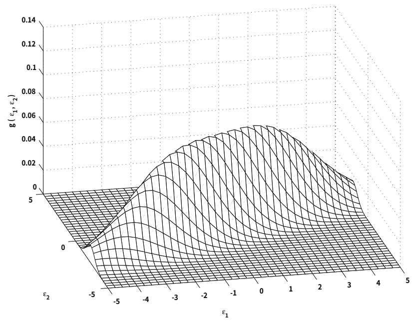
semi-nonparametric model with $K = 2$.[7] This means that advice appears to reduce the consumption of alcoholic beverages by $[\exp(-2.291) - 1] \times 100 = -89.9\%$ in the parametric model, $[\exp(-1.979) - 1] \times 100 = -86.2\%$ in the semi-nonparametric model with $K = 1$, and $[\exp(-1.566) - 1] \times 100 = -79.1\%$ in the semi-nonparametric model with $K = 2$. Compared to the result of the parametric model, the reduction of alcoholic beverages in the two semi-nonparametric models is small. Based on the LR test and estimated

[7] Using the Gauss-Hermite quadrature, the log-likelihood values of the parametric, $K = 1$, and $K = 2$ models are $-10,732.920$, $-10,460.757$, and $-10,271.797$, respectively. The coefficient values of the endogenous binary variable (ADVICE) are -1.235, -1.038, and -0.819, respectively. Moreover, the advice effects of the three models are -70.9%, -64.6%, and -55.9%, respectively.

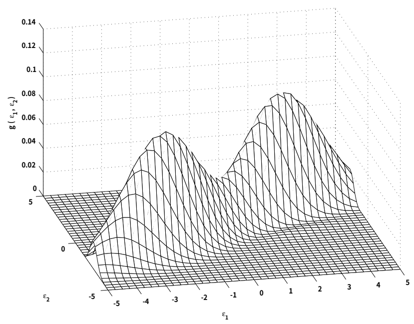
Figure 1: Estimated density of heterogeneity



(a) parametric



(b) $K = 1$



(c) $K = 2$

results, the influence of the doctor's advice has a negative effect on drinking behavior; however, it can be overestimated by the parametric model.

Figure 1 graphs the estimated densities of the three models using the 10% significant coefficients. We find that the semi-nonparametric model with $K =$

1 has a fatter tail than the normal density (parametric) model; moreover, the semi-nonparametric model with $K = 2$ is a twin-peak distribution.

4. Conclusion

This paper proposes a new semi-nonparametric count data estimation with an endogenous binary variable that generalizes bivariate correlated unobserved heterogeneity using Hermite polynomials. In an example using the 1990 NHIS data, the semi-nonparametric model with $K = 2$ overcomes the other models in terms of the LR test. The absolute values of the endogenous binary regressor coefficients of the semi-nonparametric models are smaller than that of the parametric model, and that of the semi-nonparametric model with $K = 2$ is the smallest of the three. This introduces the interpretation of the binary endogenous variable, that is, the effect of the advice variable. The parametric model overestimates the effect of doctor's advice in our example. Moreover, the estimated densities of the semi-nonparametric models have fatter tail than that of the parametric model.

One major advantage of the semi-nonparametric model is the flexibility of bivariate distributed heterogeneity. The difference between the endogenous binary variable's coefficients of the parametric and semi-nonparametric models is not negligible in our example. Therefore, it is useful to generalize bivariate heterogeneity using Hermite polynomials.

References

Cameron, A.C., and P. Johansson (1997) "Count Data Regression Using Series Expansions: With Applications" *Journal of Applied Econometrics* **12**, 203-223.

Deb, P., M.K. Munkin, and P.K. Trivedi (2006) "Private Insurance, Selection, and Health Care Use: A Bayesian Analysis of a Roy-type Model" *Journal of Business and Economic Statistics* **24**, 403-415.

Gabler, S., F. Laisney, and M. Lechner (1993) "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation" *Journal of Business and Economic Statistics* **11**, 61-80.

Gallant, A.R., and D.W. Nychka (1987) "Semi-nonparametric Maximum Likelihood Estimation" *Econometrica* **55**, 363-390.

Jochmann, M. (2003) "Semiparametric Bayesian Inference for Count Data Treatment Models" *mimeo.*

Kenkel, D., and J. Terza (2001) "The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect" *Journal of Applied Econometrics* **16**, 165-84.

Kozumi, H. (2002) "A Bayesian Analysis of Endogenous Switching Models for Count Data" *Journal of the Japan Statistical Society* **32**, 141-154.

Melenberg, B., and A. van Soest (1993) "Semi-parametric Estimation of the Sample Selection Model" Working Paper 9334, CentER, Tilburg.

Miranda, A., and S. Rabe-Hesketh (2006) "Maximum Likelihood Estimation of Endogenous Switching and Sample Selection Models for Binary, Count, and Ordinal Variables" *The Stata Journal* **6**, 285-308.

Mullahy, J. (1997) "Instrumental-variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior" *Review of Economics and Statistics* **79**, 586-593.

Munkin, M.K., and P.K. Trivedi (2003) "Bayesian Analysis of a Self-selection Model with Multiple Outcomes Using Simulation-based Estimation: An Application to the Demand for Healthcare" *Journal of Econometrics* **114**, 197-220.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2002) "Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature" *The Stata Journal* **2**, 1-21.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005) "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects" *Journal of Econometrics* **128**, 301-323.

Romeu, A., and M. Vera-Hernández (2005) "Counts with an Endogenous Binary Regressor: A Series Expansion Approach" *Econometrics Journal* **8**, 1-22.

Terza, J.V. (1998) "Estimating Count-data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects" *Journal of Econometrics* **84**, 129-154.

van der Klaauw, B., and R.H. Koning (2003) "Testing the Normality Assumption in the Sample Selection Model with an Application to Travel Demand" *Journal of Business and Economic Statistics* **21**, 31-42.

Windmeijer, F.A.G., and J.M.C. Santos Silva (1997) "Endogeneity in Count Data Models: An Application to Demand for Health Care" *Journal of Applied Econometrics* **12**, 281-294.

Appendix

Following van der Klaauw and Koning (2003), we specify the bivariate semi-nonparametric normal density as follows:

$$g(\varepsilon_{1i}, \varepsilon_{2i}) = \frac{1}{P}\left(\sum_{j=0}^{K}\sum_{k=0}^{K}\alpha_{jk}\varepsilon_{1i}^{j}\varepsilon_{2i}^{k}\right)^2 \times \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{\varepsilon_{1i}}{\sigma_1}\right)^2 - 2\rho\frac{\varepsilon_{1i}}{\sigma_1}\frac{\varepsilon_{2i}}{\sigma_2} + \left(\frac{\varepsilon_{2i}}{\sigma_2}\right)^2\right\}\right], \quad \text{(A1)}$$

where

$$P = \iint_{-\infty}^{\infty}\left(\sum_{j=0}^{K}\sum_{k=0}^{K}\alpha_{jk}\varepsilon_{1i}^{j}\varepsilon_{2i}^{k}\right)^2 \times \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{\varepsilon_{1i}}{\sigma_1}\right)^2 - 2\rho\frac{\varepsilon_{1i}}{\sigma_1}\frac{\varepsilon_{2i}}{\sigma_2} + \left(\frac{\varepsilon_{2i}}{\sigma_2}\right)^2\right\}\right] d\varepsilon_{1i}\, d\varepsilon_{2i}.$$

We now consider the case of $K = 2$. The round bracket of Eq.(A1) can be rearranged as

$$\left(\sum_{j=0}^{2}\sum_{k=0}^{2}\alpha_{jk}\varepsilon_{1i}^{j}\varepsilon_{2i}^{k}\right)^2 = \gamma_0 + \gamma_1\varepsilon_{2i} + \gamma_2\varepsilon_{2i}^2 + \gamma_3\varepsilon_{2i}^3 + \gamma_4\varepsilon_{2i}^4,$$

where

$$\gamma_0 = \left(\alpha_{00} + \alpha_{10}\varepsilon_{1i} + \alpha_{20}\varepsilon_{1i}^2\right)^2,$$
$$\gamma_1 = 2\left(\alpha_{00} + \alpha_{10}\varepsilon_{1i} + \alpha_{20}\varepsilon_{1i}^2\right)\left(\alpha_{01} + \alpha_{11}\varepsilon_{1i} + \alpha_{21}\varepsilon_{1i}^2\right),$$
$$\gamma_2 = 2\left(\alpha_{00} + \alpha_{10}\varepsilon_{1i} + \alpha_{20}\varepsilon_{1i}^2\right)\left(\alpha_{02} + \alpha_{12}\varepsilon_{1i} + \alpha_{22}\varepsilon_{1i}^2\right) + \left(\alpha_{01} + \alpha_{11}\varepsilon_{1i} + \alpha_{21}\varepsilon_{1i}^2\right)^2,$$
$$\gamma_3 = 2\left(\alpha_{01} + \alpha_{11}\varepsilon_{1i} + \alpha_{21}\varepsilon_{1i}^2\right)\left(\alpha_{02} + \alpha_{12}\varepsilon_{1i} + \alpha_{22}\varepsilon_{1i}^2\right),$$
$$\gamma_4 = \left(\alpha_{02} + \alpha_{12}\varepsilon_{1i} + \alpha_{22}\varepsilon_{1i}^2\right)^2.$$

We require the following algebraic computation to obtain Eq.(4) or P:

$$\int_{-\infty}^{\infty}\int_{\hat{a}}^{\hat{b}} f(y_i|d_i, \varepsilon_{1i})\, g(\varepsilon_{1i}, \varepsilon_{2i})\, d\varepsilon_{2i}\, d\varepsilon_{1i}$$
$$= \int_{-\infty}^{\infty} f(y_i|d_i, \varepsilon_{1i}) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_{1i}}{\sigma_1}\right)^2\right)$$

$$
\begin{aligned}
&\times \frac{1}{P}\int_{\hat{a}}^{\hat{b}} \sum_{j=0}^{4} \gamma_j \varepsilon_2^j \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \\
&\times \exp\left[-\frac{1}{\left(\sigma_2\sqrt{2\left(1-\rho^2\right)}\right)^2}\left(\varepsilon_{2i}-\rho\frac{\sigma_2}{\sigma_1}\varepsilon_{1i}\right)^2\right] d\varepsilon_{2i}\, d\varepsilon_{1i} \\
\equiv &\int_{-\infty}^{\infty} f\left(y_i|d_i,\varepsilon_{1i}\right) g_1\left(\varepsilon_{1i}\right) \int_{\hat{a}}^{\hat{b}} \frac{g_2\left(\varepsilon_{2i}|\varepsilon_{1i}\right)}{P} d\varepsilon_{2i}\, d\varepsilon_{1i} \\
\equiv &\int_{-\infty}^{\infty} f\left(y_i|d_i,\varepsilon_{1i}\right) g_1\left(\varepsilon_{1i}\right) \frac{G_2\left(\varepsilon_{1i}\right)}{P} d\varepsilon_{1i}, \qquad \text{(A2)}
\end{aligned}
$$

where $g_1(\cdot)$ is the probability density function of a normal distribution. When $\hat{a} = -\infty$ and $\hat{b} = \infty$, Eq.(A2) results in P. Substituting $\xi = \rho\sigma_2\varepsilon_{1i}/\sigma_1$, $u = \varepsilon_{2i} - \xi$ and $\delta = \sigma_2\sqrt{2\left(1-\rho^2\right)}$ into Eq.(A2) yields

$$
\begin{aligned}
\frac{G_2}{P} &= \frac{1}{\sqrt{\pi}\delta}\frac{1}{P}\int_{\hat{a}-\xi}^{\hat{b}-\xi} \sum_{j=0}^{4} \gamma_j\left(u+\xi\right)^j \exp\left[-\left(\frac{u}{\delta}\right)^2\right] du \\
&\equiv \frac{1}{\sqrt{\pi}\delta}\frac{1}{P}\int_{a}^{b} \sum_{j=0}^{4} \eta_j u^j \exp\left[-\left(\frac{u}{\delta}\right)^2\right] du,
\end{aligned}
$$

where $a = \hat{a} - \xi$, $b = \hat{b} - \xi$, and

$$
\begin{aligned}
&\eta_0 = \gamma_4\xi^4 + \gamma_3\xi^3 + \gamma_2\xi^2 + \gamma_1\xi + \gamma_0, && \eta_1 = 4\gamma_4\xi^3 + 3\gamma_3\xi^2 + 2\gamma_2\xi + \gamma_1, \\
&\eta_2 = 6\gamma_4\xi^2 + 3\gamma_3\xi + \gamma_2, && \eta_3 = 4\gamma_4\xi + \gamma_3, \\
&\eta_4 = \gamma_4.
\end{aligned}
$$

Therefore, using the following recursion formula (van der Klaauw and Koning, 2003):

$$
\begin{aligned}
I_k\left(a,b\right) &= \int_a^b u^k \exp\left(-\frac{u^2}{\delta^2}\right) du \\
&= \frac{\delta^2}{2}\left[a^{k-1}\exp\left(-\frac{a^2}{\delta^2}\right) - b^{k-1}\exp\left(-\frac{b^2}{\delta^2}\right)\right] + \frac{\left(k-1\right)\delta^2}{2} I_{k-2}\left(a,b\right)
\end{aligned}
$$

and substituting $b = -z_i'\beta_2 - \xi$, we obtain the following relation:

$$
\begin{aligned}
\frac{G_2\left(\varepsilon_{1i}|d_i=0\right)}{P} =& \frac{1}{\sqrt{\pi}\delta}\frac{1}{P}\left[\eta_0 I_0\left(-\infty,b\right) + \eta_1 I_1\left(-\infty,b\right) + \eta_2 I_2\left(-\infty,b\right)\right. \\
&\left. + \eta_3 I_3\left(-\infty,b\right) + \eta_4 I_4\left(-\infty,b\right)\right].
\end{aligned}
$$

Using the same procedure, we can calculate the term P.