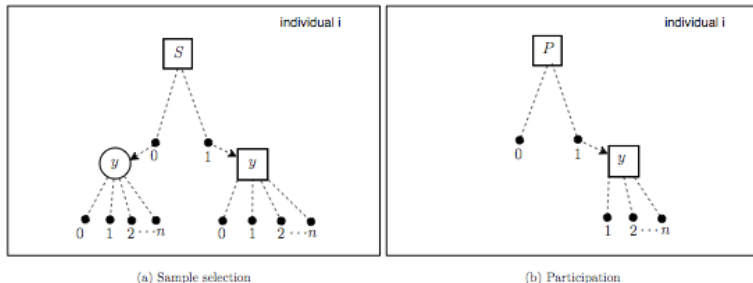


Endogenous Treatment Effects for Count Data Models with Sample Selection or Endogenous Participation

Massimilano Bratti & Alfonso Miranda

Sample selection vs Participation

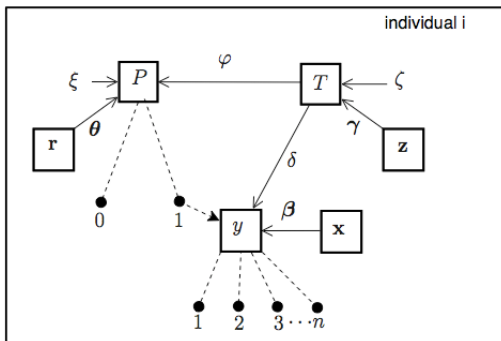
In many fields of applied work researchers need to model a count dependent variable y that is only observed for a proportion of the sample (i.e., a sample selection problem) or that is defined only for a subset of the population (i.e., a participation problem)



Note. Observed variables are written inside a square and unobserved variables are written inside a circle. S is a selection dummy and P is a participation dummy

Participation + Treatment

Often the count dependent variable is, itself, function of a dichotomous variable that indicates a treatment condition T



Note. Solid arrows indicate functional relationships. x , z , and r are vectors of explanatory variables. ξ and ψ are random errors. β , γ , and θ are vector of coefficients. And δ and φ are coefficients

Examples

- ▶ Effect of physician advice on individual alcohol or cigarettes consumption (Kenkel et al. 2001)
- ▶ Effect of health status (good vs. poor) on the number of visits to a general practitioner (Windmaijer and Santos Silva, 1997)
- ▶ Effect of health insurance coverage (additional coverage or public vs. private insurance) on the number of doctor or hospital visits (Riphahn et al. 2003)

The challenge

The participation (sample selection) and treatment dummies are both likely to be endogenous. In the drinking example:

- ▶ Health 'oriented' individuals are more likely to seek advice and less likely to participate (drink), and abuse drinking given participation
- ▶ Individuals who have been given advice against drinking are less likely to participate
- ▶ Physicians are less likely to give advice to individuals with low levels of consumption

The aim

We aim to develop methods to deal with an endogenous participation (or a sample selection) problem together with an endogenous treatment problem in a model for a count variable

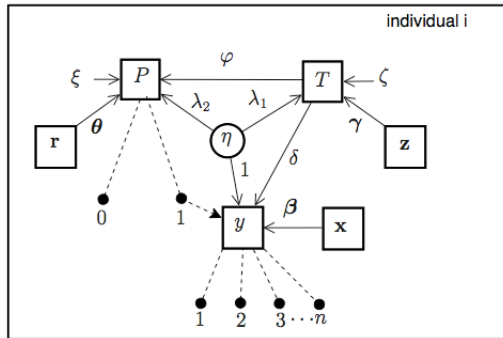
The method should be able to deal with cases in which **unobserved factors** affecting sample selection or participation are correlated with those affecting exposure to the treatment and those relevant for the intensive margin of the activity of interest (smoking, drinking, etc.)

Previous work

To the best of our knowledge, there are models that are able to address only one of these issues at the time, or that address a similar – but not exactly the same – issue

- ▶ Endogenous participation (selection) and endogenous treatment
 - ▶ Terza et al. (2008) propose a two-step estimator for a dependent variable observed in interval form to model sample selection and endogenous treatment;
 - ▶ Li and Trivedi (2009) use a Bayesian approach to estimate a model for a continuous and non negative dependent variable with endogenous participation and multivariate treatments
- ▶ Endogenous participation or endogenous treatment
 - ▶ Greene 2009 (excellent review)
 - ▶ Kenkel and Terza (2001), which will be described later in more details

Endogenous Participation + Endogenous Treatment



Note. η represents an unobserved or latent variable. λ_1 and λ_2 are parameters

Formal definition I

We start with the **Endogenous Participation Endogenous Treatment (EPET) Poisson** model we assume that the treatment dummy T and the participation dummy P are generated as follows:

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + v, \quad (1)$$

$$P^* = \mathbf{r}'\boldsymbol{\theta} + \phi T + q \quad (2)$$

with $T = 1(T^* > 0)$, $P = 1(P^* > 0)$. The count y is generated according to the following cumulative conditional distribution function,

$$G(y|\eta) \equiv \Pr(y|\eta) = \begin{cases} \text{not defined} & \text{if } P = 0 \\ [\mu^y \exp(-\mu)] / [1 - \exp(-\mu)] y! & \text{if } P = 1 \end{cases}$$

with,

$$y = \begin{cases} 0 & \text{if } P = 0 \\ 1, 2, \dots & \text{if } P = 1 \end{cases} \quad (3)$$

$$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, \quad (4)$$

Formal definition II

An important feature of the EPET Poisson model is the $y = 0$ count is generated by a different data generating mechanism from $y \neq 0$ counts. This is equivalent to saying that the decision to drink vs. not to drink is a different one from the one on the number of drinks to consume (conditional on drinking), and should be modelled separately, although the two processes may be correlated

Correlation between T , P and y is induced by imposing some structure on the residuals of equations (1) and (2),

$$\begin{aligned}v &= \lambda_1 \eta + \zeta \\q &= \lambda_2 \eta + \xi,\end{aligned}\tag{5}$$

Main assumptions

To close the model we require the covariates to be all exogenous and impose some distributional conditions

$$D(\eta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \zeta, \xi) = D(\eta) \quad (\text{C1})$$

$$D(\zeta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\zeta|\eta) \quad (\text{C2})$$

$$D(\xi|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\xi|\eta) \quad (\text{C3})$$

$$\zeta \perp \xi \mid \eta, \quad (\text{C4})$$

In what follows, we assume that $\eta \sim N(0, \sigma_\eta^2)$ and that $\zeta|\eta$ and $\xi|\eta$ are both distributed as independent standard normal variates, and that all covariates (except the treatment of interest) are exogenous

Implied correlations

Correlations between y , T , and S are functions of the factor loadings λ_1 and λ_2 . In particular, the model implies the following correlations:

$$\rho_{\eta,v} = \frac{\lambda_1 \sigma_\eta^2}{\sqrt{\sigma_\eta^2(\lambda_1^2 \sigma_\eta^2 + 1)}} \quad (6)$$

$$\rho_{\eta,q} = \frac{\lambda_2 \sigma_\eta^2}{\sqrt{\sigma_\eta^2(\lambda_2^2 \sigma_\eta^2 + 1)}} \quad (7)$$

$$\rho_{v,q} = \frac{\lambda_1 \lambda_2 \sigma_\eta^2}{\sqrt{(\lambda_1^2 \sigma_\eta^2 + 1)(\lambda_2^2 \sigma_\eta^2 + 1)}}. \quad (8)$$

Estimation

We estimate the model by Maximum Simulated Likelihood (MSL) and it is identified by restrictions on the covariance matrix and by functional form. So, \mathbf{x} , \mathbf{z} , and \mathbf{r} can all have the same elements. However, specifying some exclusion restrictions for the selection and/or treatment equations is always advisable when it is possible

Likelihood function

$$\log(L) = \sum_{i, P_i=0} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \Pr_P(0|\eta) \Pr_T(\tau|\eta) \phi(\eta) d\eta \right\} \\ + \sum_{i, P_i=1} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \Pr_P(1|\eta) \Pr_T(\tau|\eta) G(y|\eta) \phi(\eta) d\eta \right\}$$

Note 1 $\Pr_P(0|\eta)$ is the conditional probability of $P = 0$ given η and $\Pr_P(1|\eta)$ is the conditional probability of $P = 1$ given η

Note 2 $\Pr_T(\tau|\eta)$ represents the probability of $T = \tau$ given η , with $\tau = 0, 1$

Note 3 $G(y|\eta)$ is the cumulative distribution of y

Note 4 $\omega_0 = 1(T = 0)$, and $\omega_1 = 1(T = 1)$

Note 5 To simplify notation, conditioning on observable variables is not explicitly written

The model implies overdispersion

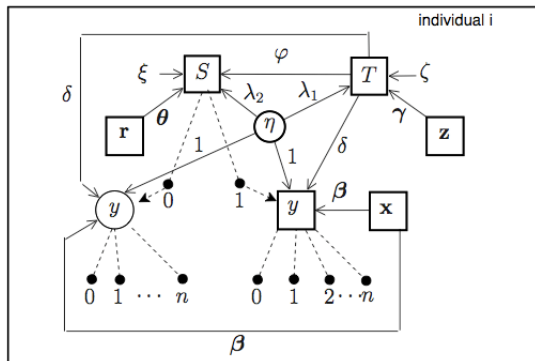
The use of the Poisson distribution for the analysis of count data has been criticized in the past due to the unattractive feature that conditional mean and conditional variance are restricted to be equal, also known as equidispersion. In the present model, however, the introduction of the unobserved heterogeneity term η in the log-linear model for $\mu \equiv E[y|x, T, \eta]$ forces the count variable y to exhibit overdispersion. In fact, it can be shown that:

$$\kappa \equiv E[y|x, T] = E_{\eta} [E[y|x, T, \eta]] = \exp \left(\mathbf{x}'\beta + \delta T + \frac{\sigma_{\eta}^2}{2} \right) \quad (9)$$

$$\text{Var}[y|x, T] = \kappa \{ 1 + \kappa [\exp(\sigma_{\eta}^2) - 1] \} \quad (10)$$

so that, in general, $\text{Var}[y|x, T] \geq E[y|x, T]$ because $\sigma_{\eta}^2 \geq 0$ by definition. This implies, therefore, that the methods here described cannot be used to fit underdispersed count data because in that case $\text{Var}[y|x, T] < E[y|x, T]$, and η cannot have negative variance

Sample selection + endogenous treatment count model



SET-Poisson

In the **Sample Selection Endogenous Treatment (SET) Poisson** model we assume that the treatment dummy T and the sample selection dummy S are generated as follows:

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + \lambda_1\eta + \zeta, \quad (11)$$

$$S^* = \mathbf{r}'\boldsymbol{\theta} + \varphi T + \eta + \xi \quad (12)$$

with $T = 1(T^* > 0)$, $S = 1(S^* > 0)$ and,

$$F(y|\eta) \equiv \Pr(y|\eta) = \begin{cases} \text{not defined} & \text{if } S = 0 \\ [\mu^y \exp(-\mu)] / y! & \text{if } S = 1. \end{cases} \quad (13)$$

$$y = \begin{cases} \text{missing} & \text{if } S = 0 \\ 0, 1, 2, \dots & \text{if } S = 1, \end{cases} \quad (14)$$

$$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, \quad (15)$$

Application: Kenkel and Terza 2001

We apply the EPET Poisson model to Kenkel and Terza's (2001) data (KT, hereafter). KT analyze the effect of physician advice on drinking using the 1990 National Health Interview Survey (U.S.)

The authors drop from the analysis lifetime abstainers and former drinkers with no drinking in the past year. Because the physician advice to cut drinking was recommended as a way of reducing high blood pressure, the authors focus only on men who have drunk alcohol at least once in the last 12 months and report having been told at some time that they had high blood pressure

In spite of this sample selection, KT observe in their sample that 21% of current drinkers (according to their definition) did not drink at all in the last two weeks (excess of zeroes)

Application: Kenkel and Terza 2001

Excess zeros:

- ▶ recent quitters or people who were actively trying to stop drinking all together in the last 12 months
- ▶ individuals who drink only in very special occasions such as weddings, birthdays, or Christmas day (occasional drinkers)
- ▶ 'frequent' drinkers who, by chance, did not drink any alcohol in the past two weeks; although this last scenario is less likely

Since KT compute drinking as 'the product of self-reported drinking frequency (the number of days in the past two weeks with any drinking) and drinking intensity (the average number of drinks on a day with any drinking)', (p. 171-172), the measure refers to the last two weeks, a sufficiently long time-span. For this reason, we expect the first two explanations above to be more relevant in this specific case

Application: Kenkel and Terza 2001

KT account for this excess zero by using a flexible functional form for the conditional mean of drinking based on the inverse Box-Cox transformation. In contrast, here we account for it using a standard Poisson model, but treating the zeros and the positive drinking outcomes as if they were generated by two separate DGPs.

Our idea is that the DGP for occasional drinkers or quitters should be a different one from that ruling day-by-day drinking by frequent drinkers

Results

We estimated several models:

- ▶ A Poisson model
- ▶ An Endogenous Treatment (ET) Poisson model (neglecting endogenous participation)
- ▶ An EPET model

In what follows all estimations by MSL use 1600 Halton draws to perform the integration. Re-estimating the models with 2000 draws did not change significantly coefficient or standard errors

Sample selection + endogenous treatment count model

Table 2 Marginal effects of physician advice on the number of drinks consumed in the last two weeks and on the probability of drinking

	Poisson	Probit	ET- Poisson	EP- Poisson		EPET- Poisson	
	$y^{(a)}$ (1)	$\Pr(y>0)^{(b)}$ (2)	$y^{(a)}$ (3)	$\Pr(y>0)^{(b)}$ (4)	$y>0^{(a)}$ (5)	$\Pr(y>0)^{(b)}$ (6)	$y>0^{(a)}$ (7)
Physician advice (T)	3.679*** (.558)	0.079*** (.017)	-5.395*** (.386)	0.079*** (.017)	3.377*** (.557)	-0.045 (.049)	-4.072*** (.864)
$\hat{\rho}_{\eta,v}$			0.832*** (.029)			0.689*** (.092)	
$\hat{\rho}_{\eta,q}$.082 (.141)		0.378*** (.088)	
$\hat{\rho}_{v,q}$						0.261*** (.084)	
$\hat{\sigma}_{\eta}^2$			2.190*** (.069)	1.207*** (.023)		1.456*** (.099)	
N.obs.	2,467	2,467	2,467	2,467		2,467	
Log-likelihood	-32,263	-1,247	-10,184	-8,660		-10,062	
$\widehat{\Pr}(y=0)^{(c)}$	0.00	0.21	0.11	0.21		0.22	
$BIC^{(d)}$		70,360	23,418	20,675		20,671	

Notes

- ▶ *** significant at 1%. Eicker-White robust standard errors in parentheses.
- ▶ (a) The marginal effect is computed for a discrete change in T (from 0 to 1) at the sample median of the dependent variable, in analogy to KT;
- ▶ (b) The marginal effect is computed for a discrete change in T (from 0 to 1) at the sample mean of the other independent variables;
- ▶ (c) Probability of the zero outcome predicted by the model.
- ▶ (d) Bayesian information criterion. For the sake of comparability all BICs refer to a three-equation model. In the ET-Poisson the BIC refers to the ET-Poisson and the probit equation for exogenous participation; in the EP-Poisson the BIC refers to the EP-Poisson and the probit equation for exogenous treatment. The BIC in columns (1)-(2) refers to the probit models for exogenous participation and exogenous treatment and the Poisson model with exogenous participation and exogenous treatment (including the zeros).
- ▶ Note. y is the number of alcoholic drinks consumed in the last two weeks, $\Pr(y>0)$ the probability of drinking in the last two weeks, $y>0$ the number of alcoholic drinks consumed in the last two weeks conditional on drinking and T a dichotomous indicator of individual treatment status. Estimation refers to the 1990 NHIS with the sample selection and covariates used in KT. ET, EP and EPET stand for Endogenous Treatment, Endogenous Participation and Endogenous Participation Endogenous Treatment, respectively. ET-Poisson, EP-Poisson and EPET-Poisson models were estimated using MSL and 1, 600 Halton draws. The joint Wald test statistic for $\rho_{y,T} = \rho_{y,P} = \rho_{T,P} = 0$ in the EPET-Poisson model, distributed as a $\chi^2(3)$, is 57.904 (p-value=0.00).

Main findings

- ▶ Neglecting the potential endogeneity of the treatment produces wrongly signed estimates of the treatment effects
- ▶ Neglecting endogenous participation leads to an overestimate of the treatment effects
- ▶ The EPET Poisson model produces estimates of the treatment effects that are in between those produced by the ET Poisson model and the estimates by KT

Remark

The EPET Poisson model also allows the same covariates to have different effects on the the intensive and the extensive drinking margins

Conclusions In this paper, we:

- ▶ Propose a model suitable for all cases in which an endogenous treatment affects a count outcome variables, and in which there are endogenous participation or sample selection issues
- ▶ Show that accounting for the endogeneity of the treatment status and endogenous participation issues is paramount to obtaining consistent estimates of treatment effects, using data from KT study of physician advice on drinking

In the future there are plans to make the model's code available to the general public through a Stata's ado file

Comments are welcome, Thanks!

References

- ▶ Bratti, M., Miranda, A., 2011. Endogenous treatment effects for count data models with endogenous participation or sample selection. *Health Economics* (forthcoming). [Working paper version](#).
- ▶ Greene, W., 2009. Models for count data with endogenous participation. *Empirical Economics* 36, 133–173.
- ▶ Kenkel, D., Terza, J., 2001. The effect of physician advice on alcohol consumption: Count regression with an endogenous treatment effect. *Journal of Applied Econometrics* 16, 165–184.
- ▶ Li, Q., Trivedi, P. K., 2009. Impact of prescription drug coverage on drug expenditure of the elderly – evidence from a two-part model with endogeneity, mimeo.
- ▶ Riphahn, R., Wambach, A., Million, A., 2003. Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Econometrics* 4, 387–405.
- ▶ Terza, J. V., Kenkel, D. S., Lin, T.-F., Sakata, S., 2008. Care-giver advice as a preventive measure for drinking during pregnancy: Zeros, categorical outcome responses, and endogeneity. *Health Economics* 17, 41–54.
- ▶ Windmijer, F., Santos Silva, J., 1997. Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics* 12, 281–94.