DISCUSSION PAPER SERIES

IZA DP No. 5372

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

Massimiliano Bratti
Alfonso Miranda

December 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Massimiliano Bratti
Università degli Studi di Milano
and IZA

Alfonso Miranda
IoE, University of London
and IZA

Discussion Paper No. 5372
December 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5372
December 2010

ABSTRACT

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

In this paper we propose an estimator for models in which an endogenous dichotomous treatment affects a count outcome in the presence of either sample selection or endogenous participation using maximum simulated likelihood. We allow for the treatment to have an effect on both the sample selection or the participation rule and the main outcome. Applications of this model are frequent in – but are not limited to – health economics. We show an application of the model using data from Kenkel and Terza (2001), who investigate the effect of physician advice on the amount of alcohol consumption. Our estimates suggest that in these data (i) neglecting treatment endogeneity leads to a wrongly signed effect of physician advice on drinking intensity, (ii) neglecting endogenous participation leads to an upward biased estimate of the treatment effect of physician advice on drinking intensity.

JEL Classification: C35, I12, I21

Keywords: count data, drinking, endogenous participation, maximum simulated likelihood, sample selection, treatment effects

Corresponding author:

Alfonso Miranda
Department of Quantitative Social Science
Institute of Education
University of London
20 Bedford Way
London WC1H 0AL
United Kingdom
E-mail: A.Miranda@ioe.ac.uk

* Both authors equally contributed to this paper. Participants to presentations held at the 19th European Workshop on Econometrics and Health Economics (Lausanne), the National Autonomous University of México, and the College of México are acknowledged for their suggestions. We are grateful to João Santos Silva and Stefanie Schurer for their useful comments. The usual disclaimers apply.

1 Introduction

There are often cases in health economics in which one is interested in the effect of an endogenous dichotomous treatment on an outcome which takes on non-negative integer values with cardinal interpretation (count data). Examples include — but are not limited to — the effect of physician advice on individual alcohol consumption (Kenkel and Terza 2001), the effect of health status on the number of visits to a general pratictioner (Windmaijer and Santos Silva 1997), or the effect of health insurance coverage on the number of doctor or hospital visits (Riphahn et al. 2003).[1]

In all these applications the treatment of interest is likely to be endogenous (*endogenous treatment*).[2] Receiving physician advice to reduce drinking is certainly not exogenous with respect to the intensity of those activities by patients. Health conditions may not be exogenous with respect to the number of visits to a doctor since individuals who are less concerned with their health may engage in health damaging behavior and at the same time be less prone to see a doctor. Similarly, demand for health insurance is clearly endogenous as high-risk types are expected to buy more comprehensive coverage.

In addition to an endogeneity problem, in all these cases one is likely to have also an *endogenous participation* problem.[3] Indeed, participation to an activity, such as smoking, drinking or seeing a doctor (the extensive margin) and the intensity of the activity (the number of cigarettes or drinks consumed or the number of visits, i.e. the intensive margin) may be two very different processes. For this reason, one might want let the two processes to be produced by different data generating processes (DGPs, hereafter). For instance, one may be likely to see a doctor only if she is ill. And the amount of health insurance coverage may have an effect only on the intensive margin of the activity, not on the fact that one sees or does not see a doctor.

In other cases, one may have a *sample selection* issue. In a sample of smokers or drinkers, for instance, data on cigarette or alcohol consumption may have not been reported by all individuals, and the data may not be missing at random (NMAR) with respect to the level of drinking or smoking. For instance, heavy smokers or drinkers may have not answered the survey. In this case, neglecting sample selection will lead to inconsistent estimates of

[1] See Winkelmann (1998) and Greene (2009) for a review of count data models with selectivity.

[2] This issue is acknowledged and addressed in all the articles we cited.

[3] Here we use a terminology different from that in Greene (2009), who considers self-selection into the treatment as an instance of 'endogenous participation.' In what follows by 'participation' we generally mean participation in the activity measured by the main (count) outcome variable (e.g., drinking, smoking).

the treatment effects of interest (e.g., physician advice).

A number of previous papers have suggested strategies for estimating count data models with either sample selection or endogenous treatment, though not both at the same time. Greene (1997), Terza (1998), Winkelmann (1998), Miranda (2004) and Miranda and Rabe-Hesketh (2006) discuss fully parametric methods for estimating count data models based on the Poisson distribution and normally distributed unobserved heterogeneity. Kenkel and Terza (2001) use a flexible Box-Cox specification for the count and normally distributed unobserved heterogeneity to develop a two-step method for estimating the endogenous treatment model. Windmaijer and Santos Silva (1997) discuss a GMM strategy that only requires the specification of the conditional mean of the count y and it is thus less restrictive in terms of the distributional assumptions about y that the researcher needs to impose to achieve a consistent estimator.

The endogenous participation model is closely related to the double-hurdle model of Cragg (1971), the Tobit-type estimator for censored Poisson regression of Terza (1985), the hurdle model of Mullahy (1986), the double-hurdle model of Jones (1989), the two-part model of Mullahy (1998), the endogenous hurdle of Greene (2009), and the zero-inflated count model of Melkersson and Rooth (2000). All these models are motivated by the idea that individuals must cross one or two hurdles before a strictly positive value of the dependent variable y is observed. Further, the zero outcome is thought to be special in the sense that a large proportion of the individuals in the sample chooses $y = 0$ and that the *participation* decision is qualitatively different from the *intensity of consumption* decision. For these reasons the models above suggest specifying a different data generating mechanism for zero and strictly positive y. These models have been used to analyze smoking, drinking, and fertility behavior among other applications. Endogenous participation is allowed in Greene (2009) and Mullahy (1998). However, none of the aforementioned models allow endogenous participation and endogenous treatment at the same time.

To the best of our knowledge, to date only Terza et al. (2008) and Li and Trivedi (2009) have suggested strategies to address endogenous treatment and endogenous participation at the same time. Terza et al. (2008) put forward a two-step estimator for a *grouped* dependent variable which relies on a joint normality assumption. In this paper, in contrast, we propose an estimation method which is appropriate to deal with endogenous treatment affects and with either sample selection or endogenous participation when the dependent variable is a *count*. Li and Trivedi (2009), on their side, use a Bayesian approach and a two-part model to estimate a model for a *continuous and non negative* dependent variable with endogenous participation and multivariate

treatments. Also in this case, multivariate normality is required. The estimator we propose is similar in spirit to the ones proposed by Terza et al. (2008) and Li and Trivedi (2009), relies on the same distributional assumptions, and addresses both treatment endogeneity and endogenous participation or both treatment endogeneity and sample selection at the same time.[4] In addition to the different type of dependent variable that is used in the two aforementioned studies, our approach differs from that of Terza et al. in the sense that we use Maximum Simulated Likelihood (MSL). As a consequence, we gain in efficiency with respect to the two-step estimator of Terza et al. and obtain correct standard errors in the usual way unlike the two-step approach where standard errors need to be corrected after estimation. Our approach differs from Li and Trivedi (2009), as we use a frequentist rather than a Bayesian approach.

We illustrate our estimator using data from Kenkel and Terza (2001), who study the effect of physician advice on drinking.

The structure of the paper is as follows. In the next section we report a description of the econometric model, focusing on the case of estimating endogenous treatment and endogenous participation (the case of endogenous treatment and sample selection is reported in Appendix A). In section 3 we apply our estimator to Kenkel and Terza (2001) study on physician advice and drinking. Section 4 summarizes our main findings.

2 The econometric model

We aim to develop a model for a count variable y_i that is function of a dummy variable T_i representing the i-th individual treatment status, with $T_i = 1$ if the individual has been treated and $T_i = 0$ if she has not been treated. The treatment dummy is always observed and, from a theoretical point of view, is a potentially genuine (causal) shifter of the conditional distribution of y_i. We say that T_i is an *endogenous treatment* if treatment status is not random, but there are unobservable individual characteristics affecting T_i that also affect the outcome y_i.

We define a second dummy that represents either a sample selection rule or a participation rule. The second dummy is denoted as P_i when it represents a participation rule and as S_i when it represents a sample selection indicator. Although we will refer to models using individual-level data, the individual i subscript is omitted throughout to simplify notation.

It may be useful to take an example from a specific case to clarify the

[4] The model does not address treatment endogeneity, endogenous participation and sample selection at the same time.

problem. Imagine that we want to study the effect of physician advice on drinking, like in Kenkel and Terza (2001). Physician advice is the treatment of interest, i.e. $T = 1$ if the individual received advice and $T = 0$ otherwise. The outcome of interest y is the number of drinks consumed in a given interval of time. However, not all individuals drunk during this period. Some individuals did not drink because they are non-drinkers and others, although they generally drink, simply consumed zero alcoholic beverages during the time interval of the study. In such a context one may model the first or both types of zeros as 'non-participation' in the drinking activity (where $P = 0$ are non-participants while $P = 1$ are participants) considering zero counts as generated by a different DGP from the one determining strictly positive number of drinks consumed by occasional or frequent drinkers. This approach will make it easy to account for the potential problem of endogenous participation.

In other cases, the most pressing data problem may be that a non-negligible proportion of the interviewed individuals did not responded to the drinking question. In this context, response ($S = 1$) and non-response ($S = 0$) may be correlated with the level of drinking or factors related to drinking and therefore there is potentially a sample selection problem.

In the following subsection we describe the main features of the model addressing endogenous participation while the description of the model with sample selection is reported in Appendix A. Both models address the potential endogeneity of the treatment.

2.1 Endogenous participation: $y = 0$ when $P = 0$

The model with (potentially) endogenous participation considers the case where the dependent count variable y for a given individual is *always* zero if the participation dummy P takes on value zero and *can* be positive or null if $P = 1$. In the example above — the effect of physician advice on drinking — we may have three different types of individuals who did not consume alcohol in a given period of time: non-drinkers, drinkers who only drink occasionally and frequent drinkers. Were the three types of individuals distinguishable (e.g., they declare their type in the survey), a first approach could be to classify occasional and frequent drinkers as participants and non-drinkers as non participants. In this case the count variable for the number of drinks will include also the zeros contributed by occasional or frequent drinkers. However, in case there is not enough information to distinguish between individuals — e.g., they were asked about their level of alcohol consumption without enquiring whether they are non-drinkers, frequent, or occasional drinkers — the natural choice is to pull together all zeros and

consider as participants ($P = 1$) only 'current participants', i.e. those who consumed a positive number of drinks ($y > 0$) during the period of study. We will use this second approach in which consumption for individuals with $P = 1$ is always positive, and $y = 0$ when $P = 0$ as it is more in line with the features of the data used in Kenkel and Terza (2001).[5]

The endogenous treatment is denoted as T. The endogenous treatment and the participation dummies are generated according to a continuous latent variable model:

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + v, \tag{1}$$
$$P^* = \mathbf{r}'\boldsymbol{\theta} + \varphi T + q \tag{2}$$

with $T = 1(T^* > 0)$, $P = 1(P^* > 0)$, and vectors $\mathbf{z}$ and $\mathbf{r}$ represent a set of explanatory variables (including the constant term) with dimension $K_T \times 1$ and $K_P \times 1$, respectively. $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are conformable vectors of coefficients, φ is the coefficient of the treatment dummy in the participation equation, and v and q are residual terms. We assume that the count y is generated according to the following conditional cumulative distribution function,

$$G(y|\eta) \equiv \mathbb{P}(y|\eta) = \begin{cases} \text{not defined} & \text{if } P = 0 \\ [\mu^y \exp(-\mu)] / [1 - \exp(-\mu)] y! & \text{if } P = 1. \end{cases} \tag{3}$$

with,

$$y = \begin{cases} 0 & \text{if } P = 0 \\ 1, 2, \ldots & \text{if } P = 1 \end{cases} \tag{4}$$

and where $\mathbb{P}(.)$ denotes 'probability of,' η is a random variable representing unobserved individual heterogeneity, and $\mu \equiv E[y|\mathbf{x}, T, \eta]$. We use a zero-truncated Poisson distribution for y given $P = 1$. This is done to meet the requirements of the Kenkel and Terza's data, where it is not possible to distinguish among individuals who did not have any drink because they quitted drinking or are trying to quit, and those who just did not have any drink by chance.[6] We use a log-linear model for specifying the conditional mean of y given T, P, and η:

$$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, \tag{5}$$

where, again, vector $\mathbf{x}$ represents a $K_y \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of conformable coefficients, and δ is the coefficient of the treatment

[5] As we will see below, Kenkel and Terza (2001) drop from the analysis 'non-drinkers', i.e. according to their definition those individuals who did not drink in the last 12 months.

[6] Were this piece of information available, a Poisson distribution can be used instead of the zero-truncated Poisson, allowing for two different types of zeros.

dummy in the equation of the main response count y. Finally, correlation between T, P, and y is allowed by imposing some structure on the residuals of equations (1) and (2),

$$\begin{aligned} v &= \lambda_1 \eta + \zeta \\ q &= \lambda_2 \eta + \xi, \end{aligned} \tag{6}$$

where ζ and ξ are 'idiosyncratic' error terms and $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2\} \in \mathbb{R}^2$ are free *factor loadings* to be estimated along the other parameters.

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

$$D(\eta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \zeta, \xi) = D(\eta) \tag{C1}$$
$$D(\zeta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\zeta|\eta) \tag{C2}$$
$$D(\xi|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\xi|\eta) \tag{C3}$$
$$\zeta \perp \xi \mid \eta, \tag{C4}$$

where $D(.)$ stands for 'distribution of.' Condition C1 is the usual random effects assumption, which requires the unobserved individual heterogeneity term η to be independent of all explanatory variables in the system as well as independent of errors ζ and ξ. The conditional independence assumptions in C2 and C3 are weaker than calling for ζ and/or ξ to be independent of the explanatory variables and thus accommodate some limited dependence between control variables and idiosyncratic errors.[7] C2-C3 together ensure the exogeneity of all explanatory variables $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{r}$. Finally, condition C4 requires the idiosyncratic errors to be independent of each other conditional on η. Again, this does not rule out some dependence between ζ and ξ. 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.

The model 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. Note that in this parametrization $\text{Var}(v_i) = (\lambda_1^2 \sigma_\eta^2 + 1)$ and $\text{Var}(q_i) = (\lambda_2^2 \sigma_\eta^2 + 1)$ instead of the usual probit normalization of $\text{Var}(v_i) = \text{Var}(q_i) = 1$. As a consequence,

[7] If claiming independence between all explanatory variables and the unobserved heterogeneity term η is judged untenable for a particular application, instead of requiring condition C1 one could follow Mundlak (1978) and Chamberlain (1980) approach and assume $\eta|\mathbf{w} \sim N(\mathbf{w}'\boldsymbol{\psi}, \sigma_a^2)$, for a vector $\mathbf{w}$ that can contain some elements of $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{r}$ and where $\boldsymbol{\psi}$ is a vector of conformable coefficients. This assumption imposes some restrictions to the way explanatory variables and the unobserved heterogeneity term η can be related (namely, some cross-equation coefficient restrictions) but allows at least some dependence.

coefficients in (1) and (2) will be larger than the usual probit coefficients. After estimation, one can recover the usual probit parametrization multiplying coefficients in (1) and (2) by a factor of $1/\sqrt{\lambda_1^2\sigma_\eta^2+1}$ and $1/\sqrt{\lambda_2^2\sigma_\eta^2+1}$, respectively.

The use of the Poisson distribution for the analysis of count data has been criticized in the past due to the unattractive feature that the conditional mean and the conditional variance are restricted to be equal, a property also known as *equidispersion* (see, for instance, Winkelmann 2008). In contrast, in the present model the introduction of the random term η in the log-linear model for μ allows the count variable y to exhibit *overdispersion* whenever $\sigma_\eta \neq 0$, that is if significant unobserved heterogeneity is detected (see Gourieroux 2000, Miranda and Rabe-Hesketh 2006). However, our model cannot be used to analyze dependent variables which usually exhibit *underdispersion*, such as the number of children.[8]

The correlations between the error terms in y, T^*, and S^* are functions of the factor loadings (λ_1, λ_2) and σ_η^2. In particular, the model implies the following correlations:

$$\rho_{\eta,v} = \frac{\lambda_1\sigma_\eta^2}{\sqrt{\sigma_\eta^2(\lambda_1^2\sigma_\eta^2+1)}} \tag{7}$$

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

$$\rho_{v,q} = \frac{\lambda_1\lambda_2\sigma_\eta^2}{\sqrt{(\lambda_1^2\sigma_\eta^2+1)(\lambda_2^2\sigma_\eta^2+1)}}. \tag{9}$$

The treatment dummy T is an exogenous variable in the main response equation whenever $\rho_{\eta,v}=0$. Similarly, if $\rho_{\eta,q}=0$ participation is exogenous in the main response equation. If $\rho_{\eta,v}=\rho_{\eta,q}=0$, one can obtain consistent estimates of δ on the basis of a simple Poisson regression fitted on the subsample for which $y>0$ (that is 'participants').

Let $\mathbb{P}_P(0|\eta)$ denote the conditional probability of $P=0$ given η and $\mathbb{P}_P(1|\eta)$ the conditional probability of $P=1$ given η. Here, to simplify notation, we do not explicitly write the conditioning on observable variables. In a similar fashion, $\mathbb{P}_T(\tau|\eta)$ represents the probability of $T=\tau$ given η, with $\tau=\{0,1\}$. Finally, denote by $G(y|\eta)$ the cumulative distribution of y

[8] To the knowledge of the authors no method has been suggested in the literature that could deal with underdispersed count data and either sample selection or an endogenous treatment effect, let alone the two problems together.

given η which is defined by equations (3) and (5) together. The log-likelihood function is then:

$$\begin{aligned}\log(L) = & \sum_{i,P_i=0}\sum_{\tau}\omega_\tau \ln\left\{\int \mathbb{P}_P(0|\eta)\mathbb{P}_T(\tau|\eta)\phi(\eta)d\eta\right\} \\ & + \sum_{i,P_i=1}\sum_{\tau}\omega_\tau \ln\left\{\int \mathbb{P}_P(1|\eta)\mathbb{P}_T(\tau|\eta)G(y|\eta)\phi(\eta)d\eta\right\},\end{aligned} \tag{10}$$

where $\phi(\cdot)$ is the density of a normal variate with mean zero and variance σ_η^2, $\omega_0 = 1(T = 0)$, $\omega_1 = 1(T = 1)$ and $1(.)$ is the indicator function. We will refer to this model as to the the Endogenous Participation Endogenous Treatment (EPET) Poisson model.

The integrals in equation (10) do not have a closed form solution and must be numerically evaluated. We use MSL (for a detailed discussion on MSL, see, Train 2003). To evaluate the integrals we use Halton sequences instead of uniform pseudorandom sequences. Halton draws have been shown to achieve high precision with fewer draws than uniform pseudorandom sequences because they have a better coverage of the [0, 1] interval. A modified Newton-Ramphon algorithm is used for maximization, using analytical first derivatives and numerical second derivatives. At convergence Eicker-Huber-White robust standard errors are computed.

The use of a common latent factor structure like the one written in (6) has four main advantages over the alternative of specifying a multivariate normal distribution for v, q, and η (see Deb and Trivedi 2006). First, the common latent variable approach can be used quite flexibly to combine appropriately chosen conditional and marginal distributions that generate the joint distribution that the researcher wants to use. Second, latent factors have a natural interpretation as proxies for unobserved covariates since they enter into the equations in the same way as observed covariates. The factor loadings can therefore be interpreted in much the same way as coefficients on observed covariates can. Third, it provides a parsimonious representation of error correlations in models with a large number of equations. Related to this, and quite importantly for computational feasibility, the latent variable approach transforms a problem in which calculation of the log-likelihood involves the computation of a three dimensional integral into a problem where only a one dimensional integral needs to be computed.

The cost of using the one (latent) factor structure is that it implies some variance-covariance restrictions.[9] In our case, there are three implicit

[9]See for instance, Carneiro et al. (2003). The authors note that these restrictions can have an arbitrary content and conclude that it is important to appeal to economic theory to justify any specific identification scheme. Latent factor models are very popular in health economics; for some recent examples of normally distributed one-latent factor models see Kenkel and Terza (2001), Terza et al. (2008) and Terza (2009).

variance-covariance restrictions as the six elements of the variance-covariance matrix must be recovered from only three parameters (λ_1, λ_2 and σ^2_η).[10] Hence, the model we propose is especially useful in cases in which the endogeneity of both the treatment and the participation decisions are likely to be determined by a common unobservable variable entering all error terms. In our early example, we might think of an individual's intertemporal discount rate. Individuals with a high intertemporal discount rate are more likely to engage in health damaging activities such as drinking, to be in bad health, and to receive drinking advice when seeing a doctor. In particular, they are both more likely to drink and to drink a high number of drinks. To put it simply, the omission of a single unobservable which enters all three equations is likely to generate a positive correlation between all three error terms. The model is also suitable to cases in which two or more highly correlated unobservable variables enter all equations. In such a case using a one latent factor model approach may be a good strategy because after conditioning on a single *summary* latent factor η, the remaining components of the error terms in the three equations can be expected to be approximately orthogonal (see condition C4).

In Appendix B we investigate the finite sample performance of the EPET-Poisson model and of the Endogenous Treatment Poisson (ET-Poisson) model, a model in which only the endogeneity of the treatment status (but not of participation) is addressed, in a Monte Carlo simulation study.

3 An application to the effect of physician advice on drinking

In this section we apply the EPET-Poisson model to the problem of estimating the treatment effect of physician advice on alcohol consumption using data from Kenkel and Terza (2001).

[10] Let us define the variance-covariance matrix of the error terms as:

$$V = \begin{bmatrix} \sigma^2_\eta & \sigma_{\eta,v} & \sigma_{\eta,q} \\ & \sigma^2_v & \sigma_{v,q} \\ & & \sigma^2_q \end{bmatrix}. \tag{11}$$

The three restrictions are (i) $\sigma^2_v = \lambda^2_1\sigma^2_\eta + 1$, (ii) $\sigma^2_q = \lambda^2_2\sigma^2_\eta + 1$, and (iii) $\sigma_{v,q} = (\sigma_{\eta,v} \cdot \sigma_{\eta,q})/\sqrt{\sigma^2_v\sigma^2_q}$ (or $\rho_{v,q} = \rho_{\eta,v} \cdot \rho_{\eta,q}$).

3.1 The Kenkel and Terza's study

For the purpose of illustration, we use the same data — the 1990 National Health Interview Survey — and adopt the same empirical specification and exclusion restrictions used by Kenkel and Terza (2001).[11] Our aim is to show how the estimates of treatment effects are sensitive to various assumptions about endogeneity of treatment status and participation, and which model better fits the data. In the Kenkel and Terza's study drinking is measured as the number of drinks consumed in the last two weeks.[12] Physician advice about drinking is built from respondents' answers to the following question: 'Have you ever been told by a physician to drink less?'

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, they focus only on men who have drunk alcohol at least once in the last 12 months ('drinkers') and report having been told at some time that they had high blood pressure.[13] In spite of this, Kenkel and Terza observe in their sample that 21% of drinkers (according to their definition) did not drink at all in the last two weeks. Various reasons may be behind the excess of zeros. First, it could be that zeros are contributed by recent quitters or people who were actively trying to quit drinking in the last 12 months. Second, $y = 0$ could also be contributed by individuals who drink only in very special occasions such as weddings, birthdays, or New Year's Eve (occasional drikers). Finally, $y = 0$ could be contributed by 'frequent' drinkers who, by chance, did not drink any alcohol in the past two weeks; although this last scenario is less likely as two weeks are a period long enough to expect a strictly positive number of drinks to be consumed by 'frequent' drinkers. Unfortunately, individuals were asked only their level of alcohol consumption in the last two weeks without enquiring whether they classify themselves as quitters, frequent, or occasional drinkers. As a consequence, we cannot separate the

[11] Although the model is formally identified by covariance restrictions and functional form, Kenkel and Terza (2001) provide an additional source of identification 'through exclusion restrictions involving a set of eleven variables related to health insurance status, physician contacts, and health problems' (p. 176). The plausibility of these restrictions is discussed by the authors.

[12] Kenkel and Terza states 'This is calculated 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).

[13] A potential problem with this sample selection is that the decision to quit drinking may be affected by health status, so that in the sample one is likely to observe only 'healthy' drinkers. Kenkel and Terza argue that this is likely to induce only a small selection bias as in the the National Health Interview Survey only 12% of individuals declare not to drink because of health problems.

different types of zeros.

Clearly, the fact that a good proportion (perhaps the majority) of these zeros are likely to be contributed by occasional drinkers and quitters suggests that the excess zeros cannot be ignored. The authors acknowledge this and account for the excess zeros by using a flexible functional form for the conditional mean of drinking based on the inverse Box-Cox transformation. An alternative way of addressing this issue, which we follow here, is to treat the zeros and the positive drinking outcomes as if they were generated by two separate DGPs (see Terza 1998). More details on the data and the covariates used are available in the original study. Table 1 reports the definitions and the means of all the variables, which match the corresponding means in Kenkel and Terza (2001). The Table also provides information on which variables are used to identify the model (the exclusion restrictions).

3.2 Results

3.2.1 The effect of the treatment

In this section we focus only on the effect of the treatment of interest (physician advice). As we said, we use the same specifications (and exclusion restrictions) as the original article for the treatment and the drinking intensity equations.

The first column of Table 2 reports the marginal effects from a Poisson model where the potential endogeneity of the treatment — physician advice — is not addressed.[14]The results are similar to those reported by Kenkel and Terza in the models where physician advice is considered exogenous (see Table III in their article): advice appears to have a counterintuitive *positive* effect on drinking that is statistically significant at 1%. Column (2) reports the marginal effects of physician advice on the probability of drinking obtained from a simple probit model, and also in this case, advice turns out to be positively correlated with drinking. Column (3) reports the marginal effects when the potential endogeneity of advice is taken into account but endogenous participation is neglected using the ET-Poisson model. This model assumes that both zeros and positive y outcomes are produced by the same DGP but accounts for the endogeneity of the treatment.[15] This

[14]In analogy to Kenkel and Terza (2001), marginal effects are evaluated at the *median* value of the dependent variable.

[15]More details are available upon request from the authors. The model consists of two equations, the drinking equation in which both zero and positive consumption are modelled through a Poisson model with unobserved heterogeneity and a physician advice equation. The model is estimated using MSL, using a one latent factor structure and normality of the error terms.

model (and the following models estimated using MSL) was estimated using 1,600 Halton draws.[16] First, note that the correlation between the errors in the drinking intensity and the physician advice equation $\rho_{\eta,v}$ is positive, as expected, and statistically significant at 1%. Hence, advice T is endogenous with respect to drinking y. In other words, individuals who have a higher latent propensity to drink are also more likely to receive advice. Second, the marginal effect of physician advice turns out to be negative, statistically significant, and amounts to a bit less than $-5\frac{1}{2}$ (-5.4) drinks per two weeks. Both results suggest that the positive effect of T on y that is reported by the Poisson model with exogenous treatment is *spuriously* driven by a positive bias which results from the fact that individuals endogenously sort themselves into the treatment. In other words, those receiving advice were also the heaviest drinkers.

When the intensive margin (i.e., drinking participation) and the extensive margin (i.e., number of drinks conditional on strictly positive drinking) are allowed to be generated by different DGPs with the EPET-Poisson model in column (5), the effect of physician advice falls in absolute value by more than one drink per week, to -4.1 (+25%), and remains highly statistically significant.[17] Physician advice turns out to be endogenous with respect to drinking participation and $\rho_{v,q}$ is positive, which is consistent with the correlation found between advice and drinking intensity (i.e., $\rho_{\eta,v} > 0$) and our theoretical predictions at the end of subsection 2.1. It is also important to notice that the EPET-Poisson model shows that physician advice has no effect on the likelihood of drinking (column (4)), a result in sharp contrast with the one obtained from the simple probit model in which the positive association between drinking and physician advice was generated by unobserved heterogeneity. Comparison of the ET and the EPET-Poisson models allows a better understanding of the effect of physician advice, which does not seem to induce people not to drink but simply to cut their two-week drinking. This could be explained by the presence in the population of light drinkers and heavy drinkers. The first may not quit, following physician advice, since they do not believe that their limited drinking is damaging their health, while the second may not quit simply because they have higher levels of addiction.

[16] Using 2,000 Halton draws produced only negligible changes in coefficients and standard errors of the estimates.

[17] We included the same set of controls both in the drinking participation and in the drinking intensity equations. The same is done, for instance, in Terza et al. (2008). In general, unlike in the sample selection model in which there might be specific factors affecting non-response but not necessarily affecting intensity of consumption, in the case of endogenous participation it is hard to think of variables affecting the intensive or the extensive margin only.

Hence, physician advice may have an effect only on the drinking intensive margin and not on drinking prevalence.

Table 2 shows the presence of significant unobserved heterogeneity in the three choices (as σ_η^2 is statistically different from zero). The last two lines of the table also clearly suggest that the EPET Poisson model is obviously much better at predicting the zeros and that it fits the data better than the two alternative models (Poisson and ET-Poisson) as indicated by the lowest value of the Bayesian information criterion (BIC).

Two notes are worth of mentioning. Firstly, in Kenkel and Terza's specific case pooling the intensive and the extensive margins and forcing the DGPs to be the same for the two choices is not very harmful because the effect of the treatment on the two outcomes goes in the same direction, although our estimates suggest that only the intensive margin is significantly affected by physician advice. Furthermore, the correlations between unobservables in the endogenous treatment and drinking intensity, and between the unobservables in the endogenous treatment and the endogenous participation, all have the same sign. Secondly, Kenkel and Terza use a flexible functional form — the non-linear inverse Box-Cox form — that, although imposing the same DGP for the intensive and extensive drinking margins, produces a marginal effect of the treatment (about -4.5 drinks) that is somewhat between the one reported by a model that only deals with the endogenous treatment problem and the one obtained from a model that deals with both endogenous treatment and endogenous participation (the EPET-Poisson model). Clearly, in other applications the consequences of neglecting endogenous participation may be more substantial.

In order to have an idea of the goodness of the exclusion restrictions, Table 3 reports the marginal effects for the physician advice equation and Wald tests for the variables identifying the model over and above functional form and covariance restrictions. In both the ET and the EPET-Poisson models Wald tests suggest that the "identifying" variables are highly statistically significant and the model is unlikely to suffer from weak identification.

3.2.2 The effect of other covariates

The main advantage of a model not imposing the same DGPs on the intensive and the extensive drinking margins is that not only the effect of the treatment but also that of other covariates are allowed to differ across the two choices. Think of, for instance, the effect of parental supervision, or strictness of parenting styles, on youngsters' smoking. In this case, parenting style is likely to affect the likelihood of smoking participation but it is rather unlikely to affect the quantity of cigarettes smoked given participation. Similarly, alcohol

(cigarette) taxes and prices are more likely to affect the quantity of drinks (cigarettes) consumed than the drinking (smoking) prevalence itself. Table 4 reports the marginal effects (at the sample mean) of the other covariates estimated in the EPET-Poisson model, and shows which is the relevant margin (intensive, extensive or both) affected by the regressors. Just to take a few examples, it is interesting to notice that years of education are positively associated with the probability of drinking but negatively associated with the average number of drinks consumed. Individuals in their forties and fifties drink less on average, but this effect is entirely accounted for by their lower probability of drinking. We also reported the marginal effects obtained from the ET-Poisson model for the sake of completeness.

4 Concluding remarks

In this paper we have proposed a Full Information Maximum Likelihood estimator for count data models with endogenous treatment effects and either sample selection or endogenous participation, which is implemented using maximum simulated likelihood. Sample selection occurs when the main outcome is missing for some individuals and the data are not missing at random. In contrast, endogenous participation occurs when participation into an activity (e.g., smoking or drinking) and the intensity of the activity are produced by two different, but correlated, DGPs.

For illustrative purposes, we have applied our proposed estimator to the Kenkel and Terza (2001)'s data on physician advice and drinking. Our estimates suggest that in these data (i) neglecting treatment endogeneity leads to a perversely signed effect of physician advice on drinking intensity, (ii) neglecting endogenous participation leads to an upward biased estimate of the treatment effect of physician advice on drinking intensity.

Appendix A. Endogenous sample selection: y missing when $S = 0$

The model with (potentially) endogenous sample selection considers the case where the dependent count variable y for a given individual is missing if the selection dummy S takes on value zero and is observed if the selection dummy takes on value one. The endogenous treatment is denoted as T. The endogenous treatment and the selection dummies are generated according to

a continuous latent variable model:

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + v, \tag{12}$$
$$S^* = \mathbf{r}'\boldsymbol{\theta} + \varphi T + q \tag{13}$$

with $T = 1(T^* > 0)$, $S = 1(S^* > 0)$, and vectors $\mathbf{z}$ and $\mathbf{r}$ represent a set of explanatory variables (including the constant term) with dimension $K_T \times 1$ and $K_S \times 1$, respectively. $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are conformable vectors of coefficients, φ is the coefficient of the treatment dummy in the selection equation, and v and q are residual terms. We assume that the count y is generated according to the following conditional cumulative distribution function,

$$F(y|\eta) \equiv \mathbb{P}(y|\eta) = \begin{cases} \text{not defined} & \text{if } S = 0 \\ \left[\mu^y \exp(-\mu)\right]/y! & \text{if } S = 1. \end{cases} \tag{14}$$

with,

$$y = \begin{cases} \text{missing} & \text{if } S = 0 \\ 0, 1, 2, \ldots & \text{if } S = 1, \end{cases} \tag{15}$$

$$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, \tag{16}$$

and all other remaining aspects of the model are the same as in subsection 2.1. The main difference from the model presented here and the one in subsection 2.1 is the fact that here we use a Poisson distribution for y given $S = 1$ whereas we used a zero-truncated Poisson for y given $P = 1$ in the endogenous participation model. This is a minor modification that reflects the fact that in the endogenous participation model we considered the $y = 0$ count as being generated by a different data generating mechanism from $y > 0$ counts.

Another important difference is the fact that now only individuals with $S = 1$ in the sample will contribute a non missing observation for y. The likelihood function is now written as follows:

$$\begin{aligned} \log(L) = & \sum_{i,S_i=0} \sum_{\tau} \omega_\tau \ln\left\{\int \mathbb{P}_S(0|\eta)\mathbb{P}_T(\tau|\eta)\phi(\eta)d\eta\right\} \\ & + \sum_{i,S_i=1} \sum_{\tau} \omega_\tau \ln\left\{\int \mathbb{P}_S(1|\eta)\mathbb{P}_T(\tau|\eta)F(y|\eta)\phi(\eta)d\eta\right\}. \end{aligned} \tag{17}$$

Again, the model can be estimated using MSL.

Appendix B. Monte Carlo simulation study

In this appendix we investigate the finite sample performance of the EPET-Poisson and the ET-Poisson models in a Monte Carlo simulation study. The

aim of the analysis is twofold: (1) to show that the EPET-Poisson estimator can recover the true population parameters when the DGP is indeed EPET-Poisson, and to offer evidence that it is approximately normally distributed in finite samples of moderate size; (2) to show how large the bias is when the researcher fits a ET-Poisson model to data that are truly generated by a EPET-Poisson process and how misleading inference can be if the analyst assumes that the misspecified ET-Poisson estimator is asymptotically normality distributed. Given the focus of the study, all experiments discussed here make assumptions that are favorable to the EPET-Poisson estimator.

Simulations are performed using a total sample size of 1,500 observations. There are three exogenous variables: x_1, x_2, and x_3. These variables are distributed as independent standard normal variates. The main count response is denoted by y, the endogenous treatment by T, and the participation dummy by P. To facilitate identification, we specify strong exclusion restrictions. Namely: (1) x_1 only enters the equation of T; (2) x_2 only enters the equation of P; and (3) x_3 only enters the equation of y. We allow the endogenous treatment T to enter both the participation equation P and the main count response y. The DGP of the simulated data is EPET-Poisson in each replication, with the true population parameters as given in Table 5. In all experiments $\eta \sim N(0, 0.1)$.

The values of the parameters in Table 5 and the variance for η were chosen so that the resulting simulated count response variable y will have approximately a mean and a standard deviation which are near those exhibited by the Kenkel and Terza (2001) drinking data. Also, these values ensure that approximately 70% of the sample in each replication will have $P = 1$ and $y > 0$ (participants) and 30% will have $P = 0$ and $y = 0$ (non participants). Again, this was chosen to more or less match the characteristics of the Kenkel and Terza drinking data where 21% of the sample have $y = 0$. We decided to have $E(P)$ slightly higher to 0.21 to compensate for the smaller sample size that is used in the simulations. The true population parameters imply that approximately half of the sample is subject to treatment ($T = 1$) in each replication. In the Kenkel and Terza drinking data the mean of advice is slightly lower, namely 0.27. Here, again, we set mean of T to 0.5 to compensate for the smaller sample size used in the simulations. In all experiments the noise/signal ratios are approximately: (a) 0.18 in the T equation; (b) 0.19 in the P equation; and (c) 0.16 in the y equation.

We would like to test that the EPET-Poisson estimator can recover the true population parameters regardless the sign that $\rho_{\eta,v}$, $\rho_{\eta,q}$, and $\rho_{v,q}$ take. We setup two groups of experiments, which only differ in the size of the correlation coefficients. Within each group the only thing that varies is the sign of the aforementioned correlation coefficients:

- Experiment 1: $\rho_{\eta,v} = 0.35$, $\rho_{\eta,q} = 0.35$, and $\rho_{v,q} = 0.13$
- Experiment 2: $\rho_{\eta,v} = -0.35$, $\rho_{\eta,q} = 0.35$, and $\rho_{v,q} = -0.13$
- Experiment 3: $\rho_{\eta,v} = 0.35$, $\rho_{\eta,q} = -0.35$, and $\rho_{v,q} = -0.13$
- Experiment 4: $\rho_{\eta,v} = -0.35$, $\rho_{\eta,q} = -0.35$, and $\rho_{v,q} = 0.13$
- Experiment 5: $\rho_{\eta,v} = 0.78$, $\rho_{\eta,q} = 0.78$, and $\rho_{v,q} = 0.61$
- Experiment 6: $\rho_{\eta,v} = -0.78$, $\rho_{\eta,q} = 0.78$, and $\rho_{v,q} = -0.61$
- Experiment 7: $\rho_{\eta,v} = 0.78$, $\rho_{\eta,q} = -0.78$, and $\rho_{v,q} = -0.61$
- Experiment 8: $\rho_{\eta,v} = -0.78$, $\rho_{\eta,q} = -0.78$, and $\rho_{v,q} = 0.61$.

Maximum simulated likelihood (MSL) estimators are asymptotically equivalent to Maximum Likelihood estimators as long as R, the number of draws used to evaluate the simulated likelihood, grows at a faster rate than the square-root of the sample size $\sqrt{N}$ (Gourieroux and Monfort 1993). Simulation delivers an unbiased approximation of the likelihood $L(.)$. However, MSL maximizes $\ln[\widehat{L}(.)]$ rather than $\widehat{L}(.)$; and simulation of $\ln[(\widehat{L}(.)]$ does not give an unbiased estimator for $\ln[L(.)]$. This bias does not affect the consistency of the MSL estimator. However, to achieve asymptotic normality it is required that $\sqrt{N}/R \to 0$. If this condition does not hold the covariance matrix estimator of MSL is incorrect; see Cameron and Trivedi (2005, p. 394-396) and Train (2003, p. 258-259). For this reason the analyst is advised to use a *large enough* number of draws R. There is no general rule to choose R and how large it should be depends on particular applications. In the case of the EPET-Poisson, preliminary Monte Carlo simulation experiments with 200 replications showed that MSL estimators using less than $1,000$ Halton draws deliver slightly underestimated standard errors and that the nominal coverage of the parameters were below the advertised 95%. Once the number of Halton draws R was set to values larger than $1,000$ the EPET-Poisson behaved well.

Tables 6 and 7 present Monte Carlo simulation results for experiments 1 to 4. In all these experiments we used $1,600$ Halton draws. Table 6 reports the performance of EPET-Poisson and ET-Poisson in terms of bias and Monte Carlo standard deviation (SD). Results show that, as expected, the EPET-Poisson delivers largely unbiased estimators of all the parameters. In fact, for all parameters, percentage bias is below 3.1%. In all four experiments the computed bias for the coefficient of main interest b_{yT} (i.e., the coefficient of the treatment T on the main count response y) is not statistically different from zero at a significance level of 5%.

In the case of the ET-Poisson model the reported bias is large with respect to the size of the parameters. To give a relative idea of the size of the bias, the reader can consider that the percentage bias of b_{yT} in ET-Poisson is 76% in experiment 1 and 78% in experiment 4. This bias is significantly different from zero at 5% in all the four experiments.

Moving to Table 7 the reader can see the performance of the EPET-Poisson and the ET-Poisson in terms of standard error and nominal coverage. From this table the reader can see that the EPET-Poisson estimator delivers an average standard error that is broadly equal to the Monte Carlo standard deviation. In other words, the finite sample variation of the EPET-Poisson estimator is broadly the same as the variation one expects under the assumption that the EPET-Poisson estimator is normally distributed. This last observation is also supported by the fact that nominal coverage of all parameters estimated by EPET-Poisson achieve approximately the advertised size of 95%. This is true across all four experiments. In conclusion, there is strong evidence that EPET-Poisson is asymptotically normally distributed as expected.

Performance of the ET-Poisson in terms of standard error and nominal coverage is disappointing. Standard errors are largely underestimated and nominal coverage is far from the advertised 95%. In fact, the true value of the population parameter for the main coefficient of interest, b_{yT}, lied outside the 95% confidence interval in every single replication and across all four experiments.

Experiments 5 to 8 (see tables 8 and 9) give similar results and for the sake of brevity we do not add any further comment on those experiments.[18] Summarising, we find that if the DGP is EPET-Poisson: (1) the EPET-Poisson estimator is consistent and asymptotically normally distributed; (2) fitting a ET-Poisson delivers a badly biased estimator and underestimated standard errors. Hence, performing inference on the basis of ET-Poisson when the DGP is EPET-Poisson may lead to severely misleading inference.

[18] The only thing to note is that the standard errors for $\widehat{\rho}_{v,q}$ are 'slightly' larger than the calculated Monte Carlo standard deviations. This overestimation, however, is relatively small with respect to the size of $\widehat{\rho}_{v,q}$ and it does not affect the nominal coverage in any case, which remains around 94%. Correlation coefficients are special in the sense that having to lie between -1 and 1 they cannot truly be distributed as normal variates. Also, $\widehat{\rho}_{v,q}$ is a function of λ_1, λ_2, and σ_η and its standard error is calculated using the Delta method. If $\widehat{\rho}_{v,q}$, as a function of λ_1, λ_2, and σ_η is approximately quadratic around the ML estimates, then a second order approximation of the standard errors for $\widehat{\rho}_{v,q}$ will be precise. However, if $\widehat{\rho}_{v,q}$ is not quadratic enough, then calculation of the standard error of $\widehat{\rho}_{v,q}$ by the Delta method may be slightly off target.

References

Cameron, A., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge university press.

Carneiro, P., Hansen, K., Heckman, J., 2003. Estimating distributions of treatment effects with an aplication to the returns to schooling and measurement of the effects of uncertainty on college choice. International Economic Review 44 (2), 361–422.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The Review of Economic Studies 47 (1), 225–238, i.

Cragg, J., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39 (5), 829–844.

Deb, P., Trivedi, P., 2006. Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: applications to health care utilization. Econometrics Journal 9, 307–331.

Gourieroux, C., 2000. Econometrics of Qualitative dependent variables. Cambridge university press.

Gourieroux, C., Monfort, A., 1993. Simulation-based inference: A survey with special reference to panel data models. Journal of Econometrics 59, 5–33.

Greene, W., 1997. Fiml estimation of sample selection models for count data. Stern School of Business, New York University. Manuscript.

Greene, W., 2009. Models for count data with endogenous participation. Empirical Economics 36 (1), 133–173.

Jones, A., 1989. A double-hurdle model for cigarette consumption. Journal of Applied Econometrics 4 (1), 23–39.

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., 2009. Impact of prescription drug coverage on drug expenditure of the elderly — evidence from a two-part model with endogeneity. mimeoPaper available at http://mypage.iu.edu/~trivedi/drug_all_0609.pdf.

Melkersson, M., Rooth, D., 2000. Modeling female fertility using inflated count data models. Journal of Population Economics 13 (2), 189–203.

Miranda, A., 2004. FIML estimation of an endogenous switching model for count data. The Stata Journal 4 (1), 40–49.

Miranda, A., Rabe-Hesketh, S., 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. Stata Journal 6 (3), 285–308.

Mullahy, J., 1986. Specification and testing of some modified count data models. Journal of Econometrics 33, 341–365.

Mullahy, J., 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. Journal of Health Economics 17 (3), 247–281.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

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 (18), 387–405.

Terza, J., 1985. A tobit-type estimator for the censored poisson regression model. Economics Letters 18, 361–365.

Terza, J., 1998. Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. Journal of Econometrics 84, 129–154.

Terza, J. V., 2009. Parametric nonlinear regression with endogenous switching. Econometric Reviews 28 (6), 555–580.

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.

Train, K., 2003. Discrete choice methods with simulation. Cambridge university press.

Windmaijer, F., Santos Silva, J., 1997. Endogeneity in count data models: An application to demand for health care. Journal of Applied Econometrics 12 (3), 281–94.

Winkelmann, R., 1998. Count data models with selectivity. Econometric Reviews 17, 339–359.

Winkelmann, R., 2008. Econometric Analysis of Count Data, 5th Edition. Springer, Berlin Heidelberg, New York.

Table 1. Variable definitions and descriptive statistics

Variable name	Definition	mean	S.D.
Dependent variable			
y	Total drinks last two weeks	14.697	22.753
Treatment variable			
T	Physician advice to reduce drinking	0.278	-
Control variables[(a)]			
EDITINC	Income ($1,000)	2.575	5.008
AGE30	$30 < \text{age} \leq 40$	0.180	-
AGE40	$40 < \text{age} \leq 50$	0.195	-
AGE50	$50 < \text{age} \leq 60$	0.182	-
AGE60	$60 < \text{age} \leq 70$	0.199	-
AGEGT70	$\text{Age} > 70$	0.122	-
EDUC	Years of schooling	12.925	3.087
BLACK	Black d.v.	0.133	-
OTHER	Non-white	0.018	-
MARRIED	Married	0.645	-
WIDOW	Widowed	0.052	-
DIVSEP	Divorced or separated	0.160	-
EMPLOYED	Employed	0.666	-
UNEMPLOY	Unemployed	0.029	-
NORTHE	Northeast	0.217	-
MIDWEST	Midwest	0.275	-
SOUTH	South	0.295	-
Excluded variables[(b)]			
MEDICARE	Insurance through Medicare	0.252	-
MEDICAID	Insurance through Medicaid	0.031	-
CHAMPUS	Military insurance	0.059	-
HLTHINS	Health insurance	0.815	-
REGMED	Reg. source of care	0.821	-
DRI	See same doctor	0.721	-
MAIORLIM	Limits on major daily activity	0.086	-
SOMELIM	Limits on some daily activity	0.077	-
HVDIAB	Have diabetes	0.061	-
HHRTCOND	Have heart condition	0.146	-
HADSTROKE	Had stroke	0.036	-

[(a)] These are the variables included in both the main equation (and the endogenous participation equation) and the endogenous treatment equation;

[(b)] These are the variables only included in the endogenous treatment equation.

Note. This table reports the definitions, the means and the standard deviations (S.D.) for the variables used in Kenkel and Terza (2001). Data refer to the 1990 National Health Interview Survey. The estimation sample includes 2,467 observations.

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	EPET-Poisson	
	$y^{(a)}$	$Pr(y>0)^{(b)}$	$y^{(a)}$	$Pr(y>0)^{(b)}$	$y>0^{(a)}$
	(1)	(2)	(3)	(4)	(5)
Physician advice (T)	3.679***	0.079***	-5.395***	-0.045	-4.072***
	(.558)	(.017)	(.386)	(.049)	(.864)
$\widehat{\rho}_{\eta,v}$			0.832***	0.689***	
			(.029)	(.092)	
$\widehat{\rho}_{\eta,q}$				0.378***	
				(.088)	
$\widehat{\rho}_{v,q}$				0.261***	
				(.084)	
$\widehat{\sigma}^2_{\eta}$			2.190***	1.456***	
			(.069)	(.099)	
N.obs.	2,467	2,467	2,467	2,467	
Log-likelihood	-32,263	-1,247	-10,184	-10,062	
$\widehat{Pr(y=0)}^{(c)}$	0.00		0.11	0.22	
$BIC^{(d)}$	64,674.9		20,759.3	20,671	

*** significant at 1%. Eicker-Huber-White robust standard errors in parentheses.

$^{(a)}$ Marginal effects are computed at the sample median of the dependent variable, in analogy to Kenkel and Terza (2001);

$^{(b)}$ Marginal effects are computed at the sample mean of the independent variables.

$^{(c)}$ Probability of the zero outcome predicted by the model.

$^{(d)}$ Bayesian information criterion.

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 and $y>0$ the number of alcoholic drinks consumed in the last two weeks conditional on drinking. Estimation refers to the 1990 National Health Interview Survey with the sample selection and covariates used in Kenkel and Terza (2001). T and P are dichotomous indicators of individual treatment status and participation to the drinking activity, respectively. ET and EPET stand for Endogenous Treatment and Endogenous Participation Endogenous Treatment, respectively. Both 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).

Table 3 Marginal effects of covariates on the physician advice equation

	ET-Poisson		EPET-Poisson	
EDITINC	0.000	(0.002)	0.000***	(0.000)
AGE30	0.078**	(0.035)	0.098***	(0.035)
AGE40	0.041	(0.035)	0.055	(0.035)
AGE50	0.023	(0.034)	0.039	(0.036)
AGE60	0.022	(0.040)	0.047	(0.041)
AGEGT70	0.041	(0.050)	0.057	(0.051)
EDUC	-0.009***	(0.003)	-0.010***	(0.000)
BLACK	0.108***	(0.029)	0.106***	(0.026)
OTHER	0.089	(0.063)	0.079	(0.063)
MARRIED	0.047*	(0.026)	0.037	(0.026)
WIDOW	0.100**	(0.049)	0.083*	(0.049)
DIVSEP	0.102***	(0.034)	0.080**	(0.033)
EMPLOYED	-0.006	(0.027)	-0.013	(0.027)
UNEMPLOY	0.084	(0.055)	0.042	(0.049)
NORTHE	0.023	(0.025)	0.030	(0.026)
MIDWEST	-0.021	(0.024)	-0.009	(0.024)
SOUTH	-0.017	(0.024)	-0.006	(0.024)
Excluded variables				
MEDICARE	0.014	(0.045)	0.011	(0.043)
MEDICAID	0.006	(0.033)	0.016	(0.037)
CHAMPUS	-0.049**	(0.022)	-0.069***	(0.024)
HLTHINS	0.041*	(0.025)	0.065**	(0.026)
REGMED	0.011	(0.021)	-0.006	(0.025)
DRI	0.051	(0.034)	0.008	(0.031)
MAIORLIM	0.011	(0.027)	0.000	(0.028)
SOMELIM	0.108***	(0.034)	0.085**	(0.034)
HVDIAB	0.063***	(0.023)	0.068***	(0.023)
HHRTCOND	0.029	(0.036)	0.059	(0.040)
HADSTROKE	-0.009	(0.029)	-0.020	(0.030)
F-test excluded variables	49.22 [0.00]		53.8 [0.00]	

*** significant at 1%; ** significant at 5%; * significant at 10%. Eicker-Huber-White robust standard errors in parentheses, p-values in brackets.

[a] Variables excluded from the drinking equations for (economic) identification;

Note. The dependent variable is the probability of receiving physician advice of not drinking. The table reports the marginal effects at the sample means of covariates included in the physician advice equation and the F-tests for the exclusion of the identifying variables in the ET and the EPET-Poisson models.

Table 4. Marginal effects of other covariates

	ET-Poisson		EPET-Poisson			
	y		$Pr(y>0)$		$y>0$	
EDITINC	0.047	(0.040)	0.001**	(0.001)	0.074**	(0.030)
AGE30	0.961	(0.735)	-0.027	(0.015)	1.298	(0.869)
AGE40	-0.158	(0.683)	-0.049***	(0.016)	0.101	(0.833)
AGE50	-0.991*	(0.520)	-0.063***	(0.018)	-1.071	(0.839)
AGE60	-0.763	(0.646)	-0.025	(0.018)	-1.396*	(0.771)
AGEGT70	-1.192*	(0.651)	0.001	(0.020)	-2.140***	(0.739)
EDUC	-0.101	(0.067)	0.008***	(0.002)	-0.324***	(0.086)
BLACK	0.298	(0.663)	0.014	(0.013)	-0.930	(0.600)
OTHER	-1.068	(0.891)	0.042	(0.026)	-3.215***	(1.093)
MARRIED	0.024	(0.541)	-0.006	(0.012)	-0.500	(0.609)
WIDOW	2.396**	(1.087)	0.011	(0.022)	1.308	(1.482)
DIVSEP	2.061**	(0.848)	0.009	(0.015)	2.103**	(0.978)
EMPLOYED	0.213	(0.538)	0.072***	(0.013)	-0.464	(0.711)
UNEMPLOY	5.240***	(1.972)	0.092***	(0.017)	3.286*	(1.887)
NORTHE	-0.192	(0.516)	-0.034***	(0.013)	0.060	(0.783)
MIDWEST	-1.284***	(0.474)	-0.046***	(0.012)	-0.965	(0.703)
SOUTH	-1.103**	(0.468)	-0.043***	(0.012)	-0.864	(0.740)

*** significant at 1%; ** significant at 5%; * significant at 10%. Eicker-Huber-White robust standard errors in parentheses.

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 and $y>0$ the number of alcoholic drinks consumed in the last two weeks conditional on drinking. Marginal effects are computed at the sample means of the independent variables.

Table 5. Monte Carlo Simulations: True population parameters

parameter	T	P	y
constant	0	0.5	1.56
T	-	1.5	1
x_1	2.5	0	0
x_2	0	-1.9	0
x_3	0	0	-0.8

Table 6. Montecarlo simulations: Bias and SD

Parameters	$\rho_{\eta,v}=0.35$ $\rho_{\eta,q}=0.35$ $\rho_{v,q}=0.13$		$\rho_{\eta,v}=-0.35$ $\rho_{\eta,q}=0.35$ $\rho_{v,q}=-0.13$		$\rho_{\eta,v}=0.35$ $\rho_{\eta,q}=-0.35$ $\rho_{v,q}=-0.13$		$\rho_{Ty}=-0.35$ $\rho_{\eta,q}=-0.35$ $\rho_{v,q}=0.13$	
	bias	SD	bias	SD	bias	SD	bias	SD
EPET-Poisson								
$\widehat{\rho}_{\eta,v}$	0.0036	0.0967	0.0047	0.0978	0.0043	0.0979	0.0057	0.0998
$\widehat{\rho}_{\eta,q}$	0.0040	0.1333	-0.0039	0.1311	-0.0013	0.1247	-0.0045	0.1292
$\widehat{\rho}_{v,q}$	0.0012	0.0567	-0.0004	0.0563	0.0005	0.0555	-0.0009	0.0553
$\widehat{b}_{PT}$	0.0136	0.1209	0.0018	0.1210	0.0071	0.1174	0.0068	0.1211
$\widehat{b}_{yT}$	0.0008	0.0390	0.0016	0.0383	0.0004	0.0392	0.0031	0.0410
ET-Poisson								
$\widehat{\rho}_{\eta,q}$	-0.3257	0.0636	-0.1345	0.0618	-0.4200	0.0633	-0.2185	0.0641
$\widehat{b}_{yT}$	0.7600	0.0657	0.7603	0.0681	0.7777	0.0643	0.7786	0.0647

Note. Monte Carlo simulation study with $1,000$ replications, N=$1,500$, and $1,600$ Halton draws. SD ≡ Monte Carlo standard deviation.

Table 7. Monte Carlo simulations: ASE/SD and nominal coverage

Parameters	$\rho_{\eta,v}=0.35$ $\rho_{\eta,q}=0.35$ $\rho_{v,q}=0.13$		$\rho_{\eta,v}=-0.35$ $\rho_{\eta,q}=0.35$ $\rho_{v,q}=-0.13$		$\rho_{\eta,v}=0.35$ $\rho_{\eta,q}=-0.35$ $\rho_{v,q}=-0.13$		$\rho_{Ty}=-0.35$ $\rho_{\eta,q}=-0.35$ $\rho_{v,q}=0.13$	
	ASE/SD	Ncov	ASE/SD	Ncov	ASE/SD	Ncov	ASE/SD	Ncov
EPET-Poisson								
$\widehat{\rho}_{\eta,v}$	1.041	95.4	1.015	95.1	1.033	95.8	1.010	95.0
$\widehat{\rho}_{\eta,q}$	0.992	93.6	0.996	94.4	1.019	94.2	0.990	93.9
$\widehat{\rho}_{v,q}$	1.022	92.9	1.032	92.5	1.039	94.2	1.023	94.0
$\widehat{b}_{PT}$	1.014	94.9	0.971	94.5	1.003	95.6	1.010	95.4
$\widehat{b}_{yT}$	1.012	94.8	0.994	94.3	1.001	95.0	0.968	94.0
ET-Poisson								
$\widehat{\rho}_{\eta,q}$	1.109	0.02	1.139	53.5	1.123	0	1.131	11.6
$\widehat{b}_{yT}$	1.623	0	1.607	0	1.648	0	1.634	0

Note. ASE/SD ≡ (average standard error / Montecarlo standard deviation). Ncov ≡ nominal coverage (%).

Table 8. Montecarlo simulations: Bias and SD

Parameters	$\rho_{\eta,v} = 0.78$ $\rho_{\eta,q} = 0.78$ $\rho_{v,q} = 0.61$		$\rho_{\eta,v} = -0.78$ $\rho_{\eta,q} = 0.78$ $\rho_{v,q} = -0.61$		$\rho_{\eta,v} = 0.78$ $\rho_{\eta,q} = -0.78$ $\rho_{v,q} = -0.61$		$\rho_{Ty} = -0.78$ $\rho_{\eta,q} = -0.78$ $\rho_{v,q} = 0.61$	
	bias	SD	bias	SD	bias	SD	bias	SD
EPET-Poisson								
$\widehat{\rho}_{\eta,v}$	0.0027	0.0667	0.0046	0.0635	0.0026	0.0683	0.0053	0.0666
$\widehat{\rho}_{\eta,q}$	-0.0024	0.0790	-0.0011	0.0797	0.0004	0.0743	-0.0016	0.0764
$\widehat{\rho}_{v,q}$	-0.0012	0.0699	0.0016	0.0709	0.0008	0.0663	0.0014	0.0669
$\widehat{b}_{PT}$	0.0053	0.1345	0.0006	0.1122	0.0108	0.1104	0.0073	0.1352
$\widehat{b}_{yT}$	-0.0014	0.0364	0.0001	0.0344	-0.0005	0.0364	0.0005	0.0384
ET-Poisson								
$\widehat{\rho}_{\eta,q}$	-0.4849	0.0602	-0.3051	0.0589	-0.9426	0.0647	-0.7292	0.0668
$\widehat{b}_{yT}$	0.7398	0.0663	0.7504	0.0679	0.7904	0.0593	0.7908	0.0616

Note. Monte Carlo simulation study with $1,000$ replications, N=$1,500$, and $1,600$ Halton draws. SD ≡ Monte Carlo standard deviation.

Table 9. Monte Carlo simulations: ASE/SD and nominal coverage

Parameters	$\rho_{\eta,v} = 0.78$ $\rho_{\eta,q} = 0.78$ $\rho_{v,q} = 0.61$		$\rho_{\eta,v} = -0.78$ $\rho_{\eta,q} = 0.78$ $\rho_{v,q} = -0.61$		$\rho_{\eta,v} = 0.78$ $\rho_{\eta,q} = -0.78$ $\rho_{v,q} = -0.61$		$\rho_{Ty} = -0.78$ $\rho_{\eta,q} = -0.78$ $\rho_{v,q} = 0.61$	
	ASE/SD	Ncov	ASE/SD	Ncov	ASE/SD	Ncov	ASE/SD	Ncov
EPET-Poisson								
$\widehat{\rho}_{\eta,v}$	0.983	93.8	1.004	94.1	0.970	93.0	0.989	93.3
$\widehat{\rho}_{\eta,q}$	1.036	94.9	0.999	94.4	1.025	94.7	1.021	95.8
$\widehat{\rho}_{v,q}$	2.337	94.1	7.157	92.8	3.472	92.9	2.523	93.0
$\widehat{b}_{PT}$	1.006	95.3	0.981	95.3	1.007	94.3	0.994	95.6
$\widehat{b}_{yT}$	0.997	95.5	0.988	94.8	0.984	95.3	0.965	93.7
ET-Poisson								
$\widehat{\rho}_{\eta,q}$	1.076	0	1.047	0	1.079	0	1.123	0
$\widehat{b}_{yT}$	1.546	0	1.650	0	1.818	0	1.670	0

Note. ASE/SD ≡ (average standard error / Montecarlo standard deviation). Ncov ≡ nominal coverage (%).