HEDG Working Paper 10/19

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

Massimiliano Bratti
Alfonso Miranda

July 2010

york.ac.uk/res/herc/hedgwp

Department of Quantitative Social Science

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

Massimiliano Bratti
Alfonso Miranda

DoQSS Working Paper No. 1005
May 2010

Leading education and social research
Institute of Education
University of London

DISCLAIMER

Any opinions expressed here are those of the author(s) and not those of the Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

DoQSS Workings Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Department of Quantitative Social Science. Institute of Education, University of London. 20 Bedford way, London WC1H 0AL, UK.

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

Massimiliano Bratti[*], **Alfonso Miranda**[†‡]

Abstract. In this paper we propose a method to estimate 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 many fields of economics, such as health, labor, and population economics. We show the performance of the model using data from Kenkel and Terza (2001), which investigates 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 perversely signed effect of physician advice on drinking intensity, (ii) neglecting endogenous participation leads to an upward biased estimator 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.

[*]Department of Economics, Business and Statistics, Università degli Studi di Milano, Via Conservatorio 7, I-20122, Milan, Italy. E-mail: Massimiliano.Bratti@unimi.it

[†]Department of Quantitative Social Science, Institute of Education, University of London. 20 Bedford Way, London WC1H 0AL, UK. E-mail: A.Miranda@ioe.ac.uk (corresponding author).

[‡]**Acknowledgements**. Both authors equally contributed to this paper. The usual disclaimers apply.

1 Introduction

There are often cases in 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 or cigarettes consumption (Kenkel and Terza 2001), the effect of health status on the number of visits to a general pratictioner (Windmaijer and Santos Silva 1997), the effect of health insurance coverage on the number of doctor or hospital visits (Riphahn et al. 2003) or the effect of female employment on fertility (Kalwij 2000).[1]

In all these applications the treatment of interest is likely to be endogenous.[2] Receiving physician advice of reducing smoking or drinking is certainly not exogenous with respect to the intensity of those activities. 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. Last but not least, female employment status is likely to be endogenous with respect to fertility, since women with a lower taste for children may decide both to participate into the labor market and have fewer children.

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) 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 is 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. There are other cases in which the treatment might have different effects on the intensive and the extensive margins. For instance, women who have a low taste for children may both work

[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 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).

and remain childless; and the negative correlation between employment and children could be caused by spurious association induced by the exclusion of an important control variable — taste for children — which is unobservable. However, conditional on overcoming the threshold of becoming mother, it is perfectly intuitive that working women may have more children than women who do not participate in the labor market due to an income effect; this positive association reflecting a causal effect.

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 may not be missing at random with respect to the level of drinking or smoking. In this case, neglecting sample selection will lead to inconsistent estimates of the treatment of interest (e.g., physician advice).

Although the applications above mentioned are very frequent, 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 an *interval 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 to estimate a model for a *continuous and non negative* dependent variable with endogenous participation and multivariate treatments. Multinomial normality is required. Our estimator is similar in spirit to the ones proposed by Terza et al. (2008) and Li and Trivedi (2009), and relies on the same distributional assumptions. The approach, however, is different as 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. With respect to Li and Trivedi (2009) the estimator presented here is different in the sense that we use a frequentist rather than a Bayesian approach.

We illustrate the performance of our proposed 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, distinguishing between models with endogenous treatment and sample selection vs. models with endogenous treatment and endogenous participation. 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 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 S_i when it represents a selection rule and as P_i when it represents a participation rule. Although we will refer to models using individual-level data, the individual i subscripts are omitted throughout to simplify notation.

2.1 Background

In the case of sample selection y is missing for a non negligible proportion of the sample and the selection rule is defined in such a way that $S = 1$ when y is observed and $S = 0$ when y is missing. S is always observed and, in contrast to the treatment dummy, the sample selection rule S has no genuine effect on the conditional distribution of y. We say that S is an *endogenous (sample) selection rule* if unobservable heterogeneity that affects S is also a determinant of y. Finally, if the second dummy represents a participation rule both y and P are, always, observed. In this case, however, $P = 0$ when $y = 0$ and $P = 1$ when $y > 0$. Again, from a theoretical point of view, the participation dummy P is not expected to be a genuine shifter of the conditional distribution of y and we say that P is an *endogenous participation rule* if unobserved heterogeneity affecting P also affects y. In what follows, we will refer to T as *sorting* individuals between the treated and non-treated groups and S and P as *partitioning* the sample between individuals for whom the outcome is observed/unobserved or between individuals who participate/don't participate in the main activity.

The terminology we have just introduced deserves a point of clarification. First, one should note that all three cases – endogenous treatment, endogenous selection, and endogenous participation – have in common the fact that there is a dummy variable that sorts/partitions the individuals into *two mutually excluding groups*: treated vs. non-treated, observed vs. unobserved,

active participants vs. non active participants. Second, in all three cases, 'endogeneity' denotes the fact that unobservable heterogeneity affecting the partitioning/sorting dummy is also a determinant of the main response y. Now, if the similarities are important, the differences are equally relevant. In *endogenous sample selection* there is a problem of non ignorable missing data (y is missing when $S = 0$) and the partitioning dummy has no effect on the conditional distribution of y (see, for instance, Heckman 1979, Little and Rubin 2002). In contrast, *endogenous treatment* has nothing to do with missing data but is rather a problem of an endogenous explanatory dummy variable that is a genuine shifter of the conditional distribution of y (see, for instance, Heckman 1978; 1992). Finally, in *endogenous participation* the partitioning dummy has nothing to do with missing data and it is neither an endogenous dummy explanatory variable. It is, in other words, a midway case in which the same unobservables may affect both the intensive and the extensive margins of an activity, which are, however, two separate choices.

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 most cross one or two hurdles before a strict 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 choose $y = 0$ 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). None of the aforementioned models allow endogenous participation and endogenous treatment at the same time.

Deb and Trivedi (2006) have proposed a multinomial endogenous treatment model that accommodates correlated endogenous sorting into different treatments — though neither endogenous sample selection nor endogenous participation. In particular, they consider J mutually exclusive endogenous treatments that affect the count dependent variable y, which is always observed. Correlated endogenous sorting is present in this model because at any time only one of the J treatments is active and so the treatments play also the role of a set of, correlated, partitioning dummies that split the sample into different sub-samples. The main drawback of this approach is that the treatment dummies are, by definition, equal to the sample partitioning dummies. Hence, for instance, one cannot be assigned to the sample of treated individuals and to the sub-sample of individuals non-participating in the main activity at the same time. The models discussed in this paper are different from Deb and Trivedi multinomial endogenous treatment model in two important aspects. First, the role of the sample partitioning dummy deals explicitly with a pressing feature of the data: either sample selection or endogenous participation. Second, the treatment dummy is different from the selection (participation) dummy. Hence, an individual can be assigned to treatment $T = 1$ (e.g., having received physician advice) and sub-sample $P = 0$ (e.g., not smoking).

Terza et al. (2008) and Li and Trivedi (2009) contribute the only pieces of previous work that, to the best of our knowledge, are capable of dealing with endogenous treatment and endogenous participation *at the same time*. Terza et al. (2008) propose a two-step method for analyzing the effect of prenatal care-giver advice on alcohol consumption by pregnant women, where zero vs. strictly positive consumption plays the role of the participation rule and the care-giver advice the role of the endogenous treatment. Alcohol consumption is an interval coded count. This is an extension of Mullahy (1998) Modified Two Part Model. Despite being a two-step approach, Terza et al.'s method is not a Limited Information Maximum Likelihood (LIML) estimator but relies on joint multivariate normality. Unlike Terza et al. (2008), we describe in subsection 2.3 a model for a count rather than for an interval coded count. Also, because we use Full Information Maximum Likelihood (FIML) techniques, our methods deliver an estimator that is more efficient than the two-step method suggested by Terza et al. (2008) and, unlike two-step methods, directly provides correct standard errors.

Li and Trivedi (2009) estimate a model with multinomial endogenous

treatments in the context of a two-part model for a continuous dependent variable. They apply their methods to estimate the impact of prescription drug coverage on drug expenditure of the elderly. The authors implement both a GMM and a Bayesian estimator. The GMM estimator allows for endogenous multivariate treatments but does not control for endogenous participation. The Bayesian estimator, in contrast, deals with both problems at the same time. Their findings show that the GMM estimator appears to overestimate the impact of drug coverage due to positive selection. With respect to Li and Trivedi (2009)'s Bayesian estimator, our approach to modeling both endogenous treatment effects and endogenous participation (or sample selection) is based on a different (but equally fully parametric) approach (FIML) and analyzes a count outcome, rather than a continuous dependent variable.

2.2 Endogenous selection: y missing when $S = 0$

The model with 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:

$$\begin{aligned} T^* &= \mathbf{z}'\boldsymbol{\gamma} + v, && (1) \\ S^* &= \mathbf{r}'\boldsymbol{\theta} + \varphi T + q && (2) \end{aligned}$$

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,

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

with,

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

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 log-

linear model for specifying the conditional mean of y given S, T, and η:

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

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

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

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. C1-C3 together ensure 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, 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.

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) correlated random effects 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 but allows at least some dependence.

The use of the Poisson distribution for the analysis of count data has been criticized in the past due to the unattractive feature that mean and variance are restricted to be equal, also known as *equidispersion* (see, for instance, Winkelmann 2008). In the present model, however, the introduction of the unobserved heterogeneity term η in the log-linear model for $\mu \equiv E[y|\mathbf{x}, T, \eta]$ forces the count variable y to exhibit *overdispersion*. In fact, it can be shown that:

$$\kappa \equiv E\left[y|\mathbf{x}, T\right] = E_\eta\left[E[y|\mathbf{x}, T, \eta]\right] = \exp(\mathbf{x}'\boldsymbol{\beta} + \delta T + \frac{\sigma_\eta^2}{2}) \tag{6}$$

$$\mathrm{Var}\left[y|\mathbf{x}, T\right] = \kappa\left\{1 + \kappa[\exp(\sigma_\eta^2) - 1]\right\}, \tag{7}$$

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

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:

$$\rho_{y,T} = \frac{\lambda_1 \sigma_\eta^2}{\sqrt{\sigma_\eta^2(\lambda_1^2 \sigma_\eta^2 + 1)}} \tag{8}$$

$$\rho_{y,S} = \frac{\lambda_2 \sigma_\eta^2}{\sqrt{\sigma_\eta^2(\lambda_2^2 \sigma_\eta^2 + 1)}} \tag{9}$$

$$\rho_{T,S} = \frac{\lambda_1 \lambda_2 \sigma_\eta^2}{\sqrt{(\lambda_1^2 \sigma_\eta^2 + 1)(\lambda_2^2 \sigma_\eta^2 + 1)}}. \tag{10}$$

The treatment dummy T is an exogenous variable in the main response equation whenever $\rho_{y,T} = 0$. Similarly, If $\rho_{y,S} = 0$ sample selection is exogenous in the main response equation. Notice that even if $\rho_{y,S} \approx 0$, such that $\widehat{\rho}_{y,S} = 0$, one cannot ignore the selection problem altogether if $\rho_{T,S} \neq 0$ and $\rho_{y,T} \neq 0$ because in that case S will be still dependent on y through the relationship between y and T and the fact that T is a control variable in the selection equation. As a consequence, y cannot be claimed to be missing at random (MAR) and ignoring the selection mechanism will lead the researcher to obtain inconsistent estimators of $\boldsymbol{\beta}$ and δ (for more on this topic see, for instance, Little and Rubin 2002). If $\rho_{y,S} = \rho_{y,T} = 0$, on the other hand, one can obtain consistent estimators of $\boldsymbol{\beta}$ and δ on the basis of a simple Poisson regression fitted on the sub-sample for which y is not missing, even if $\rho_{T,S} \neq 0$ (see Wooldridge 2002, p. 557).[4]

In a similar fashion, if $\rho_{y,T} = 0$ and $\delta \neq 0$ the endogenous treatment problem can only be neglected if either $\rho_{T,S} = 0$ or $\rho_{y,S} = 0$. Clearly, if both $\rho_{y,T} \neq 0$ and $\rho_{y,S} \neq 0$ then neither endogenous selection nor endogenous treatment can be ignored.

Let $\mathbb{P}_S(0|\eta)$ denote the conditional probability of $S = 0$ given η and $\mathbb{P}_S(1|\eta)$ the conditional probability of $S = 1$ 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 $F(y|\eta)$ the cumulative distribution of y given η which is defined by equations (3) and (4) together. The log-likelihood

[4]Selection is ignorable if $D(y|\mathbf{x}, \mathbf{z}, \mathbf{r}, S) = D(y|\mathbf{x}, \mathbf{z}, \mathbf{r})$. Joint normality and condition $\rho_{y,S} = \rho_{y,T} = 0$ ensures this regardless of the value that $\rho_{T,S}$ may take.

function is then:

$$\begin{aligned}\log(L) = & \sum_{i,S_i=0} \sum_{\tau} \omega_\tau \ln \left\{ \int \mathbb{P}_S(0|\eta) \mathbb{P}_T(\tau|\eta) \phi(\eta) d\eta \right\} \\ & + \sum_{i,S_i=1} \sum_{\tau} \omega_\tau \ln \left\{ \int \mathbb{P}_S(1|\eta) \mathbb{P}_T(\tau|\eta) F(y|\eta) \phi(\eta) d\eta \right\}\end{aligned} \tag{11}$$

where $\phi(\cdot)$ is the density of a normal variate with mean zero and variance σ_η^2, $\omega_0 = 1(T = 0)$, and $\omega_1 = 1(T = 1)$.

The integrals in equation (11) 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. Standard errors for marginal effects are computed using the Delta method.

The use of a common latent factor structure like that written in (5) has four main advantages over the alternative of specifying a multivariate normal distribution for v, q, and η (see also 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 write. Second, it has 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. Last but not least, and quite importantly for computational feasibility, the latent variable approach transforms a problem in which calculation of the log-likelihood involves evaluating a three dimensional integral into a problem where only a one dimensional integral needs to be evaluated.

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

The model with endogenous participation is very similar to the model that deals with sample selection but here the participation dummy P has nothing to do with missing data. Instead, endogenous participation explicitly addresses the problem that zero values and strictly positive values of the count y reflect two qualitatively different decisions. While a zero count implies

a potentially conscious decision of non participation into a given activity (say, drinking or smoking), the strictly positive count implies that a potentially conscious decision of participating was taken before the actual level of consumption was set. As we discussed earlier, these two decisions are potentially affected by different variables and could follow, therefore, two different DGPs. As before, we assume that the treatment dummy T and the participation dummy P are generated as follows:

$$\begin{aligned} T^* &= \mathbf{z}'\boldsymbol{\gamma} + v, && (12) \\ P^* &= \mathbf{r}'\boldsymbol{\theta} + \varphi T + q && (13) \end{aligned}$$

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

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

with,

$$\begin{aligned} y &= \begin{cases} 0 & \text{if } P = 0 \\ 1, 2, \ldots & \text{if } P = 1 \end{cases} && (15) \\ \ln\left(\mu\right) &= \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, && (16) \end{aligned}$$

and all other remaining aspects of the model are the same as in subsection 2.2. The main difference from the model presented here and the one in the previous subsection is the fact that here we use a zero-truncated Poisson distribution for y given $P = 1$ whereas we used a Poisson for y given $S = 1$ in the endogenous selection model. This is a minor modification that reflects the fact that in the endogenous participation model the $y = 0$ count is generated by a different data generating mechanism from $y > 0$ counts.

Another important difference is the fact that now every single individual in the sample will contribute, always, a non missing observation for y. The likelihood function is now written as follows:

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

where $G(y|\eta)$ is the cumulative distribution of y given η, $\omega_0 = 1(T = 0)$, and $\omega_1 = 1(T = 1)$. Again, the model is estimated by MSL.

3 An application to the effect of physician advice on drinking

In this section we apply the count data model with endogenous treatment effects and endogenous participation to the problem of estimating the treatment effect of physician advice on alcohol consumption using data from Kenkel and Terza (2001). However, before doing it, we introduce a very simple behavioral model to motivate our work and interpret the empirical results.

3.1 An underlying behavioral model

Let us assume that an individual's latent (continuous) index of 'bad' health status T^*, i.e. higher T^* means worse health, depends on some observable characteristics $\mathbf{x}$, some specific past health problems $\mathbf{z_1}$ (diabetes, high pressure, etc.), her intertemporal discount rate η,[5] and an unpredictable lifetime health endowment ζ, that is

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + \lambda_1\eta + \zeta \tag{18}$$

where $\mathbf{z} = (\mathbf{x}, \mathbf{z_1})$ and $\mathbf{x}$ includes an intercept.[6] Our theoretical expectation is that $\lambda_1 > 0$ because individuals with high discount rates are expected to engage in health-damaging behavior such as smoking or drinking at higher rates than individuals with low discount rates. Hence, people with high η are expected to be in worse health.

We assume that a physician can observe T^*, for instance using her knowledge and specialized medical exams, and she gives her patient the advice to reduce drinking if T^* goes above a given threshold a. We normalize the threshold to zero (this is innocuous since $\mathbf{z}$ includes an intercept term). Thus, the physician advice's rule becomes $T = 1(T^* > 0)$, where T is the observed treatment dummy for physician advice.[7] In particular,

[5] We will often refer to η_i as to the intertemporal discount rate (r_i), although it should be considered as a strictly increasing transform of the discount rate, $\eta_i = f(r_i)$, such that $f : R^+ \longmapsto R$ (e.g., the logarithm transform), where we assume that r_i can be arbitrarily close to zero but not exactly zero.

[6] The vector of variables $\mathbf{z_1}$ represents a set of exclusion restrictions, which provides an additional source of identification over and above functional form and covariance matrix restrictions.

[7] We use a rather simplified mechanism underlying the physician's advice. In reality, giving advice to reduce drinking may involve a physician's choice and depend on her characteristics. However, these characteristics are not observed in the data we will use in the next section, and we prefer to use a simplified approach.

$$\begin{aligned}\mathbb{P}(T=1) &= \mathbb{P}(T^* > 0) \\ &= \mathbb{P}(\zeta \leq \lambda_1 \eta + \mathbf{z}'\boldsymbol{\gamma}),\end{aligned} \tag{19}$$

which we label the *treatment equation* or physician advice equation (in the writing of equation (19) we suppose that the distribution of ζ is symmetric). Hence, it is clear that individuals with high discount rates are more likely to have worse health, and to receive drinking advice (as $\lambda_1 > 0$) because for them receiving a relatively small health endowment ζ will trigger the health threshold more often than for individuals with low discount rates.

The individual derives utility from drinking, and the demand for alcohol consumption, given the intertemporal discount rate, η is

$$\ln\{E[y|\mathbf{x}, T, \eta]\} = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta, \tag{20}$$

which represents the drinking intensity equation. Our theoretical prediction is that $\delta < 0$. That is, physician advice is expected to reduce drinking.

Denote the utility of being a current drinker as $U_{y>0}$, that of being a non-drinker as $U_{y=0}$, and their difference as $P_i^* = U_{y>0} - U_{y=0}$. We also define an indicator P which equals one if the individual is a current drinker, that is if drunk alcohol in the last two weeks, and zero otherwise. The difference in the two utilities is parametrized as:

$$P^* = \mathbf{x}'\boldsymbol{\theta} + \varphi T + \lambda_2 \eta + \xi. \tag{21}$$

Physician advice should reduce the probability of actively participating in the drinking activity. Hence, we expect $\varphi < 0$. At the same time, we expect that people with a high intertemporal discount rate will be less willing to refrain from participating in an activity that gives them positive utility in exchange for future better health and a longer life-span. So, we expect $\lambda_2 > 0$.[8] Then,

$$\begin{aligned}\mathbb{P}(P=1) &= \mathbb{P}(P^* > 0) \\ &= \mathbb{P}(\xi \leq \mathbf{x}'\boldsymbol{\theta} + \varphi T + \lambda_2 \eta).\end{aligned} \tag{22}$$

In other words, a high intertemporal discount rate η increases the likelihood that an individual will drink alcohol, $P = 1$. We label (22) the *participation equation* or drinking participation equation (writing (22) we suppose, again,

[8] With respect to the model in section 2.3, here we have assumed that the same covariates enter the difference in utilities and the demand for alcohol consumption.

that the distribution of ξ is symmetric).

This simple behavioral model could explain why we are likely to have a non-zero correlation between the unobservables in the drinking intensity, the endogenous treatment, and the endogenous participation equations. In particular, our theoretical predictions suggest that the unobservable intertemporal discount rate η is likely to induce a positive correlation between the error term in the physician advice equation and the error term in the drinking intensity equation, causing the treatment dummy T be an endogenous variable in (20). In other words, we expect $\rho_{y,T} > 0$. Similarly, the discount rate η is likely to induce a positive correlation between the errors in the drinking intensity and the drinking participation equation, creating an endogenous participation problem. Hence, our prediction is $\rho_{y,P} > 0$. Last but not least, in the model above, the intertemporal discount rate η also makes the physician advice dummy T endogenous in the drinking participation equation. As a consequence, $\rho_{T,P} > 0$ is expected.

The model in this section also represents an example of economic problems where one unobservable individual characteristic (η, the discount factor in our specific case) is likely to affect many processes, induce correlation between them, and raise endogeneity issues. In all these cases, using the latent factor structure outlined in section 2 may be justified not only in terms of convenience but also by economic theory.

As we will see, this very simple behavioral model is able to explain many of the features of the Kenkel and Terza's study that we aim to replicate in the next section, and will be useful to interpret our empirical results.

3.2 The Kenkel and Terza (2001) study

Building on the behavioral model sketched in the previous section here we specify the econometric model that in this specific case takes the form of a Poisson model with endogenous treatment effects and endogenous participation.

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).[9] Our aim is to show how estimates of treatment effects are sensitive to various assumptions about endogeneity of treatment status and participation. In the Kenkel and Terza's study drinking is measured as the number of drinks consumed in the last two

[9] Although the model is formally identified by 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.

weeks.[10] 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, 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.[11] In spite of this, Kenkel and Terza observe in their sample that 21% of current 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 $y = 0$ are contributed by recent quitters or people who were actively trying to stop drinking all together 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 Christmas day (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. Clearly, the fact that a good proportion of these zeros are contributed by occasional drinkers suggests that the excess zeros cannot be ignored. The authors acknowledge this and account for the excess of 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 within the standard Poisson model, 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 control variables 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 indentify the model (exclusion restrictions).

[10] 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).

[11] 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.

3.3 Results

3.3.1 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) of 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 treatment – physician advice – is considered as exogenous.[12] 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 an Endogenous Treatment (ET) Poisson model. This model assumes that both zeros and positive y outcomes are produced by the same DGP. This model (and the following models estimated using MSL) was estimated using 400 Halton draws.[13] First, note that the correlation between the errors in the drinking intensity and the physician advice equation $\rho_{y,T}$ 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 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 Endogenous Participation Endogenous Treatment (EPET) Poisson model in column 5, the effect of physician advice falls by more than one drink per week, to 4 (-25%), and

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

[13] Using more Halton draws did not produce important changes in coefficients and/or standard errors.

remains highly statistically significant.[14] Physician advice turns out to be endogenous with respect to drinking participation and $\rho_{T,P}$ is positive, which is consistent with the correlation found between advice and drinking intensity (i.e., $\rho_{y,T} > 0$) and our theoretical predictions in section 3.1. 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 that 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 to quit drinking but simply to cut their two-week drinking.

The fact that the marginal effect of T on y falls after endogenous participation is accounted for suggests that treatment effects in the sub-populations of frequent and occasional drinkers are likely to be different. This will be observed, for instance, if the individuals who continue to drink after receiving physician advice are those who have a high taste for drinking (that is, high η) and are, as a consequence, inherently less willing to cut their drinking. Therefore, the response to advice in the sub-population of 'happy drinkers' is likely to be lower than among the sub-population of occasional drinkers because the marginal rate of substitution of drinks for years of healthy life is different in these two sub-populations. Again, the finding is in line with the behavioral model outlined in section 3.1.

Two considerations are worth 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 affected by physician advice. Furthermore, correlations between unobservables in the endogenous treatment and drinking intensity, and between 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 marginal effects (about -4.5 drinks) that are somewhat between the ones reported

[14] Consistently with the behavioral model in section 3.1, 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.

by a model that only deals with the endogenous treatment problem and the ones 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.

3.3.2 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 report the marginal effects obtained from the ET Poisson model for the sake of completeness.

4 Concluding remarks

In this paper we have proposed a FIML 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 (NMAR). 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 estimator of the treatment effect of physician advice on drinking intensity.

References

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.

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

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

Heckman, J., 1979. Sample selection bias as an specification error. Econometrica 47 (1), 153–162.

Heckman, J., 1992. Randomization and social program evaluation. In: Manski, C., Garfinkel, I. (Eds.), Evaluating welfare and training programs. MA: Harvad Univesity Press, pp. 201–230.

Heckman, J. J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46 (4), 931–959.

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

Kalwij, A., 2000. The effects of female employment status on the presence and number of children. Journal of Population Economics 12 (2), 221–39.

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. K., 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.

Little, R. J. A., Rubin, D. B., 2002. Statistical anaysis with missing data.

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., 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.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

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[(a)] Poisson	EPET[(a)] Poisson	
	$y^{(b)}$	$Pr(y>0)^{(c)}$	$y^{(b)}$	$Pr(y>0)^{(c)}$	$y^{(b)}$
	(1)	(2)	(3)	(4)	(5)
Physician advice (T)	3.679***	0.079***	-5.382***	-0.040	-4.026***
	(0.558)	(0.017)	(0.330)	(0.030)	(0.473)
$\rho_{y,T}$			0.791***	0.677***	
			(0.024)	(0.050)	
$\rho_{y,P}$				0.364***	
				(0.061)	
$\rho_{T,P}$				0.208***	
				(0.044)	
σ_η			1.887***	1.422***	
			(0.051)	(0.044)	
N.obs.	2,467	2,467	2,467	2,467	
Log-likelihood	-32,263	-1,247	-10,249	-10,060	

*** significant at 1%. Eicker-Huber-White robust standard errors (RSE) in round brackets. Standard errors on marginal effects are computed using the Delta method.

[(a)] Model estimated by maximum simulated likelihood with 400 Halton draws. Adding extra Halton did not produce important changes in coefficients and/or standard errors.

[(b)] Marginal effects (ME) are computed at the sample median of the dependent variable, in analogy to Kenkel and Terza (2001).
[(c)] Marginal effects (ME) are computed at the sample mean of the independent variables.
Note. The dependent variable (y) is the number of alcoholic drinks consumed in the last two weeks. Estimation refers to the 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 ($y>0$), respectively. ET and EPET stand for Endogenous Treatment and Endogenous Participation Endogenous Treatment, respectively. Both models were estimated using MSL and 400 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 181.68 (p-value=0.00).

Table 3 Marginal effects of covariates on the physician advice equation

	ET[(a)] Poisson		EPET[(a)] Poisson	
	ME	RSE	ME	RSE
EDITINC	0.000	0.002	0.000***	0.000
AGE30	0.099***	0.033	0.094***	0.036
AGE40	0.031	0.031	0.057*	0.034
AGE50	0.054*	0.032	0.050	0.038
AGE60	0.057	0.038	0.052	0.042
AGEGT70	0.088*	0.051	0.058	0.053
EDUC	-0.012***	0.003	-0.012***	0.000
BLACK	0.102***	0.025	0.089***	0.025
OTHER	0.119*	0.058	0.077	0.059
MARRIED	0.031	0.023	0.027	0.025
WIDOW	0.07	0.046	0.063	0.046
DIVSEP	0.078**	0.032	0.081**	0.035
EMPLOYED	-0.008	0.025	-0.014	0.025
UNEMPLOY	0.109**	0.053	0.046	0.054
NORTHE	-0.007	0.023	0.024	0.024
MIDWEST	-0.02	0.021	-0.016	0.024
SOUTH	-0.027	0.021	-0.017	0.023
Excluded variables[(b)]				
MEDICARE	0.015	0.046	0.010	0.043
MEDICAID	0.007	0.033	0.017	0.035
CHAMPUS	-0.050**	0.023	-0.070***	0.023
HLTHINS	0.042*	0.025	0.064**	0.026
REGMED	0.011	0.022	-0.007	0.025
DRI	0.051	0.035	0.007	0.031
MAIORLIM	0.01	0.027	-0.000	0.027
SOMELIM	0.110***	0.036	0.085***	0.033
HVDIAB	0.065***	0.024	0.068***	0.023
HHRTCOND	0.03	0.037	0.059	0.040
HADSTROKE	-0.009	0.029	-0.021	0.030
F-test excluded variables	50.71 [0.00]		51.08 [0.00]	

*** significant at 1%; ** significant at 5%; * significant at 10%. Eicker-Huber-White robust standard errors (RSE) are reported. P-values in square brackets. Standard errors on marginal effects are computed using the Delta method.

[(a)] Model estimated by maximum simulated likelihood with 400 Halton draws. Adding extra Halton did not produce important changes in coefficients and/or standard errors.

[(b)] Variables excluded from the drinking equations for (economic) identification.

Note. The dependent variable (y) is the number of alcoholic drinks consumed in the last two weeks. The table reports the marginal effects (ME) of covariates of the physician advice equation and the F-tests for the exclusion of the identifying variables obtained with the ET and the EPET Poisson models.

Table 4. Marginal effects of other covariates

	ET[(a)] Poisson		EPET[(a)] Poisson			
	y		$Pr(y>0)$		y	
	ME	RSE	ME	RSE	ME	RSE
EDITINC	0.064***	0.024	0.001**	0.001	0.089***	0.022
AGE30	1.493***	0.410	-0.029*	0.016	1.039	1.140
AGE40	-0.509	0.325	-0.050***	0.017	0.161	0.867
AGE50	-0.016	0.340	-0.060***	0.019	-0.459	1.157
AGE60	0.359	0.364	-0.024	0.019	-1.213	1.161
AGEGT70	0.192	0.409	0.001	0.021	-2.104*	1.129
EDUC	-0.176***	0.030	0.008***	0.002	-0.416***	0.076
BLACK	-0.026	0.258	0.008	0.012	-1.753***	0.479
OTHER	-0.319	0.332	0.041	0.026	-3.267***	0.510
MARRIED	-0.567*	0.294	-0.010	0.012	-1.130**	0.488
WIDOW	0.824**	0.342	0.004	0.022	0.036	0.589
DIVSEP	0.867**	0.422	0.009	0.016	2.124	1.393
EMPLOYED	0.102	0.252	0.073***	0.013	-0.534	0.500
UNEMPLOY	6.082***	0.798	0.094***	0.018	3.593	2.232
NORTHE	-1.107***	0.286	-0.036***	0.012	-0.286	0.532
MIDWEST	-1.126***	0.203	-0.050***	0.012	-1.394**	0.566
SOUTH	-1.292***	0.189	-0.048***	0.012	-1.464***	0.483

*** significant at 1%; ** significant at 5%; * significant at 10%. Eicker-Huber-White robust standard errors (RSE) are reported. Standard errors on marginal effects (ME) are computed using the Delta method.

[(a)] Model estimated by maximum simulated likelihood with 400 Halton draws. Adding extra Halton did not produce important changes in coefficients and/or standard errors.

Note. The dependent variable (y) is the number of alcoholic drinks consumed in the last two weeks. Marginal effects (ME) are computed at the sample mean of the independent variables.