Consiglio Nazionale delle Ricerche

Working Paper Cnr-Ceris, N. 18/2012

A *CONTINUOUS TREATMENT* MODEL FOR ESTIMATING A DOSE RESPONSE FUNCTION UNDER ENDOGENEITY AND HETEROGENEOUS RESPONSE TO OBSERVABLE CONFOUNDERS: DESCRIPTION AND IMPLEMENTATION VIA THE STATA MODULE "CTREATREG"

Cerulli Giovanni

Working Paper

Consiglio Nazionale delle Ricerche

CERIS ISTITUTO DI RICERCA SULL'IMPRESA E LO SVILUPPO

WORKING PAPER CNR - CERIS

RIVISTA SOGGETTA A REFERAGGIO INTERNO ED ESTERNO

ANNO 14, N° 18 – 2012
Autorizzazione del Tribunale di Torino
N. 2681 del 28 marzo 1977

A *continuous treatment* model for estimating a Dose Response Function under endogeneity and heterogeneous response to observable confounders: Description and implementation via the Stata module "`ctreatreg`"

Giovanni Cerulli

National Research Council of Italy
Ceris-CNR
Institute for Economic Research on Firms and Growth
Via dei Taurini, 19 - 00185 Roma
Phone: +39.06.4993.7885
E-mail: g.cerulli@ceris.cnr.it

ABSTRACT: This paper presents an original econometric model for estimating a Dose Response Function though a regression approach when treatment is *continuous*, individuals may react heterogeneously to observable confounders and selection-into-treatment may be (potentially) endogenous. After the description of the model, two estimation procedures are set out: one based on OLS under conditional mean independence, and one based on IV under selection endogeneity. The paper goes on by presenting `ctreatreg`, a author's user-written Stata routine for an easy implementation of such a model. The paper proceeds by performing a Monte Carlo experiment to test the reliability of the model and of its associated Stata routine. Results show that the model and the Stata routine `ctreatreg` are both reliable as estimates consistently fit the expected results.

CONTENTS

1. INTRODUCTION

In many socio-economic contexts, policy interventions take the form of "continuous" exposure to a certain type of "treatment". In public policies to support business R&D, for instance, companies are not only selected for treatment, but also awarded a different "amount" of support. Likewise, individuals getting a grant to set-up a new business, or to escape some poverty threshold are other examples in which the amount of support can vary by individual, thereby providing ground for a different response to policy.

In short, from a program evaluation perspective, what is relevant in many settings is not only the *binary* treatment status, but also the level of *exposure* (or "dose") taken. This is in tune with the language used in epidemiology, where Dose Response Functions are usually estimated in order to checking patients' resilience to different levels of drug administration.

This paper presents an original econometric model for estimating a Dose Response Function through a regression approach when treatment is *continuous*, individuals may react heterogeneously to observable confounders and selection-into-treatment may be (potentially) endogenous.

To fix ideas, consider a policy program where the treatment is assigned not randomly (i.e., according to some "structural" rule), and where – after setting who is treated and who is not – the program provides a different "level" of treatment (dose) ranging from 0 (absence of treatment) to 100 (maximum treatment level). Two groups of units are thus formed: (i) *untreated*, whose level of treatment (or "dose") is zero, and (ii) *treated*, whose level of treatment is greater than zero.

We are interested in estimating the causal effect of the treatment variable *t* on an outcome *y* within the sample observed, by assuming that treated and untreated units may respond differently both to specific observable confounders ($\mathbf{x}$), and to the "level" of the treatment. We wish to estimate a Dose-Response-Function of *y* on *t*.

The paper presents also a STATA routine, "`ctreatreg`", for the practical estimation of the Dose-Response-Function (DRF) for such a model. In this context, the DRF is shown to be equal to the "Average Treatment Effect, given the level of treatment *t*" (i.e. ATE(*t*)), along with other "causal" parameters of interest, such as the average treatment effect (ATE), the average treatment effect on treated (ATET), the average treatment effect on non-treated (ATENT), the average treatment effect conditional on the vector ($\mathbf{x}$; *t*), i.e. ATE($\mathbf{x}$; *t*), etc..

Compared with similar models - and in particular the one proposed by Hirano and Imbens (2004) implemented in STATA by Bia and Mattei (2008)[1] - this model does not need a full normality assumption, and it is well-suited when many individuals have a zero-level of treatment. Furthermore, it may account for treatment "endogeneity", by exploiting an Instrumental-Variables (IV) estimation. The Dose-Response-Function is approximated by a third degree polynomial**.** Both OLS and IV estimation are

---

[1] See also Bia, Flores and Mattei (2011) generalizing the Hirano-Imbens (2004) model by allowing for a nonparametric estimation of the Dose-Response Function.

considered. In particular, IV is based on a Heckman bivariate selection model for *w* (the yes/no decision to treat a given unit) and *t* (the level of the treatment provided) in the first step, and a 2SLS estimation for the outcome (*y*) equation in the second step.

The reliability of the model and of `ctreatreg` is checked by a Monte Carlo experiment proving that the model and the routine both work correctly. The routine provides also an interesting graphical representation of results.

The paper is organized as follows: section 2 and subsections present the model and the related estimation techniques; section 3 presents the STATA command `ctreatreg`, its use, and the Monte Carlo experiment; section 4 concludes the paper.

2. THE MODEL

Consider two different and exclusive outcomes, one when a unit is treated, y_1, and one when the same unit is untreated, y_0. Define *w* as the treatment indicator, taking value 1 for treated and 0 for untreated units, and the functions $g_1(\mathbf{x})$ and $g_0(\mathbf{x})$ as the unit response to the vector of confounding variables $\mathbf{x}$ when the unit is treated and untreated respectively. Assume μ_1 and μ_0 to be two scalars and e_1 and e_0 two random variables assumed to have zero mean and constant variance. Finally define with $h(t)$ the response function to the level of treatment *t*. Given this, the model takes on this form[2]:

$$\begin{cases} w=1 \implies y_1 = \mu_1 + g_1(\mathbf{x}) + h(t) + e_1 \\ w=0 \implies y_0 = \mu_0 + g_0(\mathbf{x}) + e_0 \end{cases}$$

where:

$$\begin{cases} h(t)=0 & if \quad w=0 \\ h(t)\neq 0 & if \quad w=1 \end{cases}$$

Assume a parametric form for $g_0(\mathbf{x})=\mathbf{x}\boldsymbol{\delta_0}$ and $g_1(\mathbf{x})=\mathbf{x}\boldsymbol{\delta_1}$ and define the Average Treatment Effect (ATE) conditional on $\mathbf{x}$ as:

$$\text{ATE}(\mathbf{x};t) = \text{E}(y_1 - y_0 \mid \mathbf{x},t) = \begin{cases} (\mu_1 - \mu_0) + \mathbf{x}(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_0) + h(t) & if \quad t>0 \\ (\mu_1 - \mu_0) + \mathbf{x}(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_0) & if \quad t=0 \end{cases}$$

$$= \begin{cases} \mu + \mathbf{x}\boldsymbol{\delta} + h(t) & if \quad t>0 \\ \mu + \mathbf{x}\boldsymbol{\delta} & if \quad t=0 \end{cases}$$

---

[2] Such a model is the representation of a *treatment random coefficient regression* as showed by Wooldridge (1997; 2003). See also Wooldridge (2010, Ch. 18).

thereby getting:

$$\text{ATE}(\mathbf{x},t,w) = \begin{cases} \text{ATE}(\mathbf{x},t>0) & if \quad w=1 \\ \text{ATE}(\mathbf{x},t=1) & if \quad w=0 \end{cases} = \text{I}(t>0)[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]+\text{I}(t=0)[\mu+\mathbf{x}\boldsymbol{\delta}] =$$
$$= w[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]+(1-w)[\mu+\mathbf{x}\boldsymbol{\delta}]$$

By averaging on $(\mathbf{x},t,w)$, the previous formula becomes:

$$\text{ATE}=\text{E}_{(\mathbf{x},t,w)}\{\text{ATE}(\mathbf{x},t,w)\} = \text{E}_{(\mathbf{x},t,w)}\{w[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]+(1-w)[\mu+\mathbf{x}\boldsymbol{\delta}]\} =$$
$$= \text{E}_{\mathbf{x}}\{\text{E}_t\{\text{E}_w\{w[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]+(1-w)[\mu+\mathbf{x}\boldsymbol{\delta}]\,|\,\mathbf{x},t\}\,|\,\mathbf{x}\}\} =$$
$$= \text{E}_{\mathbf{x}}\{\text{E}_t\{p(w=1)[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]+p(w=0)[\mu+\mathbf{x}\boldsymbol{\delta}]\,|\,\mathbf{x},t\}\}\} =$$
$$= p(w=1)\text{E}_{\mathbf{x}}\{\text{E}_{t>0}[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)]\,|\,\mathbf{x}\}+p(w=0)\text{E}_{\mathbf{x}}\{\text{E}_{t=0}(\mu+\mathbf{x}\boldsymbol{\delta})\,|\,\mathbf{x}\} =$$
$$= p(w=1)\text{E}_{\mathbf{x},t>0}[\mu+\mathbf{x}\boldsymbol{\delta}+\overline{h}_{t>0}]+p(w=0)\text{E}_{\mathbf{x},t=0}[\mu+\mathbf{x}\boldsymbol{\delta}] =$$
$$= p(w=1)(\mu+\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})+p(w=0)(\mu+\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta}) =$$
$$= N_T/N \cdot (\mu+\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})+N_{NT}/N \cdot (\mu+\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta})$$

Since by definition ATE = $p(w\text{=}1) \cdot$ ATET + $p(w\text{=}0) \cdot$ ATENT, we can get from the last row of the previous formula that:

$$\begin{cases} \text{ATE} = p(w=1)(\mu+\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})+p(w=0)(\mu+\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta}) \\ \text{ATET} = \mu+\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0} \\ \text{ATENT} = \mu+\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta} \end{cases} \tag{1}$$

and by simple algebra (by adding and subtracting the same expressions), we obtain:

$$\text{ATE}(\mathbf{x},t,w) = w[\mu+\mathbf{x}\boldsymbol{\delta}+h(t)+(\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})-(\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})]+(1-w)[\mu+\mathbf{x}\boldsymbol{\delta}+(\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta}-\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta})] =$$
$$= w[(\mu+\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta}+\overline{h}_{t>0})+(\mathbf{x}_{t>0}-\overline{\mathbf{x}}_{t>0})\boldsymbol{\delta}+(h(t)-\overline{h}_{t>0})]+(1-w)[(\mu+\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta})+(\mathbf{x}_{t=0}-\overline{\mathbf{x}}_{t=0})\boldsymbol{\delta}]$$

that is:

$$\text{ATE}(\mathbf{x},t,w) = w \cdot [\text{ATET}+(\mathbf{x}_{t>0}-\overline{\mathbf{x}}_{t>0})\boldsymbol{\delta}+(h(t)-\overline{h}_{t>0})]+(1-w) \cdot [\text{ATENT}+(\mathbf{x}_{t=0}-\overline{\mathbf{x}}_{t=0})\boldsymbol{\delta}]$$

so that:

$$\begin{cases} \text{ATET}(\mathbf{x},t) = \text{ATE}(\mathbf{x},t,w=1) = \text{ATET}+(\mathbf{x}_{t>0}-\overline{\mathbf{x}}_{t>0})\boldsymbol{\delta}+(h(t)-\overline{h}_{t>0}) \\ \text{ATE}(\mathbf{x},t) = \text{ATE}(\mathbf{x},t,w=0) = \text{ATENT}+(\mathbf{x}_{t=0}-\overline{\mathbf{x}}_{t=0})\boldsymbol{\delta} \end{cases} \tag{2}$$

where:

$$\begin{cases} \text{ATET} = \mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0} \\ \text{ATENT} = \mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta} \end{cases} \tag{3}$$

Finally, we can define the Dose-Response-Function (DRF) simply by averaging ATE($\mathbf{x}$, t) on $\mathbf{x}$:

$$\text{ATE}(t,w) = \text{E}_{\mathbf{x}}\{\text{ATE}(\mathbf{x},t,w)\} = w\cdot[\text{ATET} + (h(t)-\overline{h}_{t>0})] + (1-w)\cdot\text{ATENT}$$

that is:

$$\text{ATE}(t) = \begin{cases} \text{ATET} + (h(t)-\overline{h}_{t>0}) & \text{if} \quad t>0 \\ \text{ATENT} & \text{if} \quad t=0 \end{cases} \tag{4}$$

The estimation of (4) is main purpose of this paper.

2.1 The regression approach

In this section we consider the conditions for a consistent estimation of previous causal parameters. We start from the Potential Outcome Model (POM):

$$\begin{cases} y_0 = \mu_0 + g_0(\mathbf{x}) + e_0 \\ y_1 = \mu_1 + g_1(\mathbf{x}) + h(t) + e_1 \end{cases}$$

The observable outcome is $y = y_0 + w(y_1 - y_0)$ that we can write, by substitution, as:

$$y = y_0 + w(y_1 - y_0) = (\mu_0 + g_0(\mathbf{x}) + e_0) - w[(\mu_1 + g_1(\mathbf{x}) + h(t) + e_1) - (\mu_0 + g_0(\mathbf{x}) + e_0)]$$

By collecting the various arguments, we get that:

$$y = \mu_0 + w\cdot(\mu_1 - \mu_0) + g_0(\mathbf{x}) + w\cdot[g_1(\mathbf{x}) - g_0(\mathbf{x})] + w\cdot h(t) + e_0 + w\cdot(e_1 - e_0)$$

By assuming $g_1(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_1$ and $g_0(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_0$, and by adding and subtracting $w\overline{\mathbf{x}}\boldsymbol{\delta}$ and $w\overline{h}$, we have that:

$$y = \mu_0 + w\cdot(\mu_1 - \mu_0) + \mathbf{x}\boldsymbol{\delta_0} + w\cdot[\mathbf{x}\boldsymbol{\delta}_1 - \mathbf{x}\boldsymbol{\delta_0}] + w\cdot h(t) + e_0 + w\cdot(e_1 - e_0) + (w\overline{\mathbf{x}}\boldsymbol{\delta} - w\overline{\mathbf{x}}\boldsymbol{\delta}) + (w\overline{h} - w\overline{h})$$

that is:

$$y = \mu_0 + w \cdot [(\mu_1 - \mu_0) + \overline{\mathbf{x}}\boldsymbol{\delta} + \overline{h}] + \mathbf{x}\boldsymbol{\delta_0} + w \cdot (\mathbf{x} - \overline{\mathbf{x}})\boldsymbol{\delta} + w \cdot (h(t) - \overline{h}) + e_0 + w \cdot (e_1 - e_0) \qquad (5)$$

By assuming Conditional Mean Independence (CMI), namely that − given $\mathbf{x}$ both w and t are endogenous in equation (5), we can write the regression line of y as:

$$\mathrm{E}(y \mid \mathbf{x}, w, t) = \mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\underbrace{[(\mu_1 - \mu_0) + \overline{\mathbf{x}}\boldsymbol{\delta} + \overline{h}]}_{\text{ATE}} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + w[h(t) - \overline{h}] \qquad (6)$$

since CMI implies that $\mathrm{E}[e_0 + w \cdot (e_1 - e_0) \mid \mathbf{x}, w, t] = \mathrm{E}[e_0 + w \cdot (e_1 - e_0) \mid \mathbf{x}] = 0$, being $\mathbf{x}$ exogenous by definition. In equation (6) we have to show that ATE $= (\mu_1 - \mu_0) + \overline{\mathbf{x}}\boldsymbol{\delta} + \overline{h}$. The following proof shows this.

Proof.

We know that: $\overline{\mathbf{x}} = p(w=1) \cdot \overline{\mathbf{x}}_{t>0} + p(w=0) \cdot \overline{\mathbf{x}}_{t=0}$ where $\overline{\mathbf{x}}$ refers to the average on the entire sample. Now call $p(w=1)=p_1$ and $p(w=0)=p_0$. We saw above that:

$$\text{ATE} = p_1(\mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0}) + p_0(\mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta}) = p_1\mu + p_1\overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + p_1\overline{h}_{t>0} + p_0\mu + p_0\overline{\mathbf{x}}_{t=0}\boldsymbol{\delta} =$$
$$(p_1 + p_0)\mu + (p_1\overline{\mathbf{x}}_{t>0} + p_1\overline{\mathbf{x}}_{t=0})\boldsymbol{\delta} + p_1\overline{h}_{t>0} = \mu + \overline{\mathbf{x}}\boldsymbol{\delta} + \overline{h}$$

since $p_1 + p_0 = 1$. □

The previous proof leads to the estimation of this regression equation:

$$\mathrm{E}(y \mid \mathbf{x}, w, t) = \mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\text{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + w[h(t) - \overline{h}] \qquad (7)$$

where the term $[h(t) - \overline{h}]$ can be estimated by linear regression, partial linear or polynomial regression.

2.2 Estimation of the Dose-Response-Function under CMI

By supposing a three degree polynomial form for the function $h(t)$ of this form:

$$h(t) = at + bt^2 + ct^3$$

We get that equation (7) becomes:

$$\mathrm{E}(y \mid \mathbf{x}, w, t) = \mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\text{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + w[(at + bt^2 + ct^3) - (a\mathrm{E}(t) + b\mathrm{E}(t^2) + c\mathrm{E}(t^3))]$$

that is:

$$E(y \mid \mathbf{x}, w, t) = \mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\text{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + a[t - E(t)]w + b[t^2 - E(t^2)]w + c[t^3 - E(t^3)]w \tag{8}$$

Under CMI, an OLS estimation of equation [8] leads to the following consistent estimates of the parameters $\hat{\mu}_0, \hat{\boldsymbol{\delta}}_0, \hat{ATE}, \hat{\boldsymbol{\delta}}, \hat{a}, \hat{b}, \hat{c}$. With these parameters at hand, we can finally estimate consistently the Dose-Response-Function, taking on this form:

$$\hat{\text{ATE}}(t_i) = w[\hat{\text{ATET}} + \hat{a}(t_i - \frac{1}{N}\sum_{i=1}^{N} t_i) + \hat{b}(t_i^2 - \frac{1}{N}\sum_{i=1}^{N} t_i^2) + \hat{c}(t_i^3 - \frac{1}{N}\sum_{i=1}^{N} t_i^3)] + (1 - w)\hat{\text{ATENT}} \tag{9}$$

with:

$$\hat{\text{ATET}}(t_i) = \hat{\text{ATE}}(t_i)_{t_i>0}$$

A simple graph of the curve $\hat{\text{ATE}}(t_i)_{t_i>0}$ as function of *t*, returns the form of the DRF.

It is interesting to calculate, for each level of the dose *t* the 95% confidence interval around the DRF. By defining $T_1 = t - E(t)$, $T_2 = t^2 - E(t^2)$ and $T_3 = t^3 - E(t^3)$, the standard error of the DFR is equal to[3]:

$$\hat{\sigma}_{\hat{\text{ATE}}(t)} = \left\{ T_1\hat{\sigma}_a^2 + T_2\hat{\sigma}_b^2 + T_3\hat{\sigma}_c^2 + 2T_1T_2\hat{\sigma}_{a,b} + 2T_1T_3\hat{\sigma}_{a,c} + 2T_2T_3\hat{\sigma}_{b,c} \right\}^{1/2}$$

so that the 95% confidence interval of $\hat{\text{ATE}}(t)$ for each *t* is given by:

$$\left\{ \hat{\text{ATE}}(t) \pm 1.96 \cdot \hat{\sigma}_{\hat{\text{ATE}}(t)} \right\}$$

that can be usefully graphed along the DRF curve.

2.3 Estimation of the Dose-Response-Function under treatment endogeneity

When *w* (and thus *t*) are endogenous (i.e., CMI hypothesis does not hold anymore), then the estimation of regression (8) by OLS is known to be biased. Nevertheless, an Instrumental-Variables (IV) estimation procedure may be implemented to restore consistency.

---

[3] This comes from the variance/covariance properties where T_1 T_2 T_3 are taken as constant and *a*, *b* and *c* as random variables.

To this aim, we need to express the model in this extensive form:

$$y = \mu_0 + \mathbf{x}\boldsymbol{\delta}_0 + w\text{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + a[t - \text{E}(t)]w + b[t^2 - \text{E}(t^2)]w + c[t^3 - \text{E}(t^3)]w + \varepsilon$$

$$w = \begin{cases} 1 & \text{if} \quad w^* > 0 \\ 0 & \text{if} \quad w^* \leq 0 \end{cases}$$

$$t = \begin{cases} t' & \text{if} \quad w^* > 0 \\ t^* & \text{if} \quad w^* \leq 0 \end{cases}$$

where w^* represent the latent unobservable counterpart of the binary variable w. For instance, w^* might be seen as the cost-benefit calculus of an agency choosing to finance specific subjects. As for t, it is fully observed only when $w=1$ (and $t=t'$) otherwise it is supposed to be unobserved (although put equal to zero). By defining $T_1=t$-E(t), $T_2=t^2$-E(t^2) and $T_3= t^3$-E(t^3), the previous model may be re-written as follows:

$$\begin{cases} y = \mu_0 + \mathbf{x}\boldsymbol{\delta}_0 + w\text{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + awT_1 + bwT_1 + wT_3 + \varepsilon_y & (10.1) \\ w^* = \mathbf{x}_1\boldsymbol{\beta}_1 + \varepsilon_w & (10.2) \\ t' = \mathbf{x}_2\boldsymbol{\beta}_2 + \varepsilon_t & (10.3) \end{cases}$$

where ε_w, ε_t and ε_y are error terms supposed to be freely correlated with zero mean. Equation (10.2) − the *selection* equation − defines the regression explaining the cost-benefit indicator w^*. The vector of covariates $\mathbf{x}_1$ are the selection criteria used, for instance, by an agency to set the treated and untreated group. In turn equation (10.3) − the *treatment-level* equation − defines how the level of unit treatment is decided, and it regards only units that were considered eligible for treatment. The vector of covariates $\mathbf{x}_2$ are those exogenous variables thought of as determining exactly the treatment level.

In equation (10.1), both w and T_1, T_2 and T_3 are endogenous. To estimate consistently the parameters of that system of equations we may proceed in two steps:

1. *First*: we estimate the last two equations (10.2)-(10.3) jointly by a "bivariate sample-selection model" that can be consistently estimated by a Heckman two-step procedure (Heckman, 1979). As said, this can be seen as a model in which first the decision to treat a given unit has been taken, and then the level of the treatment decided. The Heckman two-step procedure performs a probit of w on $\mathbf{x}_1$ in the first step and a OLS regression of t' on $\mathbf{x}_2$ augmented with the Mills' ratio obtained from the probit in the second step.

2. *Second*: we take the all sample predicted values of w (i.e. $\hat{p}_w$) and t (i.e. $\hat{t}$) from the previous Heckman estimation, and then we perform a 2SLS for equation (10.1) using as instruments the following exogenous variables $(\mathbf{x}, \hat{p}_w, \hat{p}_w[\mathbf{x} - \overline{\mathbf{x}}], \hat{p}_w\hat{T}_1, \hat{p}_w\hat{T}_2, \hat{p}_w\hat{T}_3)$, thus getting a consistent estimation of the

coefficients ($\mu_0, \boldsymbol{\delta_0}, \text{ATE}, \boldsymbol{\delta}, a, b, c$). Observe that the instruments used are based on the orthogonal projection of w and t on the vector space generated by the all exogenous variables of the model.

The problem of this procedure is with parameters' identification. To get precise estimation, we need at least one instrumental variable (z) appearing only in equation (2), that is, only able to explain directly the selection process. Thus, we run under the following identification assumption:

$\mathbf{x_1} = [\mathbf{x}; z]$
$\mathbf{x_2} = [\mathbf{x}]$

so that a full specified model (all the equations depend on the same exogenous $\mathbf{x}$) is considered, where z is the instrumental variable directly correlated with the selection, but directly uncorrelated with the level of the dose as well as the level of the outcome. This procedure indentifies correctly the parameters of interest.

2.4. Estimation of comparative dose-response-functions

Besides the Dose-Response-Function and the other causal parameters of interest as defined above, the previous model allows also for calculating the average comparative response at different level of treatment (as in Hirano and Imbens, 2004). This quantity takes this formula:

$$\text{ATE}(t, \Delta) = \text{E}[y(t + \Delta) - y(t)] \tag{11}$$

Equation (11) identifies the average treatment effect between two states (or levels of treatment): t and $t + \Delta$. Given a certain level of $\Delta = \bar{\Delta}$, we can get a particular $\text{ATE}(t, \bar{\Delta})$ that can be called as the "treatment function at $\bar{\Delta}$". Observe that the standard $\text{ATE}(t)$ is obtained from (11) when t=0.

How can we get an estimation of $ATE(t, \Delta)$ in our setting? We can observe that in our framework the "potential outcome" at different t, i.e. $\text{E}[y(t)]$, is:

$$\text{E}(y \mid t) = \text{E}_{\mathbf{x}, w=1}\{\text{E}(y \mid \mathbf{x}, w, t)\} = \text{E}_{\mathbf{x}, w=1}\{\mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\text{ATE} + w[\mathbf{x} - \bar{\mathbf{x}}]\boldsymbol{\delta} + w[h(t) - \bar{h}]\} = $$
$$\mu_0 + \bar{\mathbf{x}}\boldsymbol{\delta_0} + \text{ATE} + (h(t) - \bar{h})$$

Therefore:

$$\text{E}(y \mid t + \Delta) - \text{E}(y \mid t) = [\mu_0 + \bar{\mathbf{x}}\boldsymbol{\delta_0} + \text{ATE} + (h(t + \Delta) - \bar{h})] - [\mu_0 + \bar{\mathbf{x}}\boldsymbol{\delta_0} + \text{ATE} + (h(t) - \bar{h})] = h(t + \Delta) - h(t)$$

that is:

$$\text{ATE}(t,\Delta) = \text{E}[\,y(t+\Delta) - y(t)] = h(t+\Delta) - h(t) = [\,a(t+\Delta) + b(t+\Delta)^2 + c(t+\Delta)^3\,] - [\,at + bt^2 + ct^3\,]$$

and an estimation is thus given by:

$$\hat{\text{ATE}}(t,\Delta) = \hat{a}(t+\Delta) + \hat{b}(t+\Delta)^2 + \hat{c}(t+\Delta)^3 - [\,\hat{a}t + \hat{b}t^2 + \hat{c}t^3\,]$$

Given a predefined $\Delta = \bar{\Delta}$, for each level of *t* we can use a *bootstrap* of $\hat{\text{ATE}}(t,\Delta)$ over $(\hat{a},\hat{b},\hat{c})$ to get the standard errors of $\hat{\text{ATE}}(t,\Delta)$ and then its statistical significance at various level of *t*.

3. THE STATA ROUTINE `CTREATREG`

A software implementation for estimating the model presented in section 2 has been realized by the user-written STATA routine `ctreatreg`. This routine run both under CMI and under treatment endogeneity[4]. The Help-file of the routine shows the syntax along with the options as set out in Table 1.

Table 1. STATA help file for `ctreatreg`.

```
help ctreatreg
--------------------------------------------------------------------------------------------------------------

Title

    ctreatreg -  Dose-Response model with "continuous" treatment, endogeneity and heterogeneous response to
                 observable confounders

Syntax

        ctreatreg outcome treatment [varlist] [if] [in] [weight], model(modeltype) ct(treat_level)
            [hetero(varlist_h) iv(instrument) delta(number) graphic conf(number) vce(robust) const(noconstant)
            head(noheader) beta]
    fweights, iweights, and pweights are allowed; see weight.

Description

    ctreatreg estimates the Dose-Response-Function (DRF) of a given treatment on a specific target variable, within
    a model where units are treated with different levels. The DRF is defined as the "average treatment effect,
    given the level of the treatment t" (i.e. ATE(t)).  The routine also estimates other "causal" parameters of
    interest, such as the average treatment effect (ATE), the average treatment effect on treated (ATET), the
    average treatment effect on non-treated (ATENT), and the same effects conditional on t and on the vector of
    covariates x.The DRF is approximated by a third degree polynomial function.  Both OLS and IV estimation are
    available, according to the case in which the treatment is not or is endogenous. In particular, the implemented
    IV estimation is based on a Heckman bivariate selection model for w (the yes/no decision to treat a given unit)
    and t (the level of the treatment provided) in the first step, and a 2SLS estimation for the outcome (y)
    equation in the second step.  The routine allows also for a graphical representation of results.

Options

    model(modeltype) specifies the treatment model to be estimated, where modeltype must be one of the following two
        models: "ct-ols", "ct-iv".  it is always required to specify one model.

    ct(treat_level) specifies the treatment level (or dose).  This variable takes values in the [0;100] interval,
        where 0 is the treatment level of non-treated units. The maximun dose is thus 100.

    hetero(varlist_h) specifies the variables over which to calculate the idiosyncratic Average Treatment Effect
        ATE(x), ATET(x) and ATENT(x), where x=varlist_h. It is optional for all models. When this option is not
        specified, the command estimates the specified model without heterogeneous average effect. Observe that
        varlist_h should be the same set or a subset of the variables specified in varlist.  Observe however that
        only numerical variables may be considered.
```

---

[4] For a STATA implementation when the treatment is *binary* see Cerulli (2012).

iv(instrument) specifies the variable to be used as instrument in the Heckman bivariate selection model. This option is required only for "ct-iv".

delta(number) identifies the average treatment effect between two states: t and t+delta. For any reliable delta, we can obtain the response function ATE(t;delta)=E[y(t)-y(t+delta)].

graphate allows for a graphical representation of the density distributions of ATE(x;t) ATET(x;t) and ATENT(x;t). It is optional for all models and gives an outcome only if variables into hetero() are specified.

graphdrf allows for a graphical representation of the Dose Response Function (DRF) and of its derivative. It plots also the 95% confidence interval of the DRF over the dose levels.

vce(robust) allows for robust regression standard errors. It is optional for all models.

beta reports standardized beta coefficients. It is optional for all models.

const(noconstant) suppresses regression constant term. It is optional for all models.

conf(number) sets the confidence level equal to the specified number. The default is number=95.

modeltype_options	description
Model	
ct-ols	Control-function regression estimated by ordinary least squares
ct-iv	IV regression estimated by Heckman bivariate selection model and 2SLS

`ctreatreg` creates a number of variables:

_ws_varname_h are the additional regressors used in model's regression when hetero(varlist_h) is specified.

_ps_varname_h are the additional instruments used in model's regression when hetero(varlist_h) is specified in model "ct-iv".

ATE(x;t) is an estimate of the idiosyncratic Average Treatment Effect.

ATET(x;t) is an estimate of the idiosyncratic Average Treatment Effect on treated.

ATENT(x;t) is an estimate of the idiosyncratic Average Treatment Effect on Non-Treated.

ATE(t) is an estimate of the Dose-Response-Function.

ATET(t) is the value of the Dose-Response-Function in t>0.

ATENT(t) it is the value of the Dose-Response-Function in t=0.

probw is the predicted probability from the Heckman selection model (estimated only for model "ct-iv").

mills is the predicted Mills' ratio from the Heckman selection model (estimated only for model "ct-iv").

t is a copy of the treatment level variable, but only in the sample considered.
t_hat is the prediction of the level of treatment from the Heckman bivariate selection model (estimated only for model "ct-iv").

der_ATE_t is the estimate of the derivative of the Dose-Response-Function.

std_ATE_t is the standardized value of the Dose-Response-Function.

std_der_ATE_t is the standardized value of the derivative of the Dose-Response-Function.

Tw, T2w, T3w are the three polynomial factors of the Dose-Response-Function.

T_hatp, T2_hatp, T3_hatp are the three instruments for the polynomial factors of the Dose-Response-Function when model "ct-iv" is used.

`ctreatreg` returns the following scalars:

r(N_tot) is the total number of (used) observations.

r(N_treated) is the number of (used) treated units.

r(N_untreated) is the number of (used) untreated units.

r(ate) is the value of the Average Treatment Effect.

r(atet) is the value of the Average Treatment Effect on Treated.

r(atent) is the value of the Average Treatment Effect on Non-treated.

Remarks

The variable specified in treatment has to be a 0/1 binary variable (1 = treated, 0 = untreated).

The standard errors for ATET and ATENT may be obtained via bootstrapping.

When using the option ct-iv in **modeltype()**, be sure that the number of variables included in **hetero()** is less than the number of variables included in *varlist*. This is because otherwise instruments are too much correlated and some emerging collinearity prevent to identify the estimates. For instance, when six covariates are specified in *varlist*, at most five are to be put into **hetero().**

Following the Help-file this routine is rather straightforward to use and provides suitable graphical representation of results. In particular it provides a graph for the DRF and a combined graph for the densities of ATE($\mathbf{x}$,t), ATET($\mathbf{x}$,t) and ATENT($\mathbf{x}$,t).

3.1 A Monte Carlo experiment for testing ctreatreg's reliability

In this section we provide a Monte Carlo experiment to check whether `ctreatreg` complies with predictions from the theory and to assess its correctness from a computational point of view. The first step is that of defining a Data Generating Process (DGP) as follows:

$$\begin{cases} w = 1[50 + 60x_1 + 30x_2 + 60z + a > 0] \\ y_0 = 0.1 + 0.2x_1 + 0.3x_2 + e \\ y_1 = 0.3 + 0.6x_1 + 0.3x_2 + e \\ t = 0.4x_1 + 0.6x_2 + u \end{cases}$$

where we have assumed, for simplifying the model, that $e_1 = e_0 = e$ and:

$$\begin{cases} x_1 \sim U(0;1) \cdot 100 \\ x_2 \sim U(0;1) \cdot 100 \\ z = \ln(h) \\ h \sim 2 \cdot \chi^2(1) \end{cases}$$

with:

$$(a,u) \sim N(\mathbf{0};\mathbf{\Omega})$$

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_a^2 & \sigma_{a,u} \\ & \sigma_u^2 \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \rho_{a,u}\sigma_a\sigma_u \\ & \sigma_u^2 \end{pmatrix}$$

$$\sigma_a^2 = 1, \quad \sigma_u^2 = 6.5, \quad \rho_{a,u} = 0.8$$

Finally, we suppose that the correlation between a and e_0 can be either equal or different from zero. In the latter case, w is endogenous.

Therefore, we assume the following DGP[5]:

$$e = \eta + \gamma a + v$$
$$v \sim N(0;1)$$
$$\gamma = \sqrt{\rho^2/(1-\rho^2)}$$
$$\rho = \mathrm{corr}(e;a)$$
$$\eta = 0.0001$$

When ρ=0 the model CT-OLS would be the appropriate one; otherwise, the model CT-IV should be employed. With z we indicate the instrument, directly correlated with w, but (directly) uncorrelated with y_1 and y_0. Given these assumptions, the DGP is completed by the potential outcome equation:

$$y_i = y_{0i} + w_i (y_{1i} - y_{0i})$$

generating the observable outcome y.

The DGP is simulated 200 times using a sample size of 10,000. For each simulation we get a different data matrix (x_1, x_2, y, w, z) on which we apply the two models (CT-OLS and CT-IV) implemented by `ctreatreg`.

- *Case 1. Exogeneity*

We start by assuming ρ=0, that is, zero correlation between the error term of the outcome equation (e) and the error term of the selection equation (a). Under this assumption, w is exogenous. Moreover, we assume a strong correlation between the *selection* and the *dose* equation, as implied by a correlation between a and u equal to 0.8.

Results are set out in Table 2. It is immediate to see that the value of ATE obtained by the CT-OLS estimator is really close to the true ATE (9.22) and that the confidence interval at 5% of significance for this estimator strictly contains that value. But also the percentage bias of CT-IV is very low (0.86%) and comparable with CT-OLS (0.81%) and sufficient to imply that the 5% of significance contains the true ATE even in this case.

---

[5] The coefficient γ is equal to $(\rho^2/(1-\rho^2))^{-1/2}$, where ρ=corr(e_0;a). To get this result put x=e and y=a. We know that corr(x;y)=cov(x;y)/sd(x)sd(y). We can see that, while var(y)=1 by assumption, var(x)=γ^2+1. Moreover, cov(x;y)=cov(η+γa+v;a)=cov(η+γa;a)+cov(v;a)=cov(η+γa;a)=cov(γa;a)=γcov(a;a)=γvar(a)=γ. Thus, ρ=γ/(γ^2+1)$^{-1/2}$, that implies that γ=($\rho^2/(1-\rho^2)$)$^{-1/2}$.

Table 2. Mean test of ATE from Monte Carlo results using ctreatreg.
Exogenous selection is assumed.

	Mean	Std. Err.	[95% Confidence Interval]	
ATE (true value)	9.22	-	-	-
ATE - CT-OLS	9.21	0.01	9.19	9.22
ATE - CT-IV	9.20	0.01	9.19	9.22
% BIAS of OLS	0.81	0.04	0.73	0.90
% BIAS of IV	0.86	0.04	0.77	0.94

Note: ρ=0. Number of observations 10,000. Number of simulations 200.

These results confirm what was expected, thus showing that the option CT-OLS of ctreatreg behaves correctly. As a conclusion, when the analyst assumes exogeneity, he/she may reliably use ctreatreg with the option CT-OLS.

- *Case 2. Endogeneity*

If we assume that ρ=0.7, that is, a high positive correlation between the error term of the outcome equation (*e*) and the error term of the selection equation (*a*), then *w* becomes endogenous. For the sake of comparison, we still assume the same strong correlation between the *selection* and the *dose* equation (0.8).

Table 3 shows that results are - also in this case - very consistent with the theoretical predictions. Indeed, the percentage bias of model CT-OLS is rather high and equal to around 18%, whereas the bias of CT-IV is around 1%. Furthermore, and more importantly, the 95% mean test confidence interval for CT-IV contains the true ATE. As expected, this implies that CT-IV is an unbiased estimator in presence of selection endogeneity, thus leading to a reliable estimation of the true value of ATE.

Table 3. Mean test of ATE from Monte Carlo results using ctreatreg.
Endogenous selection is assumed.

	Mean	Std. Err.	[95% Confidence Interval]	
ATE (true value)	9.22	-	-	-
ATE - CT-OLS	7.53	0.01	7.51	7.55
ATE - CT-IV	9.22	0.01	9.20	9.24
% BIAS of OLS	18.26	0.11	18.05	18.48
% BIAS of IV	1.28	0.07	1.15	1.41

Note: ρ=0.7. Number of observations 10,000. Number of simulations 200.

Overall, these results confirm the reliability of the model and that of `ctreatreg` by allowing for a trustful use of this model and this command either under selection exogeneity or endogeneity.

Finally, figure 1 plots the Dose-Response-Function along with the 95% interval confidence lines for both models. This is done by exploiting the option "graphdrf" of `ctreatreg`.

4. CONCLUSION

In this paper, an original econometric model for estimating a Dose Response Function through a regression approach when treatment is *continuous*, individuals may react heterogeneously to observable confounders and selection-into-treatment may be endogenous has been presented. Two estimation procedures are thus set out: one based on OLS under conditional mean independence, and one based on IV under selection endogeneity.
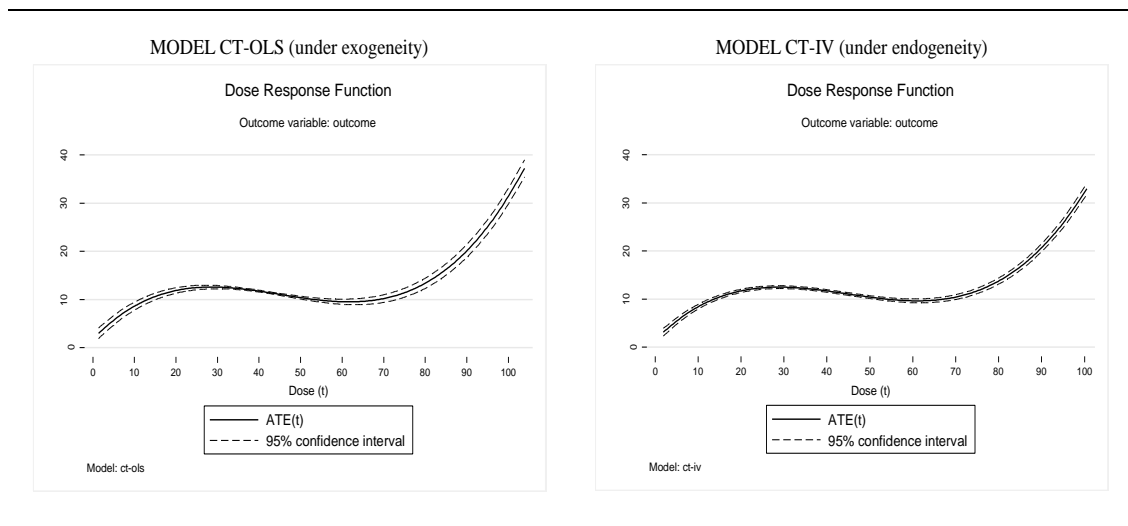

Figure 1. Graph of the Dose-Response-Function (DRF) using the `ctreatreg` *option* "ct-ols" *and* "ct-iv" *under exogeneity and endogeneity respectively.*

The paper presents also `ctreatreg`, a user-written Stata routine for an easy implementation of such a model. As application, a Monte Carlo experiment to test the reliability of the model and of its associated Stata routine has been performed. Monte Carlo results show that the model and the Stata routine `ctreatreg` are both reliable as estimates consistently fit the expected results.

REFERENCES

Bia M., and Mattei A. (2008), "A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score", *The Stata Journal*, 8, 3, 354–373.

Bia M., Flores C. and Mattei A. (2011), Nonparametric Estimators of Dose-Response Functions, CEPS/INSTEAD Working Paper Series 2011-40, CEPS/INSTEAD.

Cerulli G. (2012), "ivtreatreg: a new STATA routine for estimating binary treatment models with heterogeneous response to treatment under observable and unobservable selection", *CNR-Ceris Working Papers*, No. 03/12. Available at: http://econpapers.repec.org/software/bocbocode/s457405.htm.

Hirano K., and Imbens G. (2004), "The propensity score with continuous treatments". In Gelman, A. & Meng, X.L. (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (73-84), New York: Wiley.

Wooldridge J.M. (1997), "On two stage least squares estimation of the average treatment effect in a random coefficient model", *Economics Letters*, 56, 2, 129-133.

Wooldridge J.M. (2003), "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model", *Economics Letters*, 79, 185-191.

Wooldridge J.M. (2010), *Econometric Analysis of cross section and panel data*. Chapter 18. Cambridge: MIT Press.

Working Paper Cnr-Ceris

ISSN (*print*): 1591-0709   ISSN (*on line*): 2036-8216

Download

www.ceris.cnr.it/index.php?option=com_content&task=section&id=4&Itemid=64

Hard copies are available on request,
please, write to:

Cnr-Ceris
Via Real Collegio, n. 30
10024 Moncalieri (Torino), Italy
Tel. +39 011 6824.911   Fax +39 011 6824.966
segreteria@ceris.cnr.it   www.ceris.cnr.it